The bridge to possible

Overlay Multicast in VXLAN EVPN

Understanding fundamental concepts and architecture

Tarique Shakil, Principal Technical Marketing Engineer Cloud Networking, Cisco Systems

Cisco Webex App

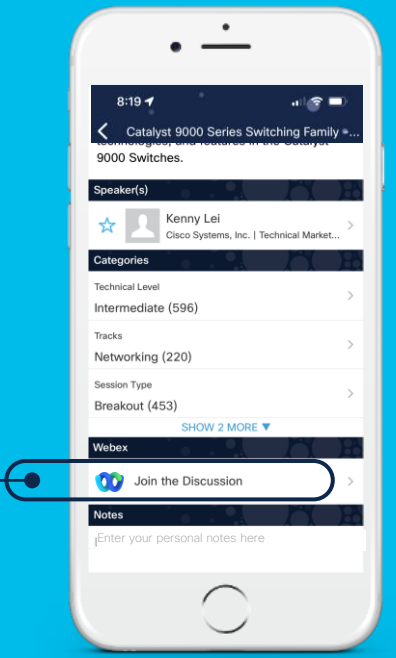
Questions?

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 until February 24, 2023.



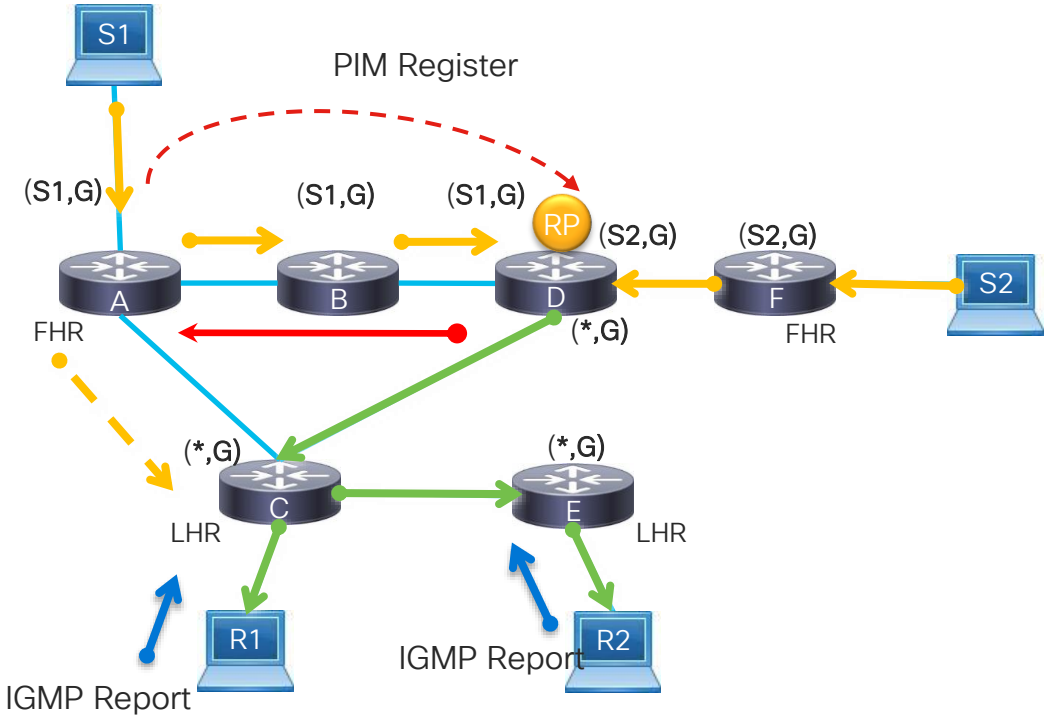
Agenda

- Multicast Routing Concepts
- VXLAN EVPN Multicast Forwarding
- MP-BGP NGMVPN Concepts
- VXLAN EVPN TRM Architecture
- VXLAN EVPN TRM Forwarding
- Configuring VXLAN EVPN TRM
- Data MDT
- Summary

Multicast Routing Concepts

Multicast Distribution Tree (MDT)

Shared Tree – PIM SM



*Usually optimized by switching to the source tree (default behavior)

FHR Generates PIM register to notify RP about new source (unicast tunnel)

RP generates PIM **register-stop (S1,G)** to notify **FHR** that the registration is complete

IGMP Report

(*,G) (**AnySource, Group**)

RP PIM Rendezvous Point

Source Tree

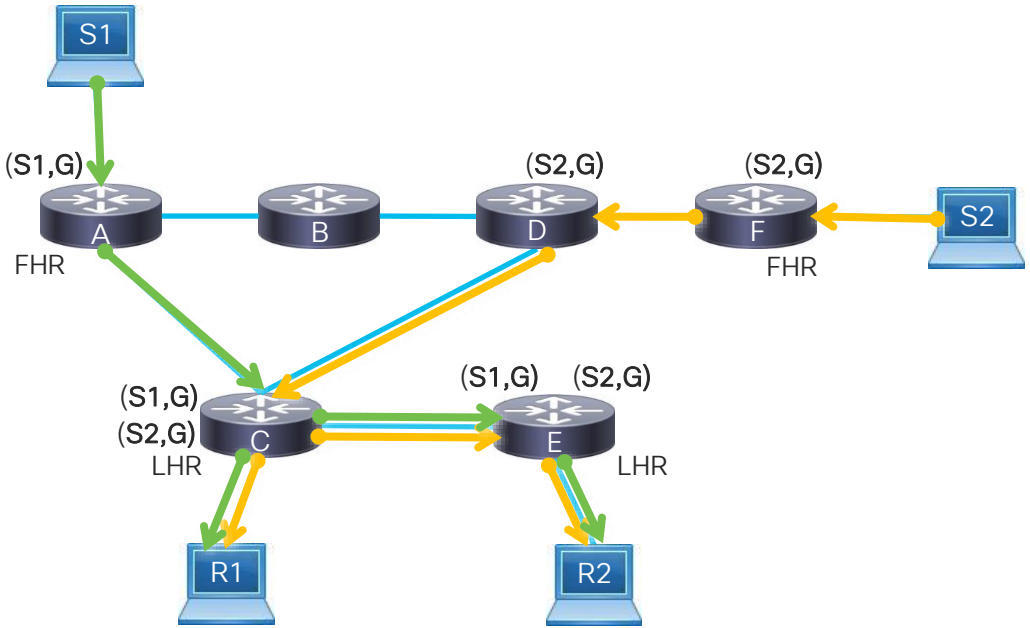
Shared Tree

SPT Switchover

- Every node should know who is the RP
- (*,G) consumes less memory, but may introduce sub-optimal path from source to all receivers*

Multicast Distribution Tree (MDT)

Shortest Path Tree – PIM SSM



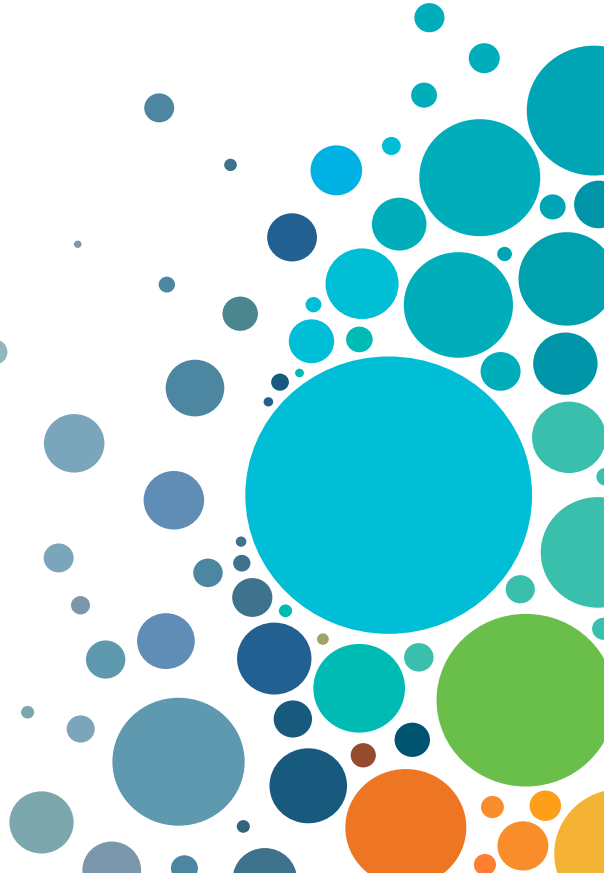
(S,G) (Source, Group)

Source Tree (S2, G)

Source Tree (S1, G)

- No need for RP
- (S,G) consumes more memory, but is always optimal. Group address can be reused

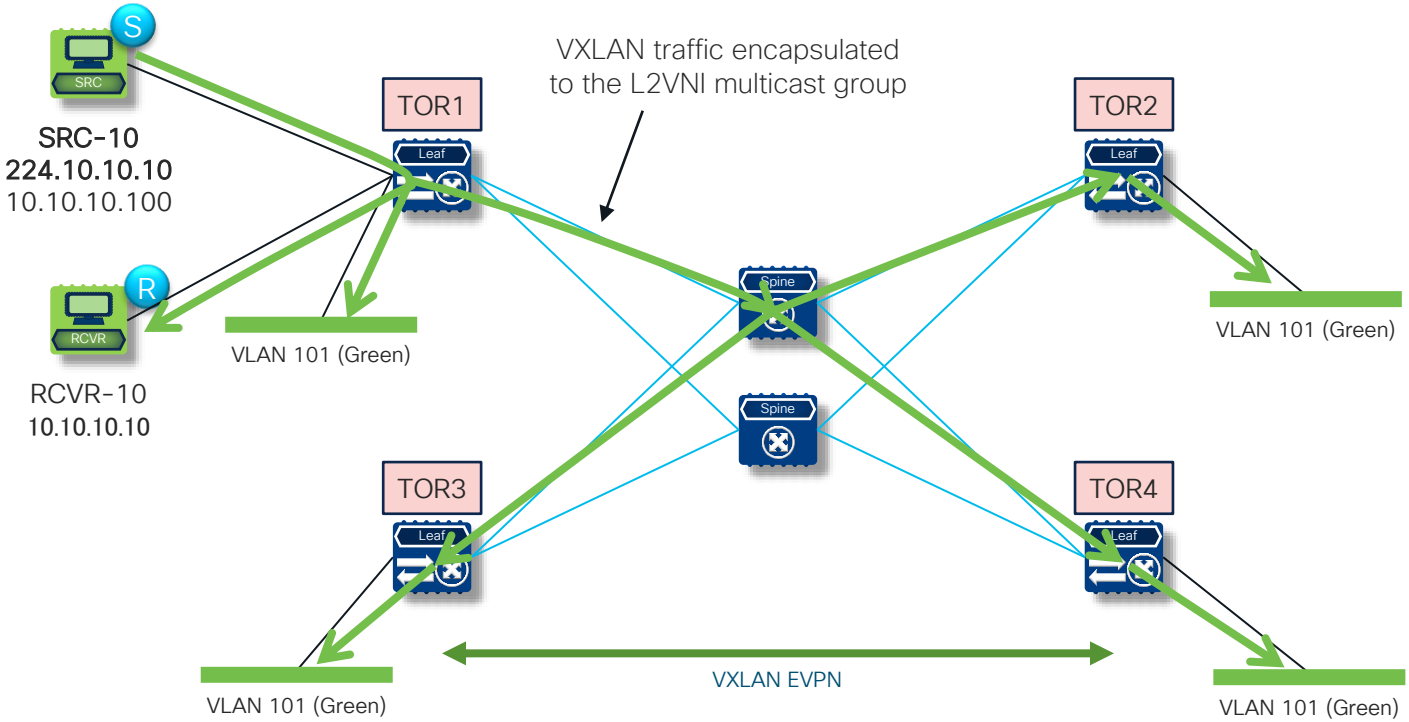
VXLAN EVPN Multicast Forwarding



cisco Live!

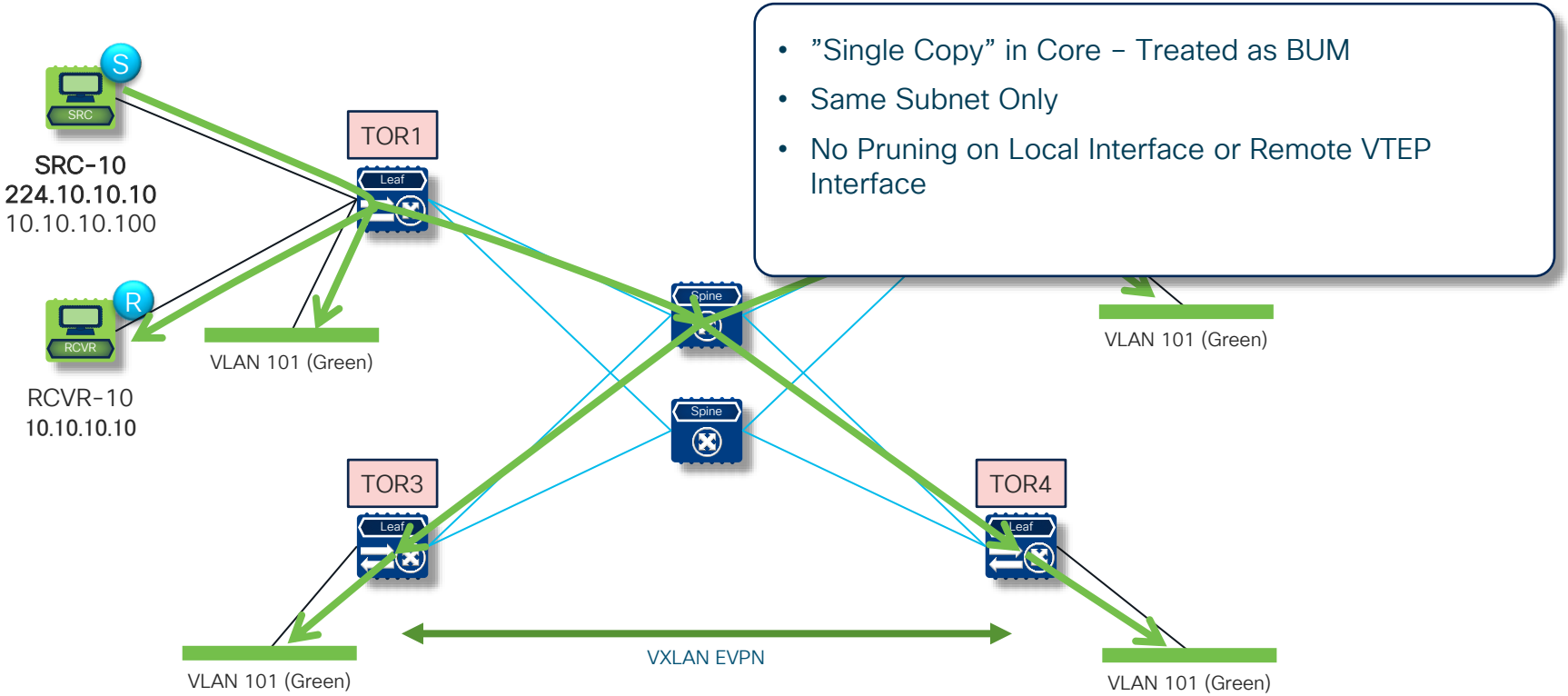
Same Subnet Forwarding no IGMP Snooping

Default Multicast Forwarding in VXLAN Overlay



Same Subnet Forwarding no IGMP Snooping

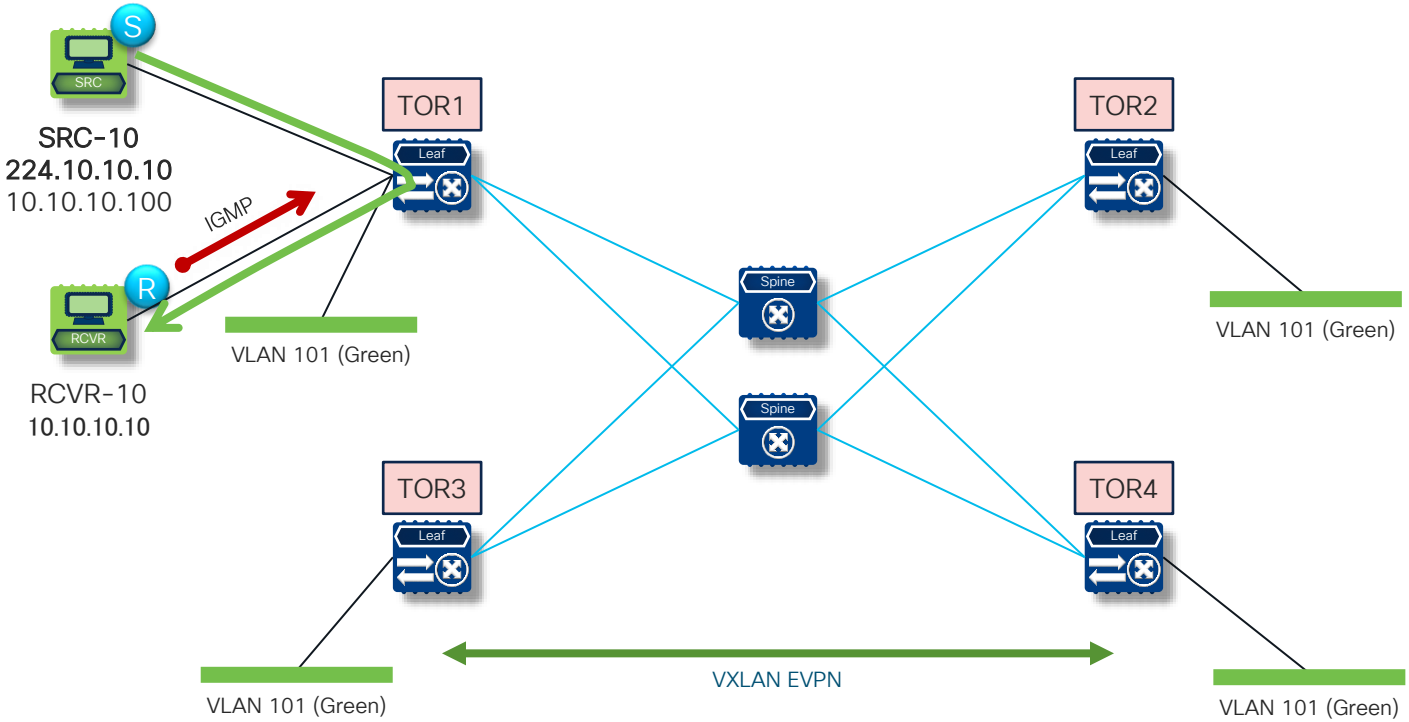
Default Multicast Forwarding in VXLAN Overlay



- ”Single Copy” in Core – Treated as BUM
- Same Subnet Only
- No Pruning on Local Interface or Remote VTEP Interface

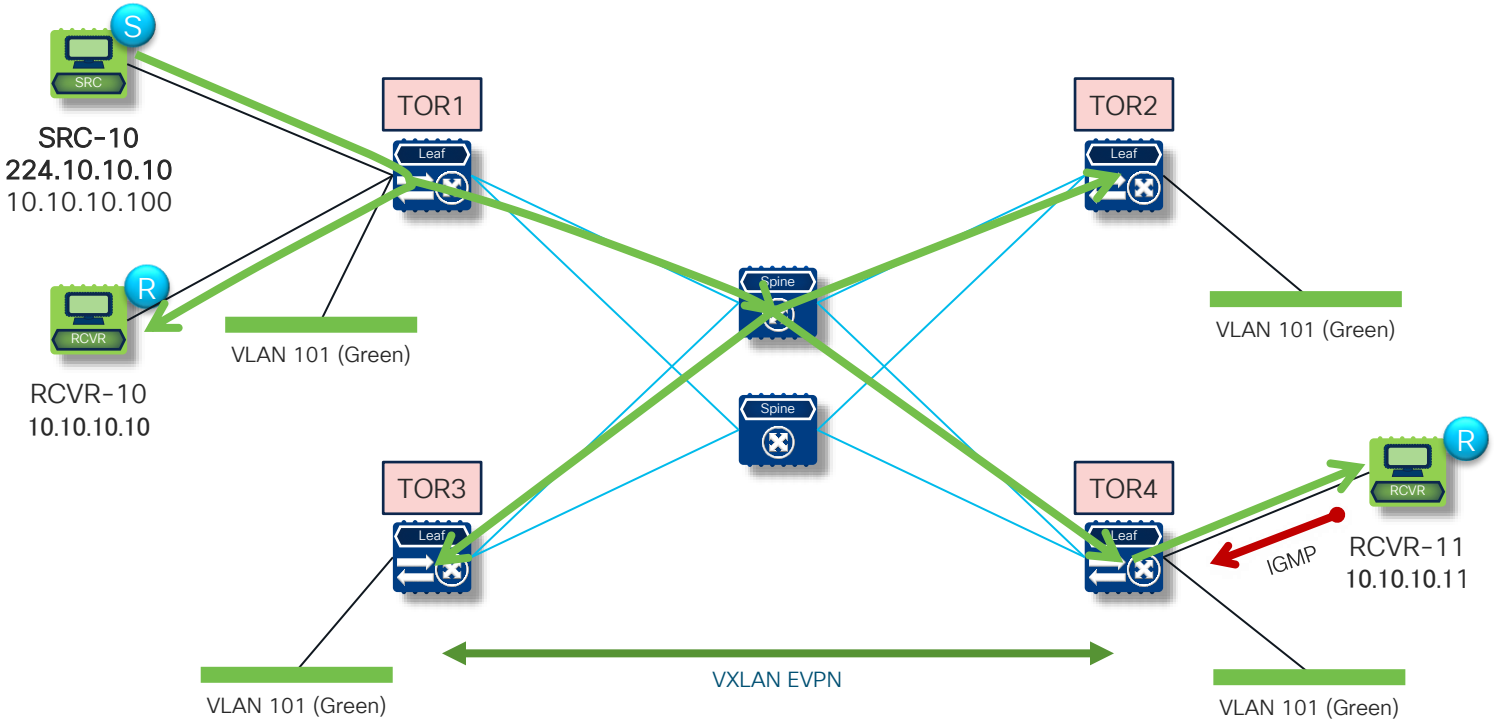
Same Subnet Forwarding with IGMP Snooping

Default Multicast Forwarding in VXLAN Overlay



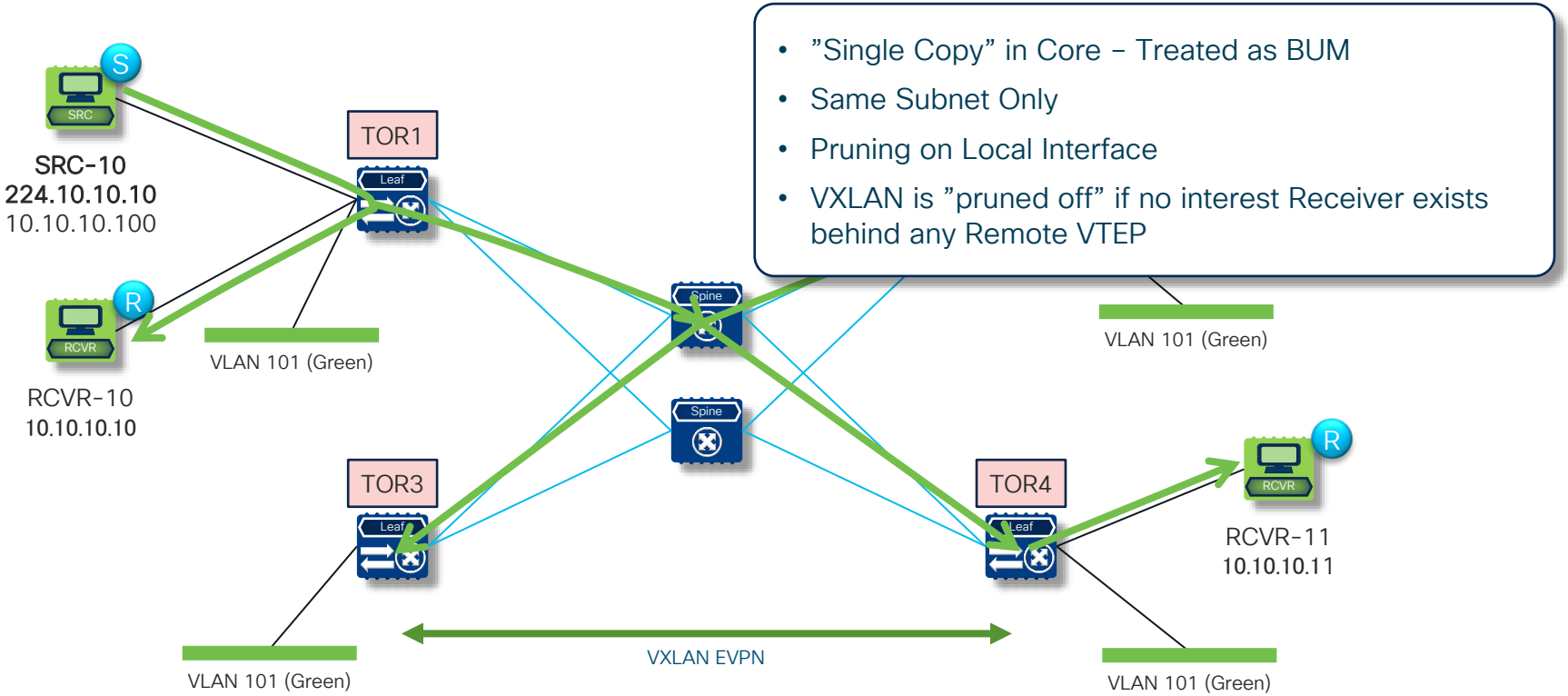
Same Subnet Forwarding with IGMP Snooping

Default Multicast Forwarding in VXLAN Overlay



Same Subnet Forwarding with IGMP Snooping

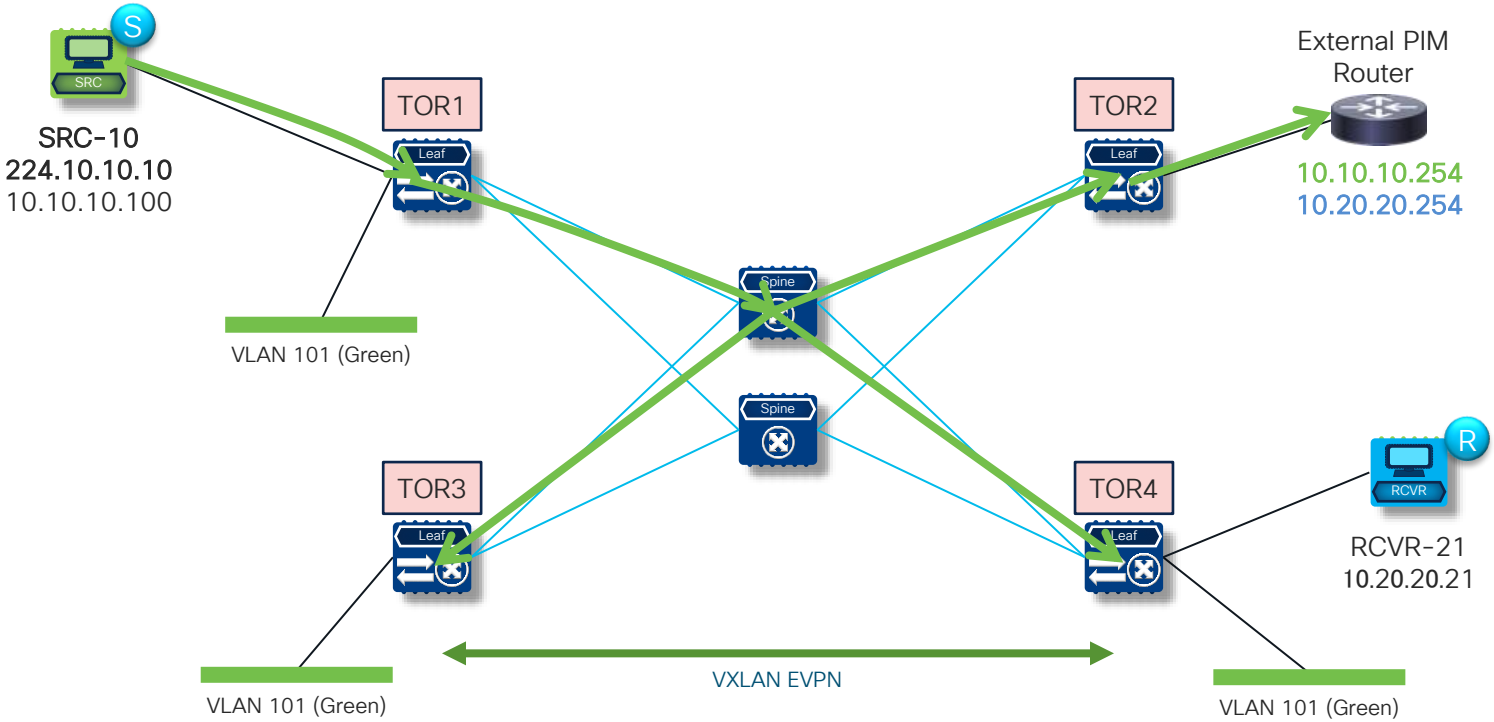
Default Multicast Forwarding in VXLAN Overlay



- ”Single Copy” in Core – Treated as BUM
- Same Subnet Only
- Pruning on Local Interface
- VXLAN is ”pruned off” if no interest Receiver exists behind any Remote VTEP

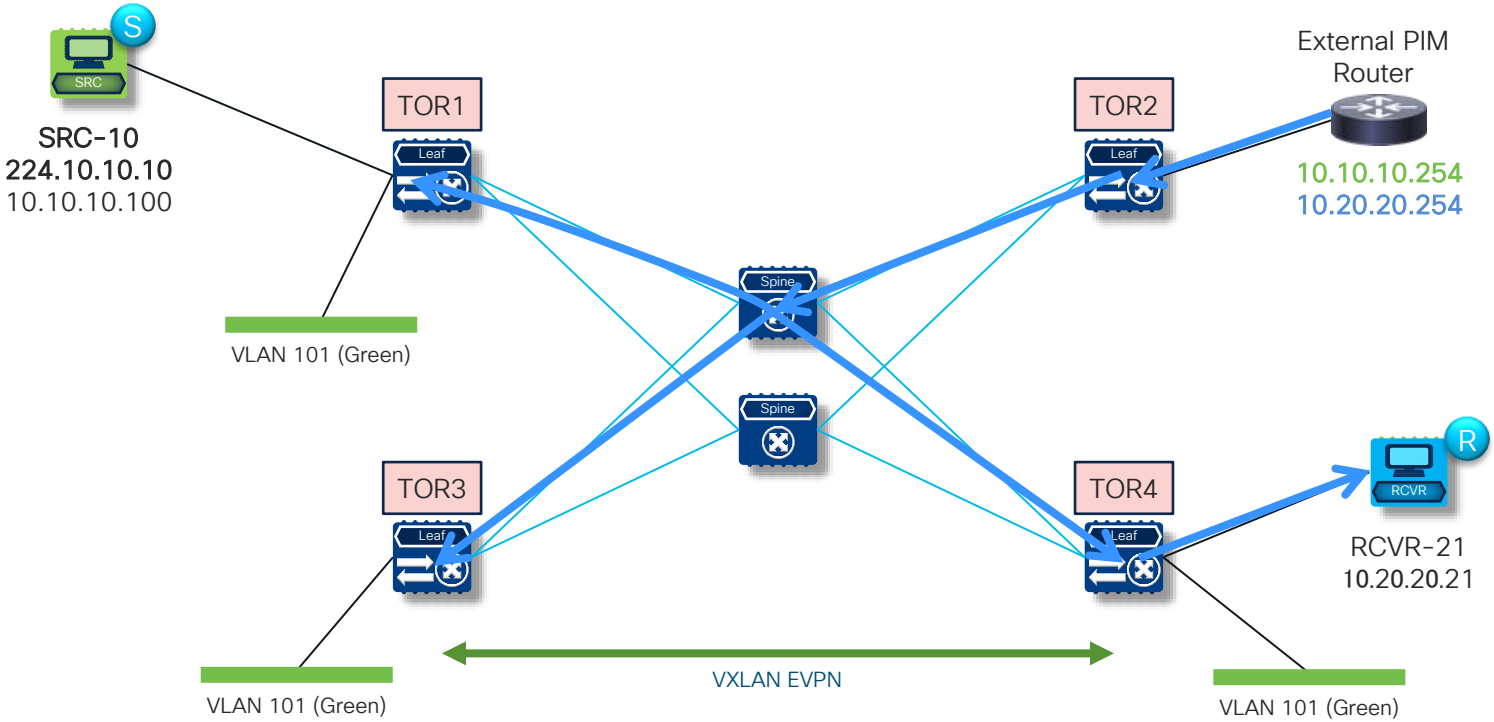
Different Subnet Forwarding – Router on-a-Stick

Default Multicast Forwarding in VXLAN Overlay



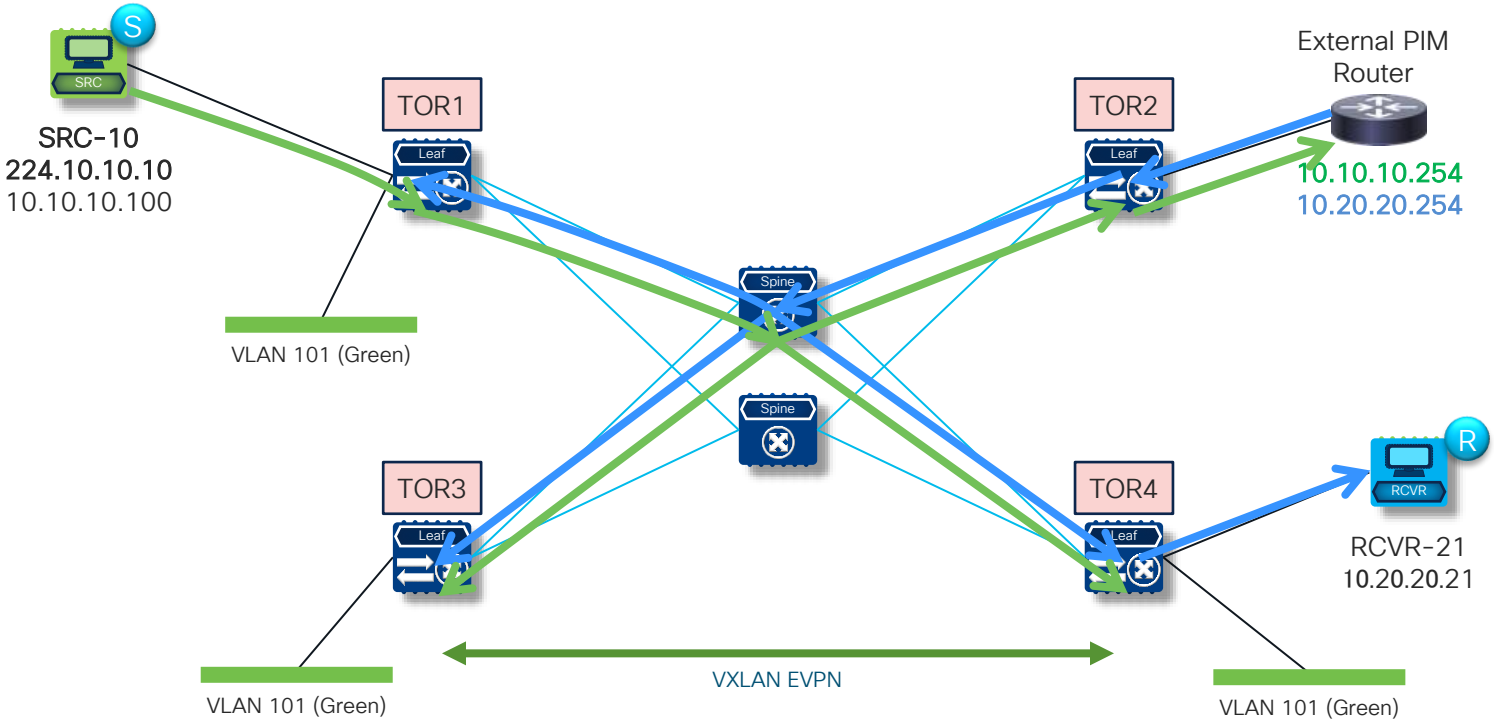
Different Subnet Forwarding – Router on-a-Stick

Default Multicast Forwarding in VXLAN Overlay



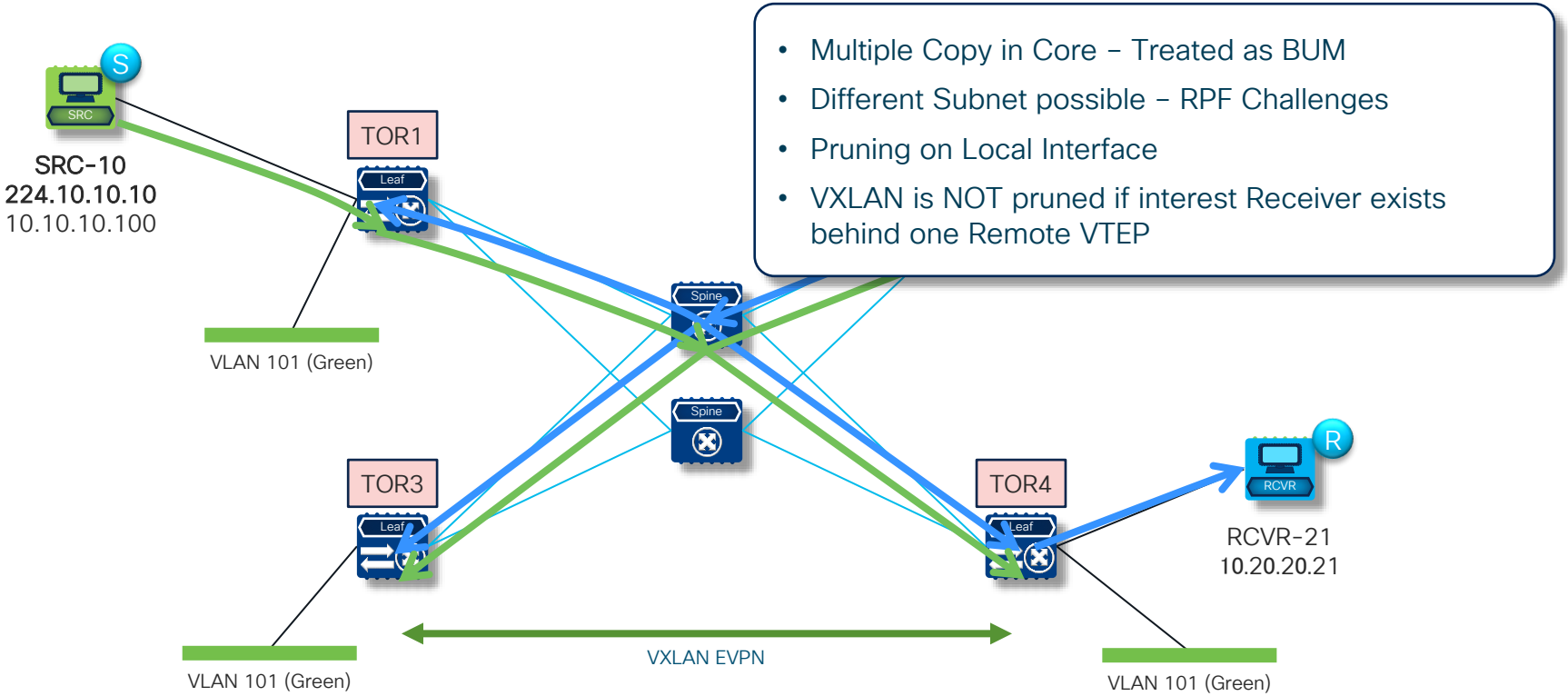
Different Subnet Forwarding – Router on-a-Stick

Default Multicast Forwarding in VXLAN Overlay



Different Subnet Forwarding – Router on-a-Stick

Default Multicast Forwarding in VXLAN Overlay

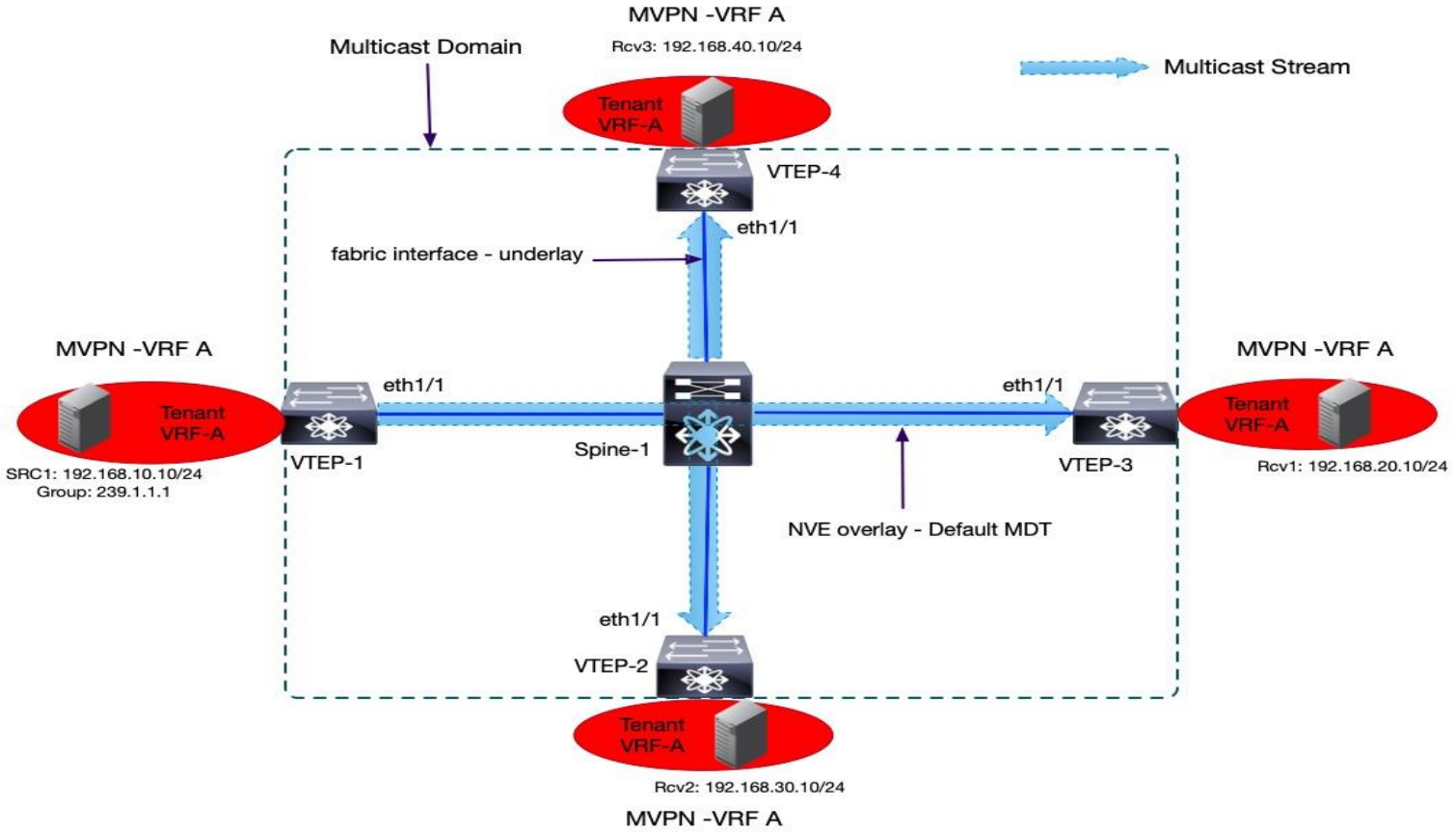


- Multiple Copy in Core – Treated as BUM
- Different Subnet possible – RPF Challenges
- Pruning on Local Interface
- VXLAN is NOT pruned if interest Receiver exists behind one Remote VTEP

MP-BGP NGMVPN Concepts

cisco Live!

MP-BGP NGMVPN Terminology



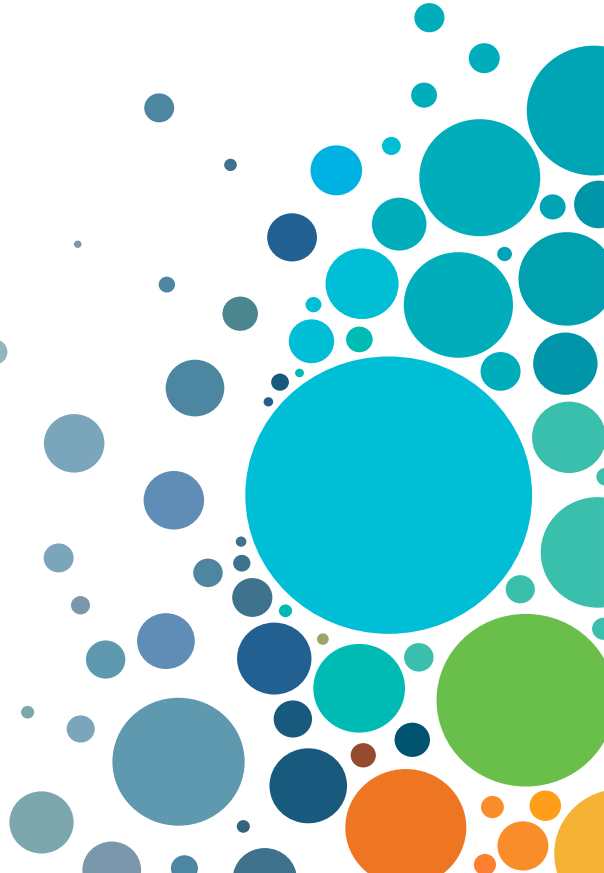
MP-BGP NGMVPN Control Plane

- MP-BGP is used to exchange both unicast (AF EVPN) and multicast (AF MVPN) route information in a VXLAN BGP EVPN fabric.
- Function:
  - “who are the members of my multicast domain?”.
  - “Which tunnel do I send my multicast traffic on?”
  - “Which multicast groups can receivers subscribe to and who are the sources for those groups?”
  - Nexus 9000 NXOS implementation based on RFC 6513 and 6514

MP-BGP NGMVPN Packet Types

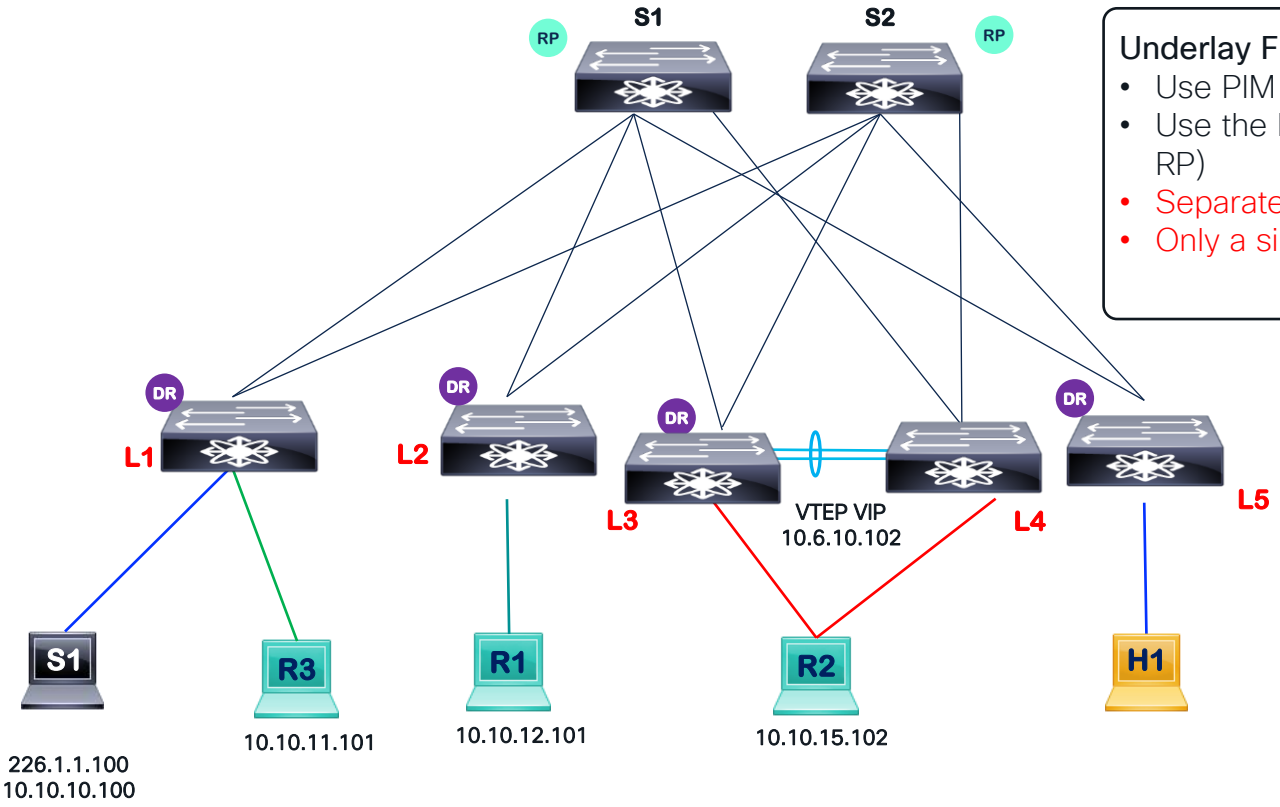
Route Type	Name	Description
5	Source Active AD Route	Originated by the FHR/VTEP with active source. Triggered by FHR/VTEP receiving multicast traffic on tenant VRF interface. Used to advertise the existence of an attached source for a specific multicast stream.
6	Shared Tree Join Route	Originated by the LHR/VTEP with an active receiver. Triggered by receiving shared tree join (C-*, C-G) on tenant VRF interface. Used in L3 mode with external RP.
7	Source Tree Join Route	Originated by the LHR/VTEP with an active receiver. Triggered by receiving a PIM join on tenant VRF interface and in response to an MVPN Type 5 route.

VXLAN EVPN TRM Architecture



cisco Live!

VXLAN EVPN TRM Underlay Routing



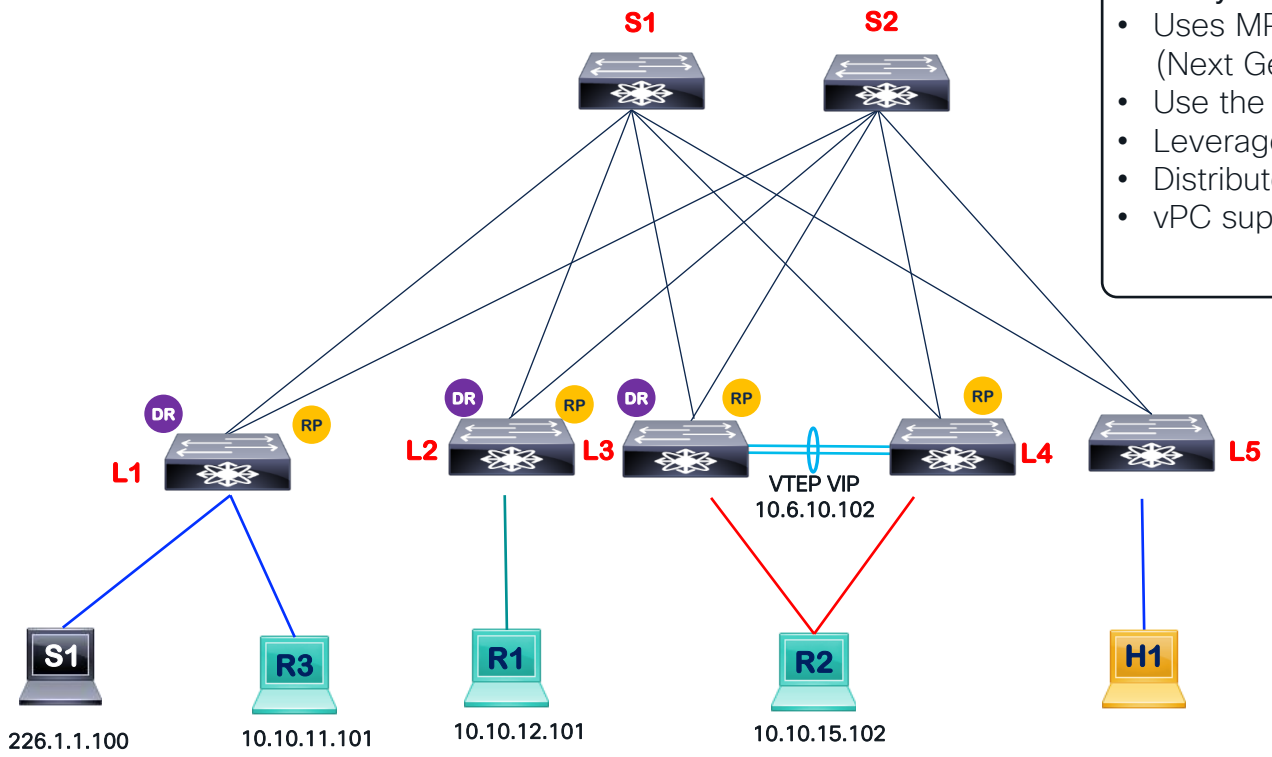
Underlay Functional Details:

- Use PIM based underlay for transport (PIM ASM)
- Use the RP defined in the underlay (PIM Anycast RP)
- Separate MCAST Group for each L3VNI(VRF)
- Only a single packet in the core

RP Underlay Rendezvous Point

DR Designated Router

VXLAN EVPN TRM Overlay Routing



Overlay Functional Details:

- Uses MP-BGP based ngMVPN control plane (Next Gen Multicast VPN)
- Use the Router Reflector in the Spine
- Leverages RP-less (in fabric RP) configuration
- Distributed Anycast Designated Router (DR)
- vPC support and non TRM VTEP Integration

RP Rendezvous Point

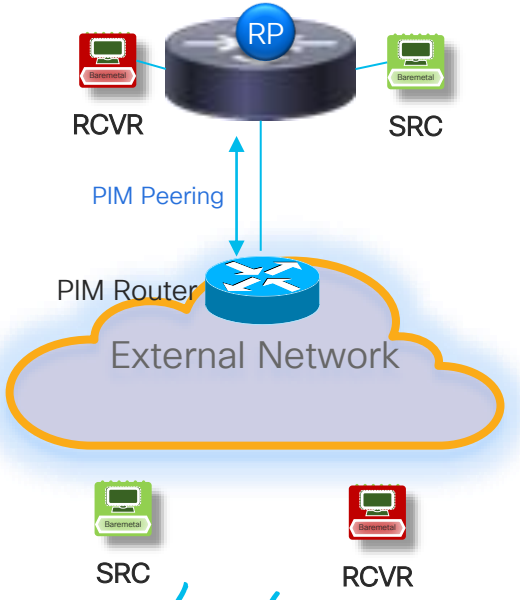
DR Designated Router

Tenant Routed Multicast

RP Deployment Models

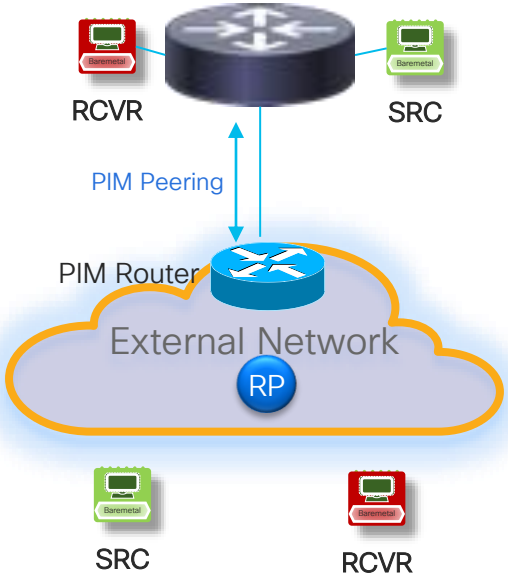
RP-Less

VXLAN EVPN Fabric as logical PIM Router (playing also the RP role)



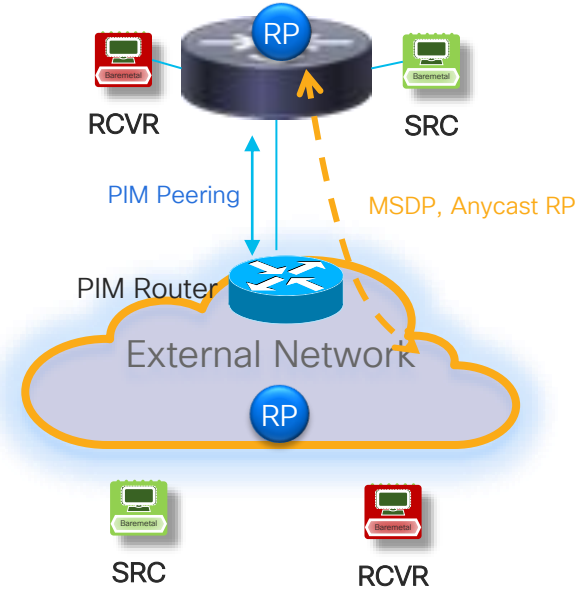
External RP

VXLAN EVPN Fabric as logical PIM Router



RP Anywhere

VXLAN EVPN Fabric as logical PIM Router (playing also the RP role)

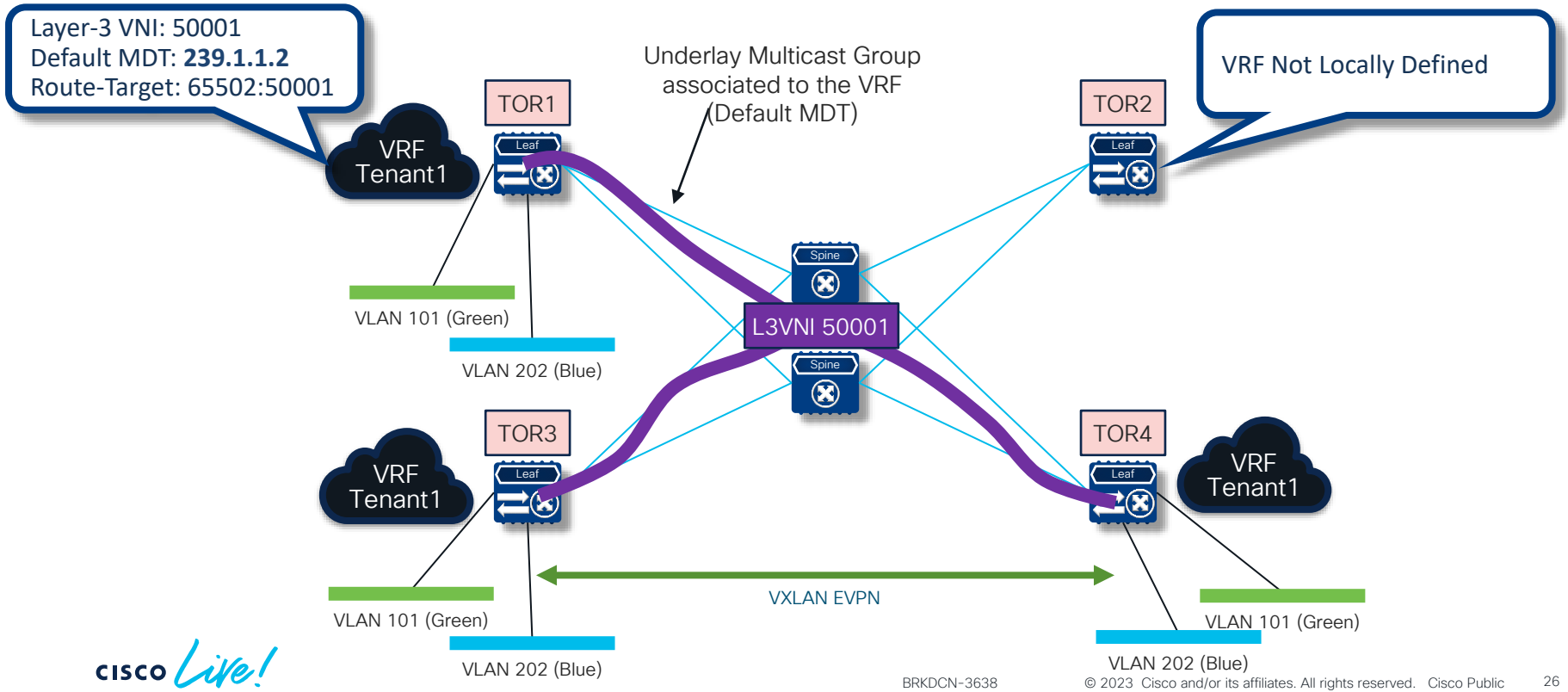


VXLAN EVPN TRM Forwarding

cisco Live!

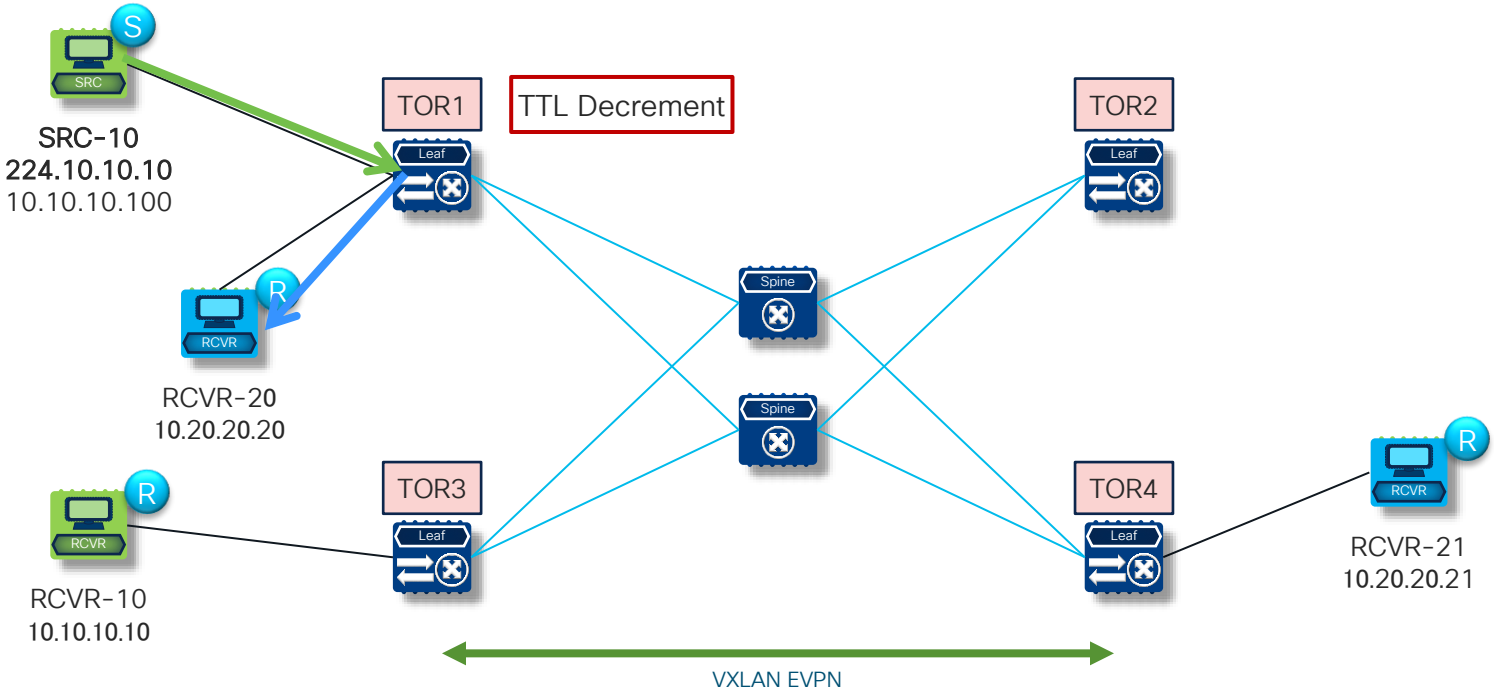
TRM Forwarding

Always Route Approach



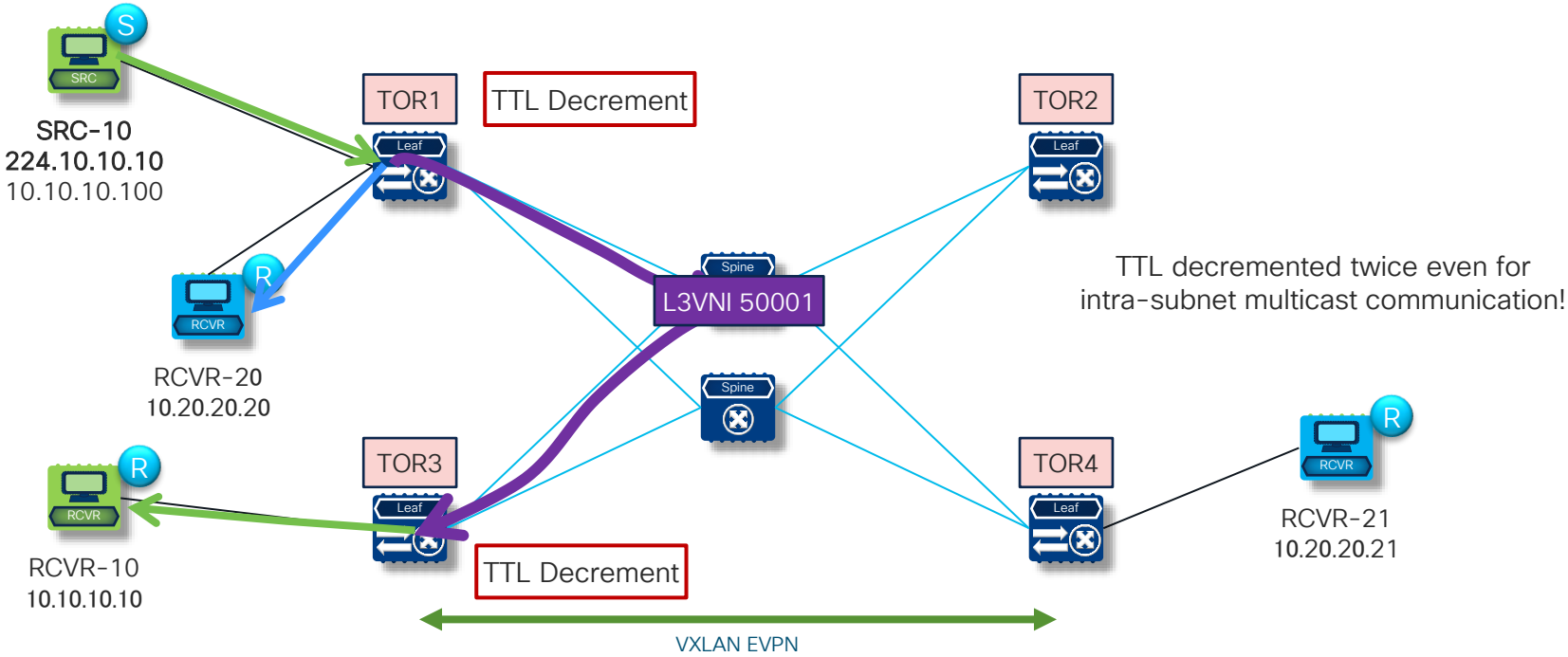
TRM Forwarding

Always Route Approach



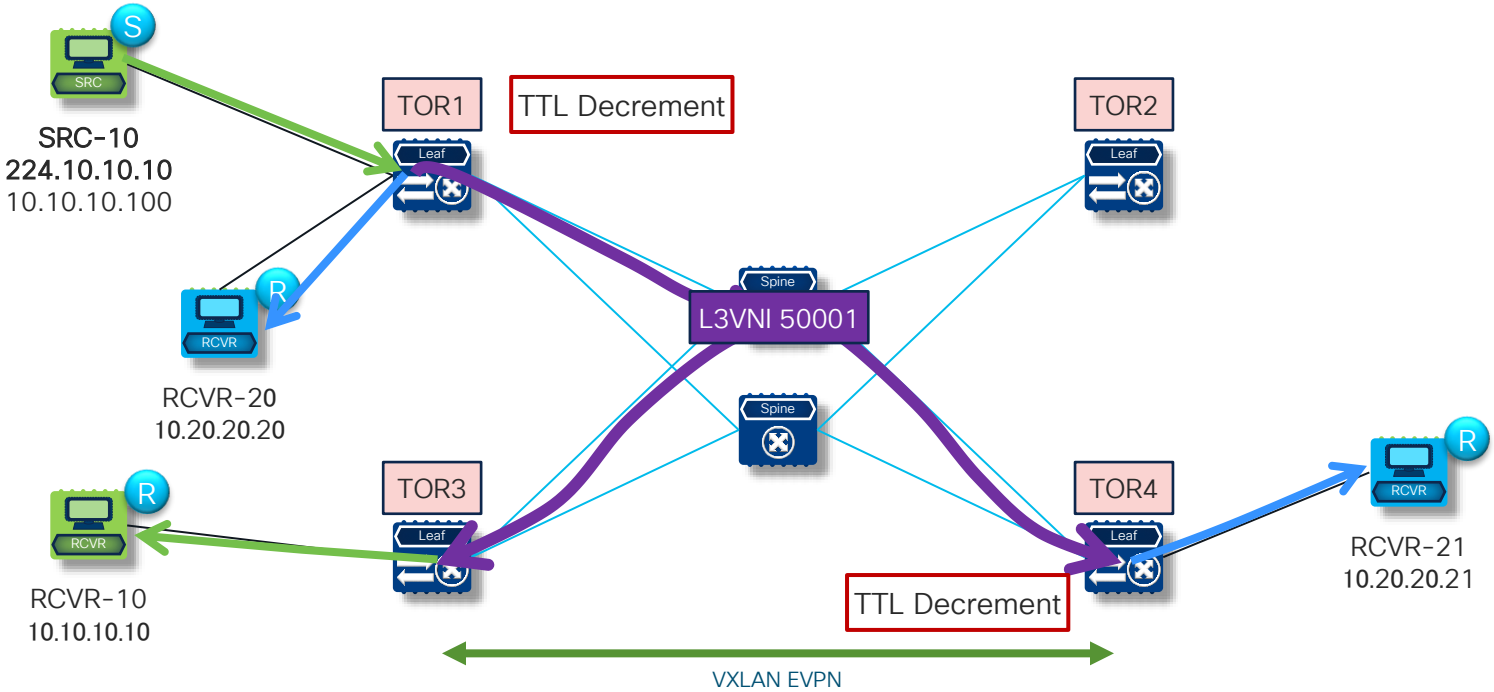
TRM Forwarding

Always Route Approach



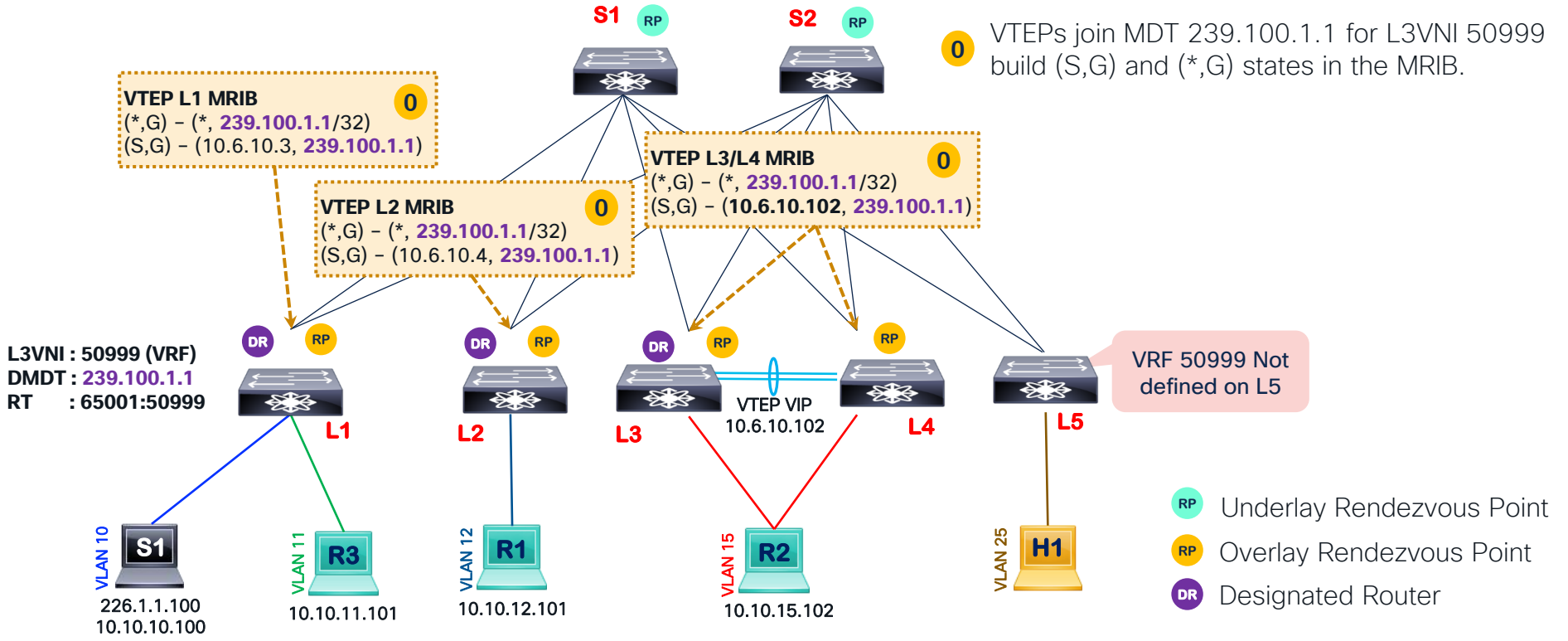
TRM Forwarding

Always Route Approach



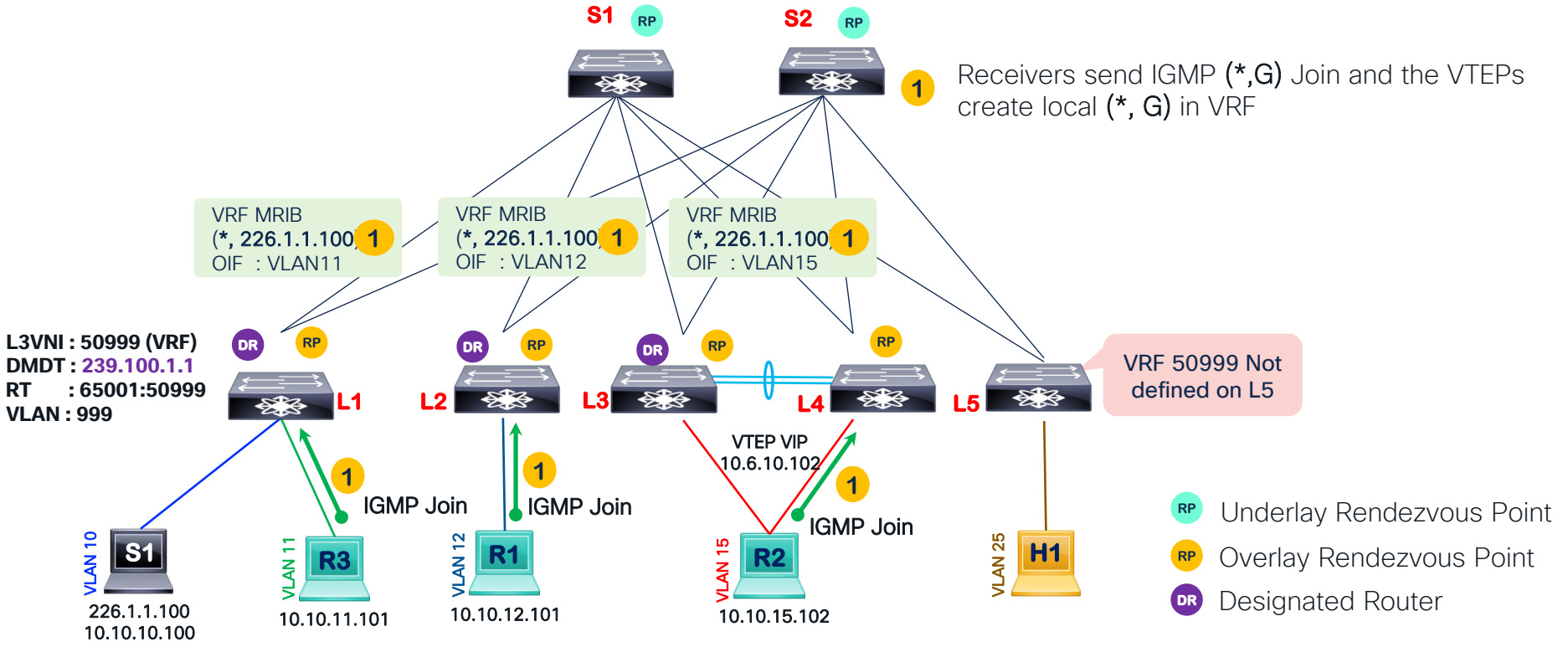
TRM Routing with Anycast RP

Underlay Multicast State



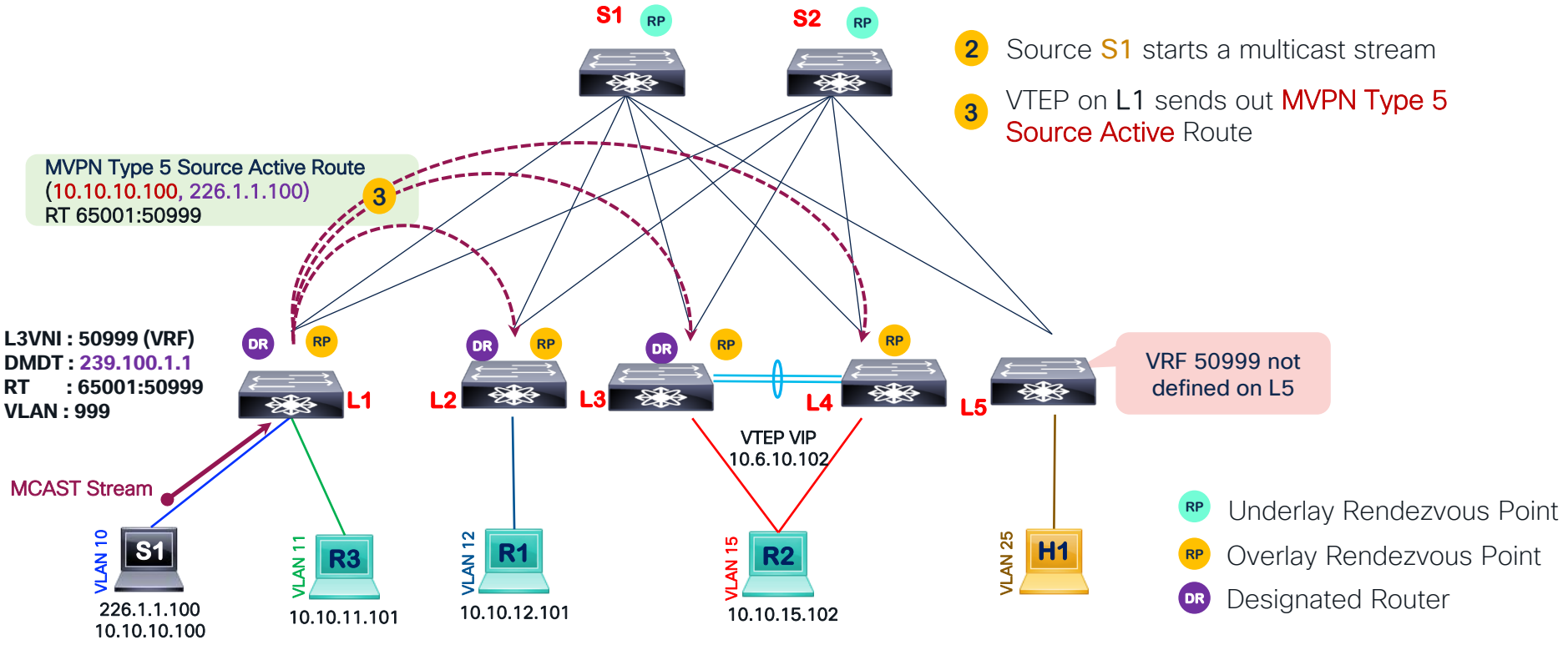
TRM Routing with Anycast RP

Packet Walk



TRM Routing with Anycast RP

Packet Walk



VRF Route Import

Extended Community

```
abcpod-1-dc1-bgw1# show bgp l2vpn evpn route-type 2 vrf tenant-1
Route Distinguisher: 10.6.10.4:3    (L3VNI 50999)
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.0000.0c00]:[32]:[10.10.10.100]/272, version 109
Paths: (1 available, best #1)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
             Imported from 10.6.10.2:32782:[2]:[0]:[0]:[48]:[0050.0000.0c00]:[32]:[10.10.10.100]/272
  AS-Path: NONE, path sourced internal to AS
    10.6.11.2 (metric 81) from 10.6.10.1 (10.6.10.1)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30015 50999
      Extcommunity: RT:65001:30015 RT:65001:50999 Route-Import:10.6.11.2:999
```

MPVPN Route Type 5

MP-BGP RIB Leaf 1 (FHR)

```
abcpod-1-dc1-leaf1# show bgp ipv4 mvpn route-type 5 detail vrf Tenant-1
Route Distinguisher: 10.6.10.2:3    (L3VNI 50999)
BGP routing table entry for [5][10.10.10.100][226.1.1.100]/64, version 7
Paths: (1 available, best #1)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in mvpn

  Advertised path-id 1
  Path type: local, path is valid, is best path, no labeled nexthop
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (10.6.10.2)
      Origin IGP, MED not set, localpref 100, weight 32768
      Extcommunity: RT:65001:50999

  Path-id 1 advertised to peers:
    10.6.10.1
```

Overlay Multicast Group

Multicast Source IP

MVPN Type 5 Route

Route Target AS:L3VNI

MPVPN Route Type 5

MP-BGP RIB Leaf 2 (LHR)

```
abcpod-1-dc1-leaf2# show bgp ipv4 mvpn route-type 5 detail vrf tenant-1
Route Distinguisher: 10.6.10.3:3    (L3VNI 50999)
BGP routing table entry for [5][10.10.10.100][226.1.1.100]/64, version 10
Paths: (1 available, best #1)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in mvpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
             Imported from 10.6.10.2:3:[5][10.10.10.100][226.1.1.100]/64
  AS-Path: NONE, path sourced internal to AS
    10.6.11.2 (metric 81) from 10.6.10.1 (10.6.10.1)
      Origin IGP, MED not set, localpref 100, weight 0
      Extcommunity: RT:65001:50999
      Originator: 10.6.10.2 Cluster list: 10.6.10.1
```

Overlay Multicast Group

Multicast Source IP

MVPN Type 5 Route

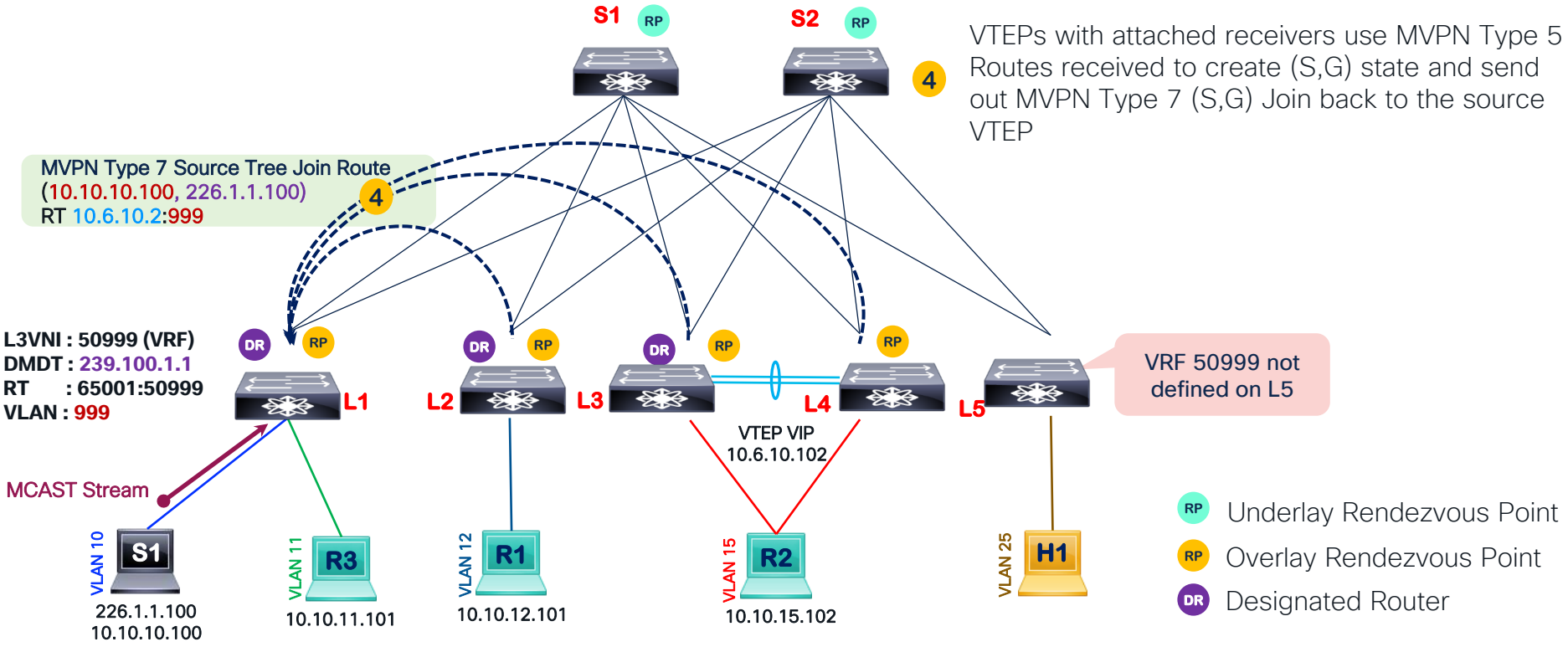
Import based on MVPN Type5 received

Where the source is attached

Route Target AS:L3VNI

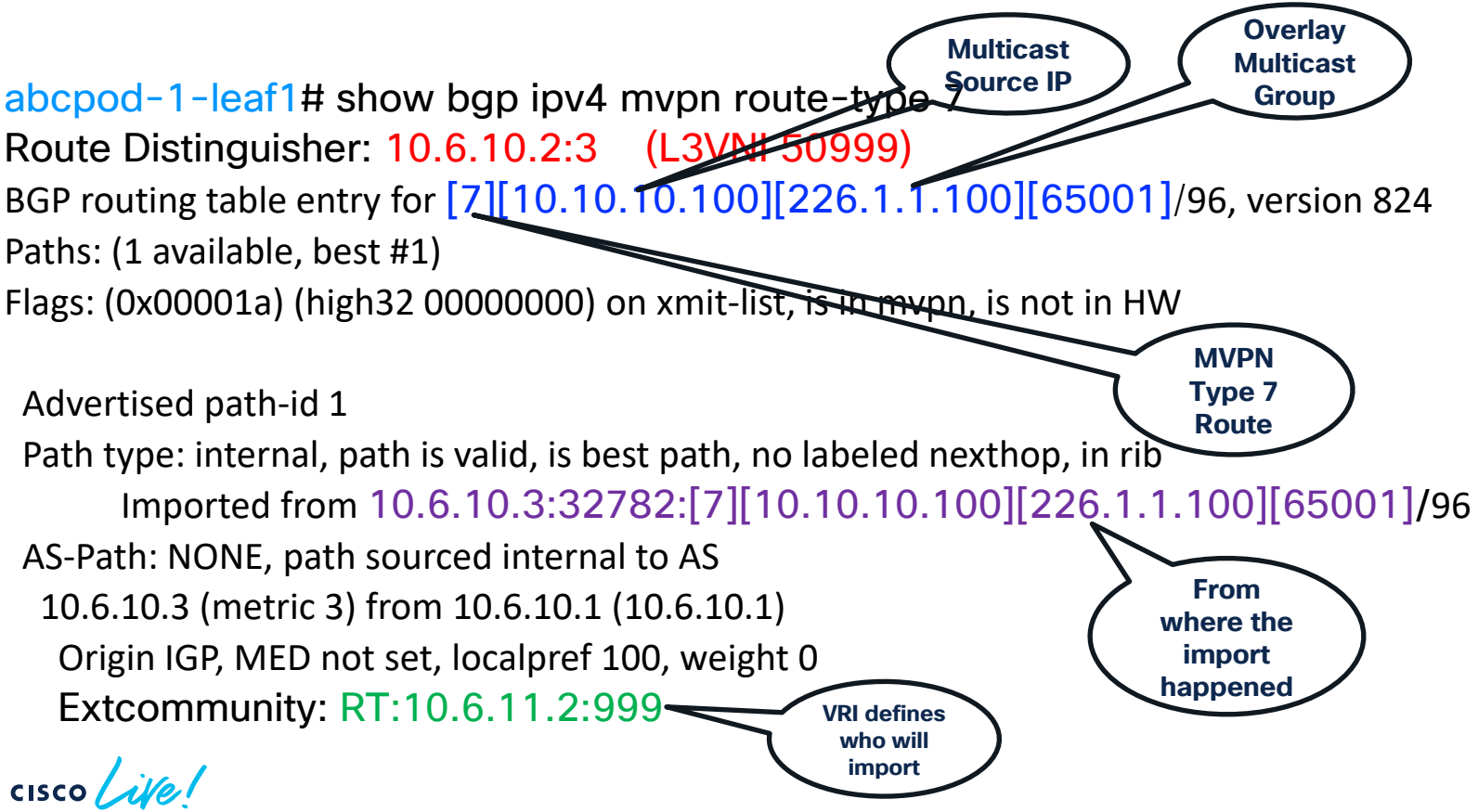
TRM Routing with Anycast RP

Packet Walk



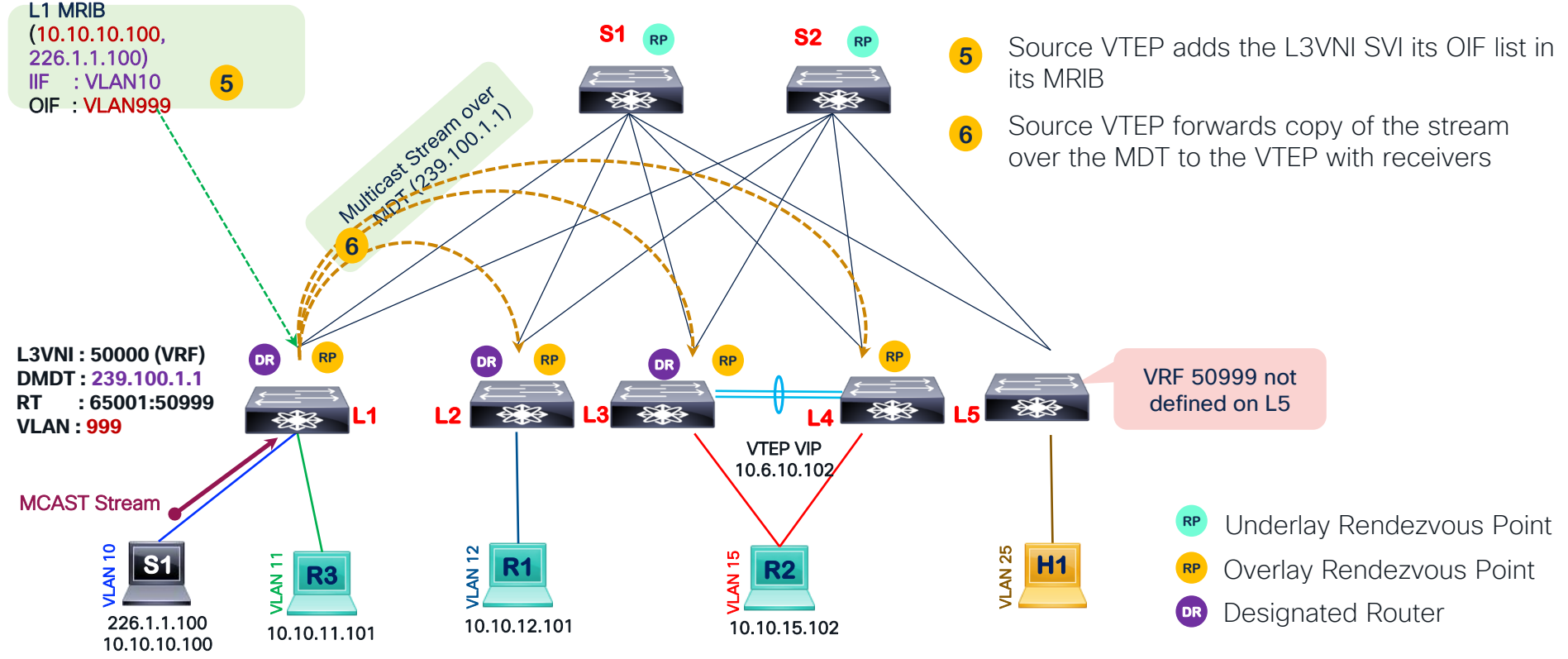
MPVPN Route Type 7

MP-BGP RIB Leaf 1 (FHR)



TRM Routing with Anycast RP

Packet Walk



5 Source VTEP adds the L3VNI SVI its OIF list in its MRIB

6 Source VTEP forwards copy of the stream over the MDT to the VTEP with receivers

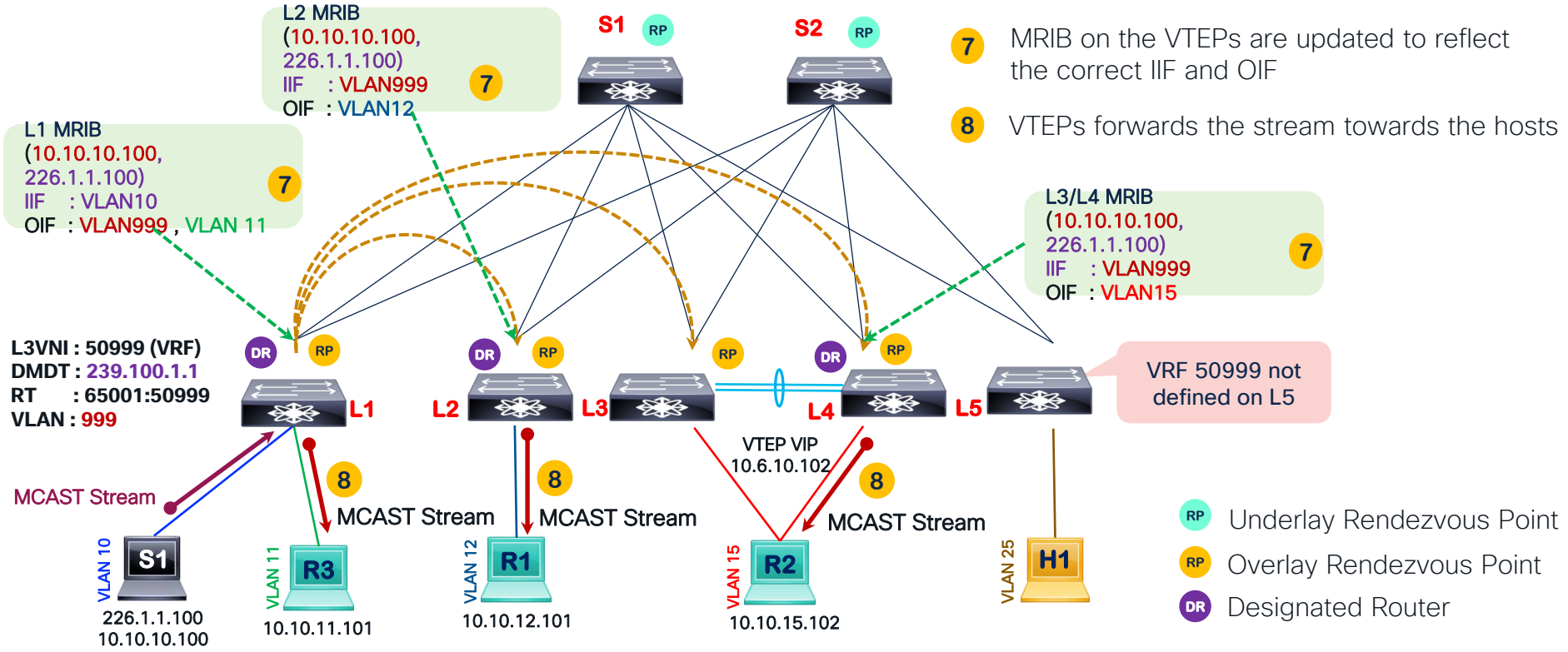
RP Underlay Rendezvous Point

RP Overlay Rendezvous Point

DR Designated Router

TRM Routing with Anycast RP

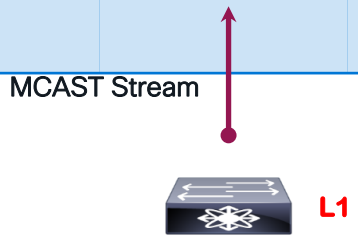
Packet Walk



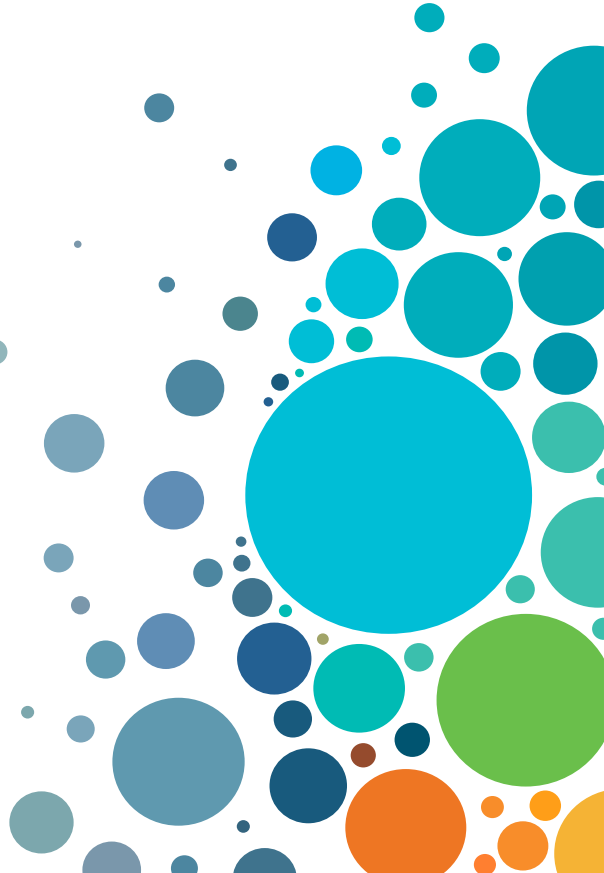
TRM Packet Encapsulation

Multicast In Multicast

Underlay SIP	Underlay DIP	VNI	Overlay SIP	Overlay DIP
10.6.11.2 NVE-PIP	239.100.1.1 Default-MDT	50999 L3 VNI	10.10.10.100	226.1.1.100



Configuring VXLAN EVPN TRM



cisco Live!

VXLAN EVPN TRM Configuration Guidelines

- TRM uses an "Always Route" approach in the overlay.
- TRM requires an IPv4 underlay.
- TRM is only supported when PIM Any Source Multicast (ASM) is used in the underlay
- TRM is not supported with PIM BiDir in the underlay
  - PIM BiDir is supported for Unicast in the underlay
- TRM also supports IPv6 Multicast in the overlay as of NXOS 10.2.1
  - MLD snooping with VxLAN VLANs with TRM
- TRM only supports PIM ASM and PIM SSM in the overlay
- 224.0.0.0/24 subnet (local scope) is excluded from TRM and is always bridged
- TRM also supports Multi-site VXLAN BGP EVPN for IPV4 and IPv6 in the overlay.

VXLAN EVPN TRM with Anycast RP

Step 1: Enable the feature for the routing processes required for VXLAN BGP EVPN and TRM on the nodes.

```
nv overlay evpn
feature bgp
feature pim
feature interface-vlan
feature vn-segment-vlan-based
feature nv overlay
feature ngmvpn
```

”feature ngmvpn” will enable the Next-Generation Multicast VPN (ngMVPN) control-plane allowing you to enable MVPN address family under BGP routing process.

nv overlay evpn, feature nv overlay and feature vn-segment-vlan-based are VXLAN EVPN unicast-routing features that must be enabled.

VXLAN EVPN TRM with Anycast RP

Step 2: Enable the MVPN address family under the BGP routing process.

```
router bgp 65501
  neighbor 10.100.100.201
    remote-as 65501
    update-source loopback0
    address-family l2vpn evpn
      send-community both
    address-family ipv4 mvpn
      send-community both
    address-family ipv6 mvpn
      send-community both
```

”address-family ipv4/v6 mvpn” enables ngMVPN Address-Family for Multicast signalization. ”send community both” ensures both standard and extended communities are exchanged for this address-family. The RT and SOO are extended communities in MVPN routes.

VXLAN EVPN TRM with Anycast RP

Step 3: Enable PIM multicast routing on the distributed anycast gateway SVI interfaces on the VTEPs.

VRF Tenant1

```
interface vlan10
 vrf member Tenant1
 ip address 10.10.10.1/24 tag 12345
 ip pim sparse-mode
 ip pim neighbor-policy NONE*
 fabric forwarding mode anycast-gateway

interface vlan20
 vrf member Tenant1
 ip address 20.20.20.1/24 tag 12345
 ip pim sparse-mode
 ip pim neighbor-policy NONE*
 fabric forwarding mode anycast-gateway

interface vlan30
 vrf member Tenant1
 ip address 30.30.30.1/24 tag 12345
 ip pim sparse-mode
 ip pim neighbor-policy NONE*
 fabric forwarding mode anycast-gateway
```

”ip pim sparse-mode” enables IGMP and PIM on the SVI VLAN used as gateways for the sources and receivers on the VTEPs.

Create a ”ip pim neighbor-policy” to avoid forming PIM neighbor relationship with PIM Routers within the VLAN (Don’t use Distributed Anycast Gateway for PIM Peering).

VXLAN EVPN TRM with Anycast RP

Step 4: Enable PIM multicast routing under the L3 VNI SVI, specify RP address and enable MVPN address family for the tenant VRF.

```
vlan 2501
 vn-segment 50001

interface vlan2501
 vrf member Tenant1
 ip forward
 ip pim sparse-mode

interface loopback250
 vrf member Tenant1
 ip address 10.51.51.254/32 tag 12345
 ip pim sparse-mode

ip multicast overlay-spt-only

vrf context Tenant1
 ip pim rp-address 10.51.51.254
 vni 50001
 rd auto
 address-family ipv4 unicast
  route-target both auto
  route-target both auto evpn
  route-target both auto mvpn
```

”ip pim sparse-mode” enables Multicast Routing on the Tenant.

”ip address 10.51.51.254” defines the Overlay Multicast Rendezvous-Point (RP) IP address in the respective VRF. This IP address has to be advertised in the BGP EVPN control-plane of the VRF (i.e. redistribute).

”ip multicast overlay-spt-only” is needed for defining the distributed RP.

”ip pim rp-address” defines the Overlay Multicast Rendezvous-Point (RP) in the VRF

”route-target both auto mvpn” defines the BGP Route-Target that is added as an Extended Community attribute to the Customer Multicast (C-Multicast) routes (ngMVPN Route-Type 6 and 7).Auto option in generating Route-Targets (RT) constructs RTs using the 2-byte Autonomous System Number and Layer-3 VNI (ASN:VNI).

VXLAN EVPN TRM with Anycast RP

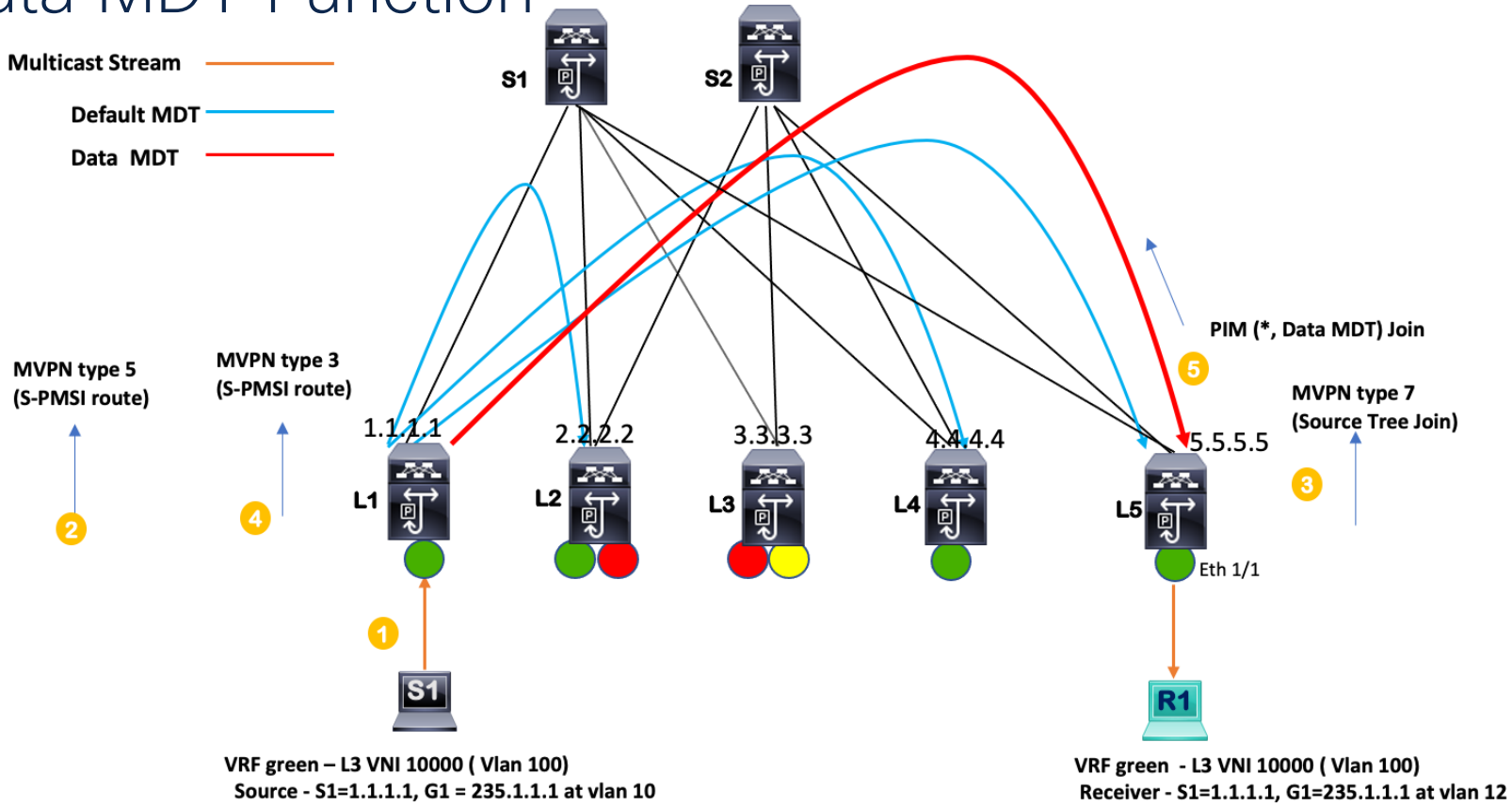
Step 5: Associate the VRF/L3VNI with the default MDT multicast group under the NVE interface.


```
interface nve1
 source-interface loopback1
 host-reachability protocol bgp
 member vni 30010
  mcast-group 239.1.1.1
 member vni 30020
  mcast-group 239.1.1.1
 member vni 30030
  mcast-group 239.1.1.2
 member vni 50001 associate-vrf
  mcast-group 239.10.1.1
```

”mcast-group” maps a multicast group to the L3VNI. The NVE becomes a multicast routed port for multicast traffic in the VRF and the NVE interface on the VTEP becomes the source/receiver for the multicast group in the default MDT tree.

Data MDT

Data MDT Function



Data MDT Configuration

```
vrf context tenant1

address-family ipv4 unicast
 route-target both auto mvpn
 route-target both auto evpn
 [no] mdt data vxlan <group-range-1> [threshold <value>] [route-map <policy-name_1>] [seq <sequence-number>]'

address-family ipv6 unicast
 route-target both auto mvpn
 route-target both auto evpn
 [no] mdt data vxlan <group-range-1> [threshold <value>] [route-map <policy-name_1>] [seq <sequence-number>]'
```

Data MDT Guidelines

- Switchover to Data MDT can be immediate or based on the traffic bandwidth (threshold-based configuration).
- ASM and SSM group ranges are supported for Data MDT.
- Data MDT supports IPv4 and IPv6 overlay multicast traffic.
- Data MDT config per L3 VRF.
- Ensure that the total number of underlay groups (L2 BUM, default MDT, and data MDT groups) is 512.

Summary

Key Takeaways

- VXLAN EVPN TRM uses open standard VXLAN data plane with MP-BGP NGMVPN control plane for tenant multicast routing.
- A single MP-BGP control plane protocol is used for both unicast (AF EVPN) and multicasting (AF MVPN) routing in tenants in a VXLAN BGP EVPN Fabric.
- VXLAN EVPN TRM forwards using an ”Always Route” approach.
- VXLAN EVPN TRM supports various RP deployments models including Anycast RP, External RP and RP Anywhere allowing redundancy and ease of migration of RPs.
- IGMP maintains its current function as Host Reporting protocol.
- PIM operates in the tenant for tenant multicast domain and underlay for Default MDT for the tenant.
- Data MDT optimizes forwarding only to remote VTEPs with attached receivers.

Complete your Session Survey

- Please complete your session survey after each session. Your feedback is important.
- Complete a minimum of 4 session surveys and the Overall Conference survey (open from Thursday) to receive your Cisco Live t-shirt.

- All surveys can be taken in the Cisco Events Mobile App or by logging in to the Session Catalog and clicking the "Attendee Dashboard" at https://www.ciscolive.com/emea/learn/sessions/session-catalog.html

Continue Your Education

Visit the Cisco Showcase for related demos.

Book your one-on-one Meet the Engineer meeting.

Attend any of the related sessions at the DevNet, Capture the Flag, and Walk-in Labs zones.

Visit the On-Demand Library for more sessions at ciscolive.com/on-demand.


cisco
The bridge to possible
Thank you
cisco Live!


CISCO Live!
ALL IN