Unicast Forwarding

Unicast

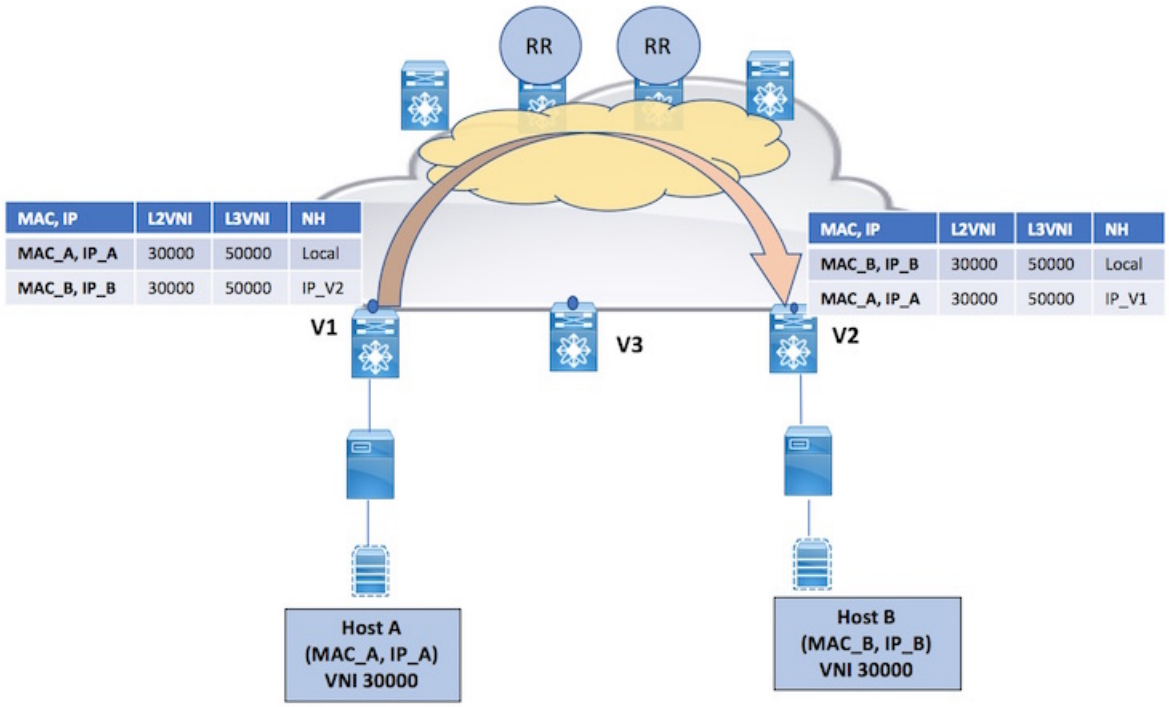
Unicast Forwarding Flows—Overview

Intra and inter subnet forwarding are the possible unicast forwarding flows in the VXLAN BGP EVPN fabric, between leaf/ToR switch VTEPs. They are explained in the subsequent sections.

Intra Subnet Forwarding (Bridging)

Traffic between end hosts within a Layer-2 virtual network is bridged. The reachability information for the 2 end hosts (subnet route, MAC address and IP address) is sent through the MP-BGP EVPN control plane, and the ARP tables in the source and target VTEPs contain this reachability information.

Figure 1: Intra subnet forwarding (Bridging)



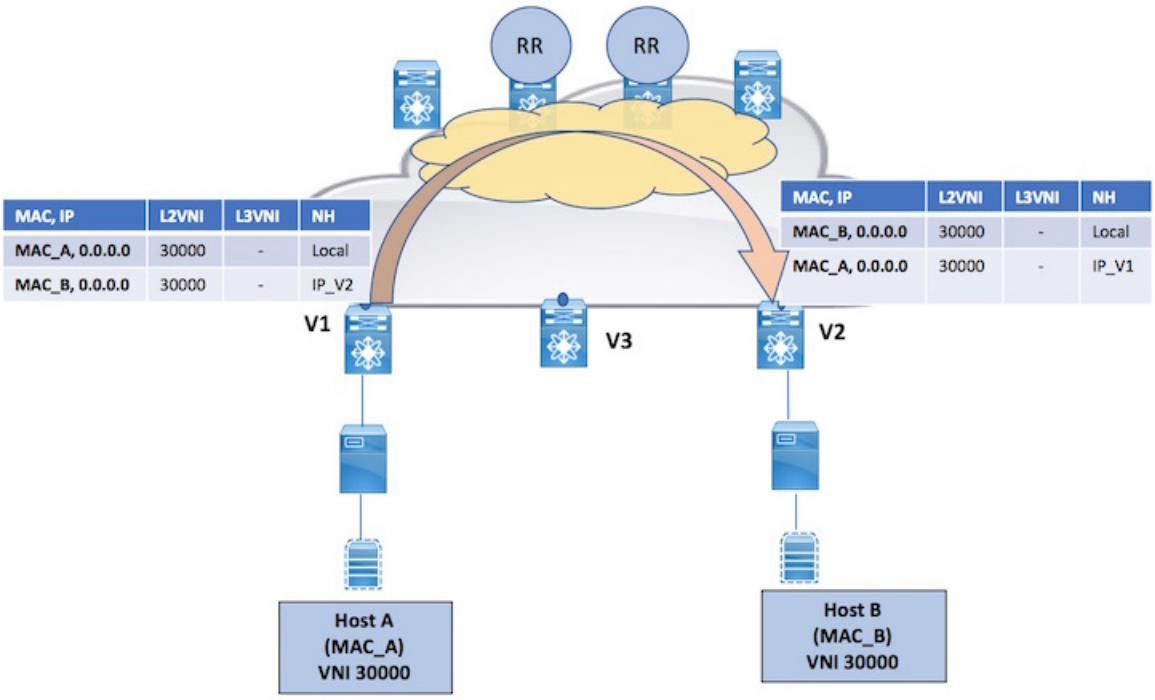
In the above image, Host A on V1 and Host B on V2 are in the same Layer-2 virtual network, represented by VNI 30000. When Host A sends traffic to Host B, the traffic is bridged from V1 to V2, through a spine switch.

Intra Subnet Non-IP Forwarding (Bridging)

This is a use case where end hosts only have a MAC address and no assigned IP address. Some assumptions for non-IP forwarding (Bridging) between 2 known end hosts are given below:

- End hosts are known to the switch VTEPs and MAC address tables are populated.
- Hosts are known to the MP-BGP EVPN control plane and the MAC addresses are populated within the control plane.
- The communication between the end hosts is on the IP layer.

Figure 2: Non-IP forwarding (Bridging)



Non-IP forwarding (Bridging)

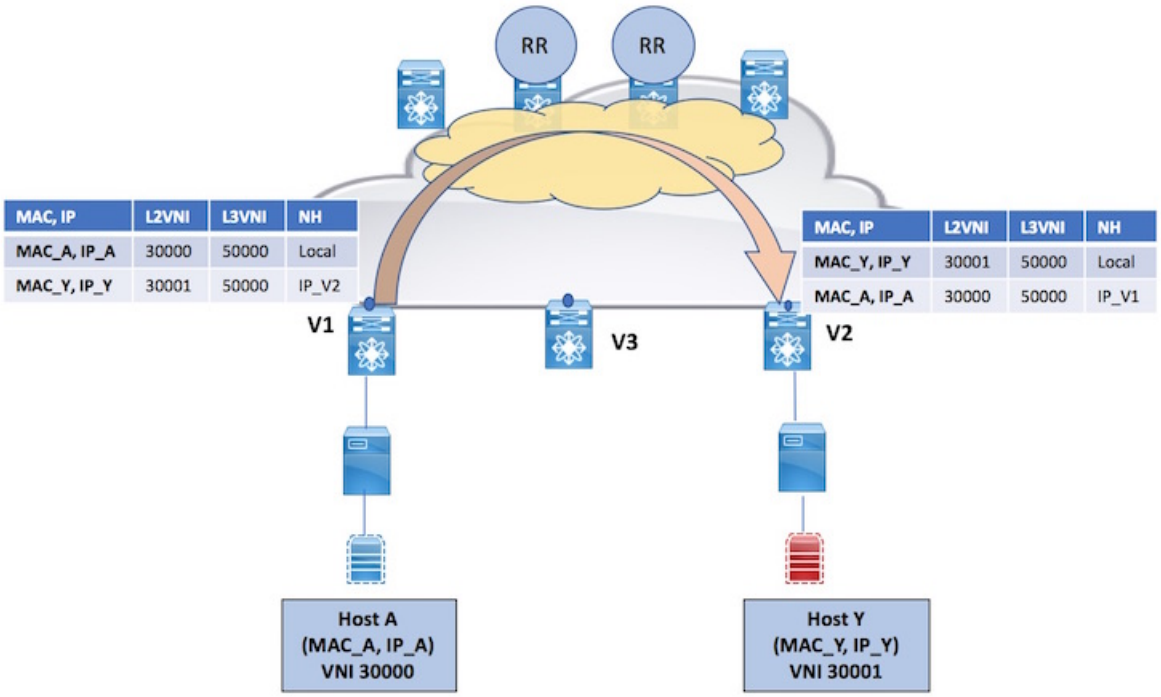
Host A and Host B belong to the same Layer-2 virtual network (with VNI 30000). When Host A sends traffic to Host B, the frame that reaches V1 only contains source and destination MAC addresses. It does not contain IP addresses and Layer-3 VNI information, unlike in the IP bridging scenario where the source and destination end hosts' IP address is updated in the inner frame.

V1 encapsulates this frame (as any incoming frame/packet) in a VXLAN packet and the packet is bridged from V1 to V2, through a spine switch. V2 decapsulates the VXLAN packet and sends the inner frame to Host B.

Inter-Subnet Forwarding (Routing)

Traffic between end hosts of different Layer-2 virtual networks is routed.

Figure 3: Routing across Layer-2 virtual networks



Some notes for the image is given below:

- Host A (attached to V1) belongs to the Layer-2 virtual network with VNI 30000, and Host Y (attached to V2) to Layer-2 VNI 30001. The subnets for the 2 networks host different IP address ranges. So, traffic between them is routed.
- The reachability information for the 2 end hosts i.e. the L3 VNI, MAC and IP addresses along with subnet address of the SVI has been sent through the MP-BGP EVPN control plane. ARP (the ARP suppression feature should be enabled for this) and forwarding tables of the source and target switch VTEPs are populated with the end hosts' reachability information.
- *Host A to V1*—When Host A sends traffic to V1, the Destination MAC (DMAC) of the packet is encapsulated with the MAC address of the (distributed IP anycast) gateway.
- *V1 to V2*—V1 does a look up, notes the VTEP that Host Y is attached to, and checks the VRF (and associated L3 VRF VNI). V1 VXLAN encapsulates the traffic sent by Host A and sends it towards V2.

  The routed traffic from V1 to V2 logically traverses through the Layer-3 VRF VNI 50000. Practically, the traffic traverses through the underlay.

Note When a packet is bridged, the target end host's MAC address is entered in the DMAC field of the inner frame. When a packet is routed, the default gateway's MAC address is entered in the DMAC field of the inner frame. This is because, though VXLAN needs inner SMAC and DMAC fields, for routing purposes, the DMAC should contain the anycast gateway address.

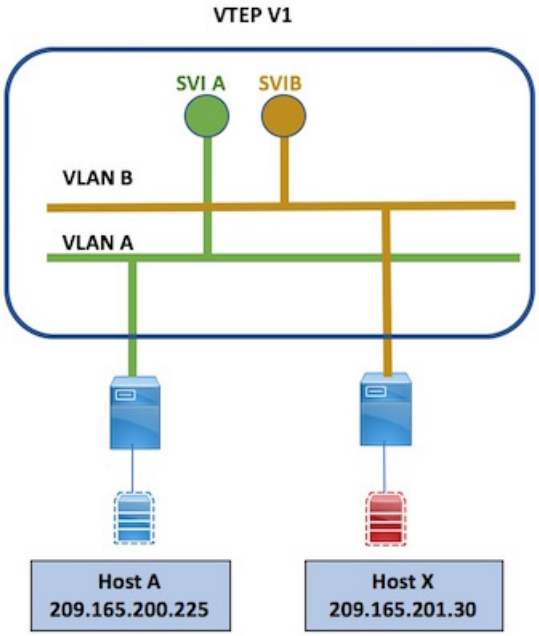
- *V2 to Host Y*—When the packet reaches V2, V2 does a control plane lookup, notes that Host Y's IP address is in VRF A, and then does a MAC table lookup for Host Y. After identifying the port to which Host Y is attached, the packet is sent to Host Y.

Other routing scenarios

Local Routing

Routing between servers belonging to different Layer-2 virtual networks, but attached to the same leaf/ToR switch (VTEP).

Figure 4: Routing across Layer-2 virtual networks on a VTEP



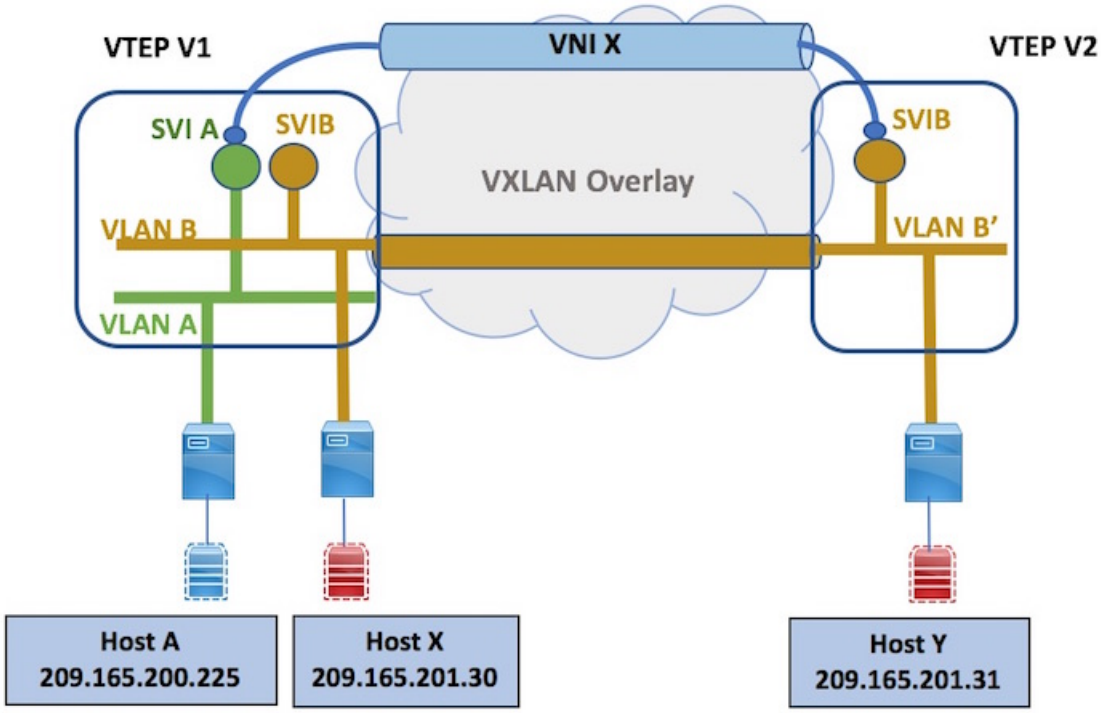
Some notes for the image is given below:

- Host A and Host X belong to different Layer-2 virtual networks, but are attached to the same ToR switch VTEP.
- In this case, the packets are routed through their VLANs (and SVIs). If Host A sends traffic to Host X, the switch does a control plane lookup, identifies that the target is local to the switch, and routes the traffic to Host X.
- The local routing scenario is normal inter VLAN routing. Neither the L2 VNI nor L3 VNI is used for local routing.

Routing to an unknown end host (scenario 1)

Routing to an unknown destination end host (or silent host), an end host that has not ARPed. However, the end host's subnet is known, and has a presence in the source ToR switch VTEP.

Figure 5: Routing to an unknown host (scenario 1)



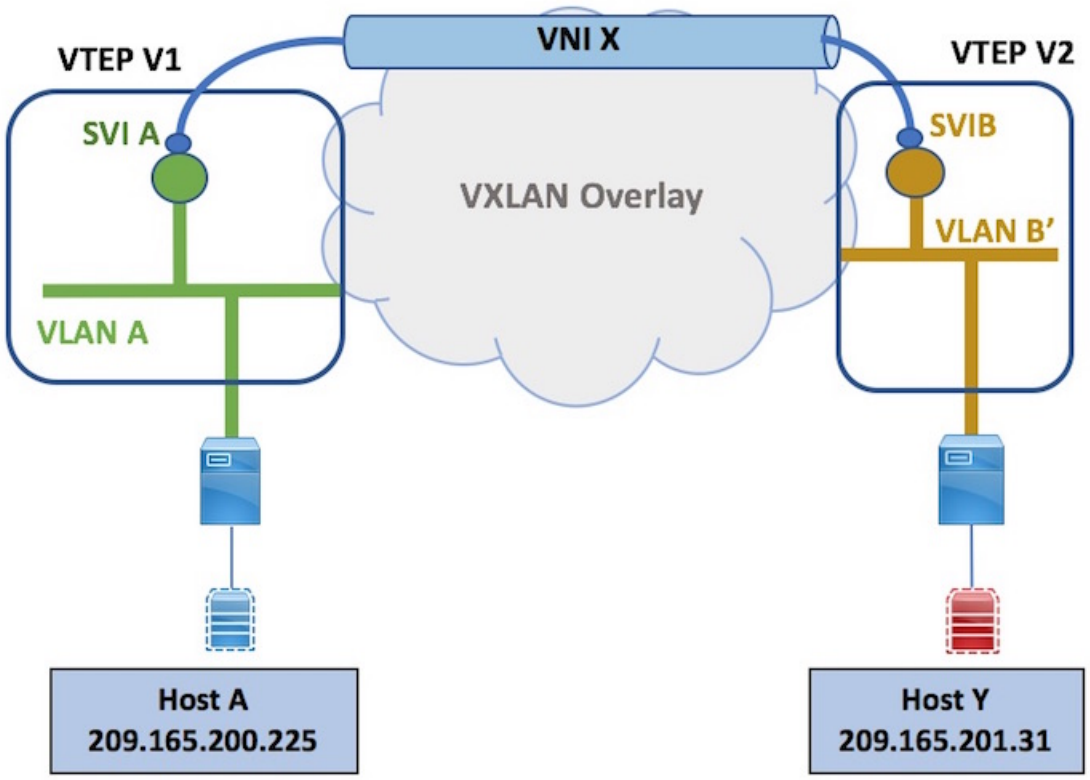
Some notes for the image is given below:

- Host A (on V1) sends traffic to Host Y (on V2). Since the 2 hosts are in different Layer-2 virtual networks, the traffic is routed. Though Host A (and V1) is aware of Host Y's IP address, V1 does not know where Host Y is located.
- V1, however, knows the subnet to which Host Y belongs, and tries to reach Host Y with this information. Since V1 has Host X attached to it, and Host X has the same subnet as that of Host Y, local routing is done within the VTEP to Host X. This is the reason for advertising local subnets on a ToR switch VTEP.
- V1 sends an ARP request for Host Y's IP address. This is only sent to those switch VTEPs that have end hosts for the L2VNI (including the VTEP that has Host Y). The ARP request's source is the anycast gateway MAC and IP address.
- In response to the ARP request, since Host Y is attached to V2, Host Y's response is sent to V2, and the response is populated in the control plane, thereby ensuring that reachability information for Host Y is made available.
- Subsequently, when Host A communicates to Host Y, the traffic is sent from V1 through the Layer-3 VNI (VNI X) to V2, and to Host Y.

Routing to an unknown end host (scenario 2)

Routing to an unknown destination. However, the end host's subnet is known, but the subnet does not have any presence in the source ToR switch VTEP.

Figure 6: Routing to an unknown host (scenario 2)



Some notes for the image is given below:

- Host A (on V1) sends traffic to Host Y (on V2). Though Host A (and V1) is aware of Host Y's IP address, V1 does not know where Host Y is located. Also, V1 does not have any end host that belongs to the same Layer-2 virtual network as that of Host Y.
- V1 knows Host Y's IP address and the subnet to which it belongs (which V1 has learnt previously through the control plane) and sends an ARP request towards the L2VNI on one of the VTEPs advertising the subnet. The ARP request's response ensures Host Y's reachability information is sent in the control plane. The direct route of the SVI is redistributed into the VXLAN BGP EVPN control plane, thereby ensuring that V1 knows where Host Y resides.
- When it is found out that Host Y's subnet is behind VTEP2, Host Y's reachability information is updated on V1. Now, Host Y is reachable across the fabric.

IPv6 address handling in the VXLAN BGP EVPN fabric

Some pointers are given below:

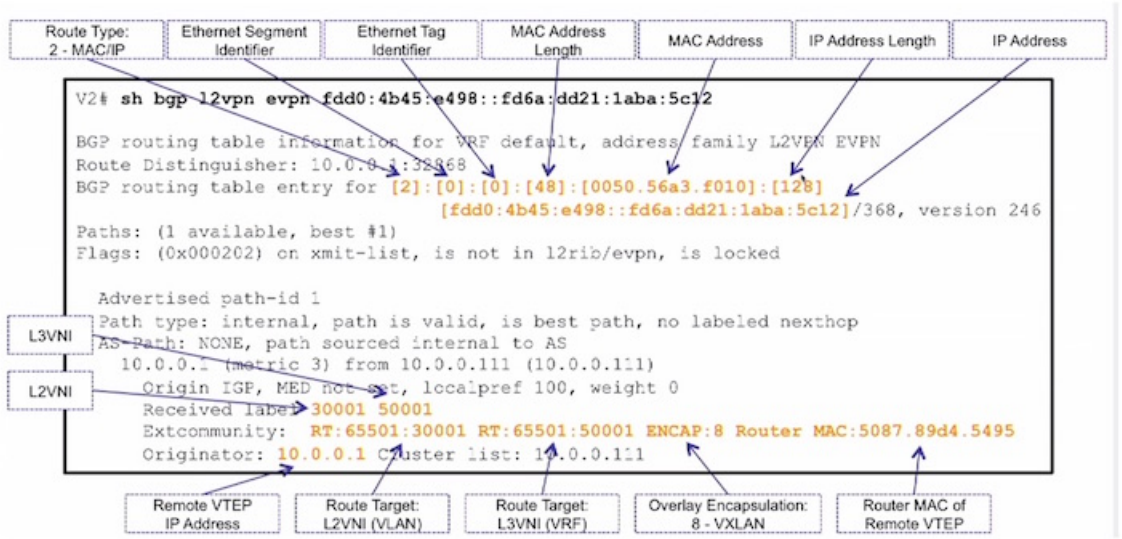
- IPv6 addresses are supported in the fabric overlay, but not in the fabric underlay. So, end hosts can have IPv6 addresses, but the underlay can only contain IPv4 addresses.

- The semantics for IPv4 and IPv6 addressing in the overlay are conceptually the same, with differences that ARP and Neighbor Discovery (ND) have.
- Address resolution - ARP (IPv4) and ND (IPv6). IPv4 addresses consume less system resources on switching hardware tables compared to IPv6 addresses due to the length of the addresses.
- Layer-3 (IPv6) information is learned based on ND snooping.
- Subnet routing is also supported for IPv6 prefixes/subnets.

IPv6 Route Types

IPv6 representations of EVPN Route Type 2 (MAC/IP advertisement routes) and Route Type 5 (IP prefix routes) are given below.

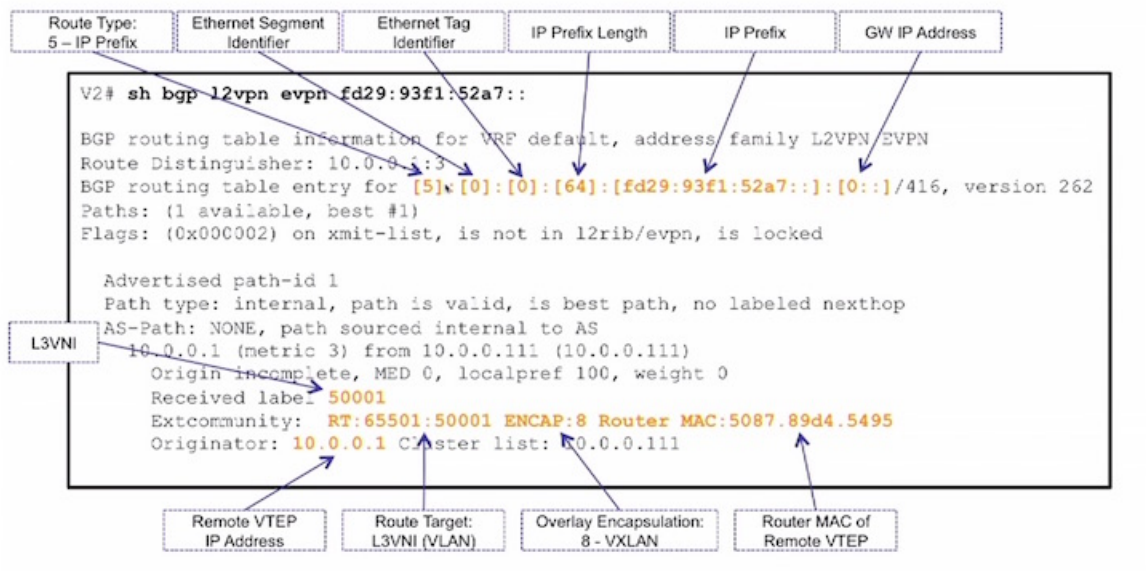
Figure 7: IPv6 host route—Route Type 2



Some notes for the image are given below:

- The fields in the image are self-explanatory.
- The IP Address Length field displays 128, indicating it's an IPv6 address. The IP Address field displays the IPv6 address. The IPv6 address is being distributed in the control plane.

Figure 8: IPv6 prefix route—Route Type 5



Some notes for the image are given below:

- The fields in the image are self-explanatory.
- The IP Prefix Length, IP Prefix, and GW IP Address fields denote IPv6 addressing semantics.