Integrated Routing and Bridging

This chapter describes the configuration of Integrated Routing and Bridging (IRB). IRB provides the ability to exchange traffic between bridging services and a routed interface using a Bridge-Group Virtual Interface (BVI).

Understanding IRB

IRB provides Layer 2 bridging service between hosts that are within a Layer 2 domain. Also, it provides routing service for hosts that are in different subnets within a Layer 3 VPN.

Bridge-Group Virtual Interface

The BVI is a virtual interface within the router that acts like a normal routed interface. A BVI is associated with a single bridge domain and represents the link between the bridging and the routing domains on the router. To support receipt of packets from a bridged interface that are destined to a routed interface, the BVI must be configured with the appropriate IP addresses and relevant Layer 3 attributes. The BVI does not support bridging itself, but acts as a gateway for the corresponding bridge-domain to a routed interface within the router.

BVI supports these attributes, and has the following characteristics:

- Uses a MAC address taken from the local chassis MAC address pool, unless overridden at the BVI interface.
- Is configured as an interface type using the **interface bvi** command and uses an IPv4 or IPv6 address that is in the same subnet as the hosts on the segments of the bridged domain.
- The BVI identifier is independent of the bridge-domain identifier.
- BVI interfaces support a number range of 1 to 4294967295.

BVI Interface and Line Protocol States

Like typical interface states on the router, a BVI has both an Interface and Line Protocol state.

The BVI interface state is Up when the following occurs:

- The BVI interface is created.
- The bridge-domain that is configured with the **routed interface bvi** command has at least one available active bridge port (Attachment circuit [AC]).

Note A BVI will be moved to the Down state if all of the bridge ports (Ethernet flow points [EFPs]) associated with the bridge domain for that BVI are down. However, the BVI will remain up if at least one bridgeport is up, even if all EFPs are down.

These characteristics determine when the the BVI line protocol state is up:

- The bridge-domain is in Up state.
- The BVI IP address is not in conflict with any other IP address on another active interface in the router.

Prerequisites for Configuring IRB

Before configuring IRB, ensure that these tasks and conditions are met:

- Know the IP addressing and other Layer 3 information to be configured on the bridge virtual interface (BVI).
- Complete MAC address planning if you decide to override the common global MAC address for all BVIs.

  You can replace the preferred MAC address for the BVI interface with the default MAC address allocated from the chassis pool. The MAC address is divided into:

  - 32 bits most significant bits called MAC prefix.

    The router has a limitation of four different MAC prefixes per system. You must not use more than four different MAC prefixes when choosing the MAC address for BVI and other L3 interfaces.

  - 16 bits least significant called MAC host. You can choose any value for the MAC host.

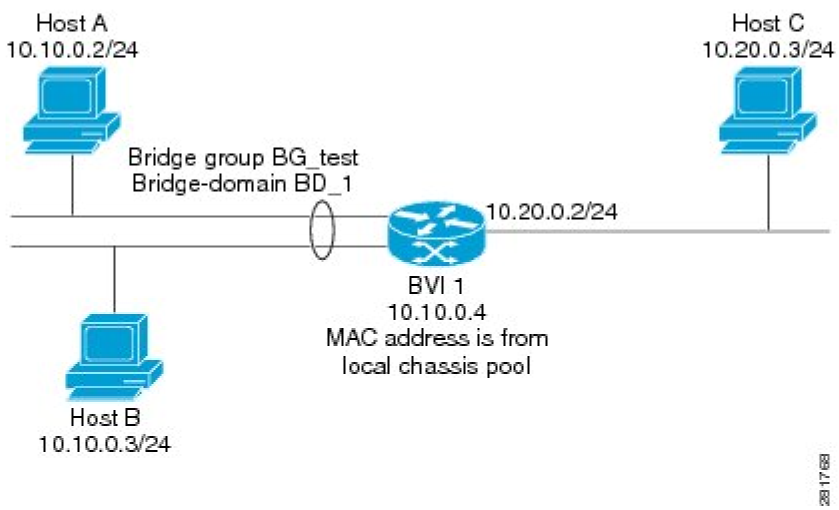
- Be sure that the BVI network address is being advertised by running static or dynamic routing on the BVI interface.

Packet Flows Using IRB

This figure shows a simplified functional diagram of an IRB implementation to describe different packet flows between Host A, B, and C. In this example, Host C is on a network with a connection to the same router. In reality, another router could be between Host C and the router shown.

Figure 1:

Figure 2: IRB Packet Flows Between Hosts



When IRB is configured on a router, the following processing happens:

- ARP requests are resolved between the hosts and BVI that are part of the bridge domain.
- All packets from a host on a bridged interface go to the BVI if the destination MAC address matches the BVI MAC address. Otherwise, the packets are bridged.
- For packets destined for a host on a routed network, the BVI forwards the packets to the routing engine before sending them out a routed interface.
- For packets that are destined for a host on a segment in the bridge domain that come in to the router on a routed interface, the BVI forwards the packet to the bridging engine, which forwards it through the appropriate bridged interface.

Packet Flows When Host A Sends to Host B on the Bridge Domain

When Host A sends data to Host B in the bridge domain on the 10.10.0.0 network, no routing occurs. The hosts are on the same subnet and the packets are bridged between their segment interfaces on the router.

Packet Flows When Host A Sends to Host C From the Bridge Domain to a Routed Interface

Using host information from this figure, the following occurs when Host A sends data to Host C from the IRB bridging domain to the routing domain:

- Host A sends the packet to the BVI (as long any ARP request the is resolved between the host and the BVI). The packet has the following information:
  - Source MAC address of host A.
  - Destination MAC address of the BVI.
- Since Host C is on another network and needs to be routed, the BVI forwards the packet to the routed interface with the following information:
  - IP source MAC address of Host A (10.10.0.2) is changed to the MAC address of the BVI (10.10.0.4).

- IP destination address is the IP address of Host C (10.20.0.3).

- Interface 10.20.0.2 sees receipt of a packet from the routed BVI 10.10.0.4. The packet is then routed through interface 10.20.0.2 to Host C.

Packet Flows When Host C Sends to Host B From a Routed Interface to the Bridge Domain

Using host information from this figure, the following occurs when Host C sends data to Host B from the IRB routing domain to the bridging domain:

- The packet comes into the routing domain with the following information:
  - MAC source address—MAC of Host C.
  - MAC destination address—MAC of the 10.20.0.2 ingress interface.
  - IP source address—IP address of Host C (10.20.0.3).
  - IP destination address—IP address of Host B (10.10.0.3).
- When interface 10.20.0.2 receives the packet, it looks in the routing table and determines that the packet needs to be forwarded to the BVI at 10.10.0.4.
- The routing engine captures the packet that is destined for the BVI and forwards it to the BVI's corresponding bridge domain. The packet is then bridged through the appropriate interface if the destination MAC address for Host B appears in the bridging table, or is flooded on all interfaces in the bridge group if the address is not in the bridging table.

Configure IRB

Follow these steps to configure an IRB:

- Configure the Bridge Group Virtual Interface
- (Optional) Configure the static MAC address on the BVI interface
- Configure the Layer 2 AC Interfaces
- Configure a Bridge Group and Assigning Interfaces to a Bridge Domain
- Associate the BVI as the Routed Interface on a Bridge Domain

Configuration Example

```
/* Configure the BVI and its IPv4 address * /
Router# configure
Router(config)#interface bvi 1
Router(config-if)#ipv4 address 10.10.0.4 255.255.255.0
Router(config-if)#ipv6 address 2001:100:1:1::1/96

/* optionally, you can configure the static MAC address */
Router(config-if))# mac-address 2001.100.2
Router(config-if))# exit
!
```

```
/* Configure the Layer 2 AC interface */
Router(config)# interface HundredGigE 0/0/0/1 l2transport
Router(config-if-l2))# exit
Router(config-if))# exit
!

/* Configure the L2VPN bridge group and bridge domain and assign interfaces */
Router(config)# l2vpn
Router(config-l2vpn)# bridge group 10
Router(config-l2vpn-bg)# bridge-domain 1
Router(config-l2vpn-bg-bd)# interface HundredGigE 0/0/0/1
Router(config-l2vpn-bg-bd-ac)# exit
!
/* Associate a BVI to the bridge domain */
Router(config-l2vpn-bg-bd)# routed interface bvi 1
Router(config-l2vpn-bg-bd-bvi)# commit


/* IRB configuration for tagged bridge ports (sub-interfaces) in a bridge domain with BVI
*/
Router# configure
Router(config)# interface HundredGigE 0/0/0/2.1 l2transport
Router(config-subif)# encapsulation dot1q 102
Router(config-subif)# rewrite ingress tag pop 1 symmetric
Router(config-subif)# exit
Router(config)# interface bvi 2
Router(config-if)# ipv4 address 56.78.100.1 255.255.255.0
Router(config-if)# ipv6 address 56:78:100::1/64
Router(config-if))# mac-address 2002.100.1
Router(config-if))# exit
Router(config)# l2vpn
Router(config-l2vpn)# bridge group 10
Router(config-l2vpn-bg)# bridge-domain 2
Router(config-l2vpn-bg-bd)# interface HundredGigE 0/0/0/2.1
Router(config-l2vpn-bg-bd-ac)# exit
Router(config-l2vpn-bg-bd)# routed interface bvi 2
Router(config-l2vpn-bg-bd-bvi)# commit
```

Note Double VLAN tagged sub-interface is not supported for IRB service.

Verification

Verify the interface status, line protocol state, and packet counters for the specified BVI:

```
Router# show interfaces bvi 1  brief
BVI1 is up, line protocol is up
Interface state transitions: 701
Hardware is Bridge-Group Virtual Interface, address is 2001.0100.0001
Internet address is 10.10.0.4/24
MTU 1514 bytes, BW 10000000 Kbit (Max: 10000000 Kbit)
reliability 255/255, txload 0/255, rxload 1/255
Encapsulation ARPA, loopback not set,
Last link flapped 2d06h
ARP type ARPA, ARP timeout 04:00:00
Last input 00:00:00, output 00:00:13
Last clearing of "show interface" counters 3d18h
30 second input rate 43721000 bits/sec, 49684 packets/sec
30 second output rate 0 bits/sec, 0 packets/sec
15428019162 packets input, 1697081244790 bytes, 0 total input drops
0 drops for unrecognized upper-level protocol
```

```
Received 0 broadcast packets, 0 multicast packets
6084259298 packets output, 669870073726 bytes, 0 total output drops
Output 0 broadcast packets, 0 multicast packets
```

EVPN IRB with distributed anycast gateway

EVPN Integrated Routing and Bridging (IRB) provides seamless Layer 3 forwarding across different IP subnets while maintaining the multi-homing capabilities of EVPN, allowing for efficient communication between EVPN hosts and IP VPNs.

Table 1: Feature History Table

Feature name	Release information	Feature description
EVPN IRB with distributed anycast gateway	Release 25.1.1	Introduced in this release on: Modular Systems (8800 [LC ASIC: P100])(select variants only*) EVPN Integrated Routing and Bridging (IRB) facilitates efficient Layer 3 communication across subnets, leveraging PE routers for connectivity over MPLS or IP networks. It supports single and multi-homing, processes packets using VRF table lookups, and enables seamless EVPN to IP VPN communication without route stitching or re-origination. A distributed anycast gateway enhances routing by sharing IP/MAC addresses for load balancing and redundancy, ensuring optimal performance and reduced latency. * This feature is supported on: • 88-LC1-12TH24FH-E • 88-LC1-52Y8H-EM

EVPN IRB feature enables a Layer 2 VPN and an Layer 3 VPN overlay that allows end hosts across the overlay to communicate with each other within the same subnet and across different subnets within the VPN.

These functionalities are part of EVPN IRB with distributed anycast gateway feature:

- Single-homing
- All-active multi-homing
- Single-active multi-homing
- Port-active multi-homing
- MAC/IP advertisement route
- IP prefix route
- MAC aging
- MAC freezing
- Symmetric IRB forwarding
- RT-stitching of subnet route (RT2) into VPNv4 / VPNv6

- Subnet route (RT5) interconnect to VPNv4 / VPNv6

EVPN IRB benefits

The benefit of EVPN IRB is that it allows hosts in an IP subnet to be provisioned anywhere in the data center. When a virtual machine (VM) in a subnet is provisioned behind an EVPN PE, and another VM is required in the same subnet, it can be provisioned behind another EVPN PE. The VMs do not have to be localized; they need not be directly connected or be in the same complex. The VM is allowed to move across the same subnet. Availability of IP MPLS network across all the EVPN PEs enables the provisioning of VM mobility. The EVPN PEs route traffic to each other through MPLS encapsulation.

Distributed anycast L3 gateway

A distributed anycast L3 gateway performs routing on the first hop. Each device shares the same IP and MAC addresses for the gateway, allowing for efficient load balancing and redundancy. This setup ensures that the closest gateway can route traffic, reducing latency and improving performance. This means routing starts on PE nodes as soon as traffic arrives from a CE. If a customer's subnet is stretched to several PE nodes via EVPN bridging, all these PE nodes provide IRB service for the same subnet. There is no active or standby role assignment on the PE nodes; all of them are active L3 gateways.

Anycast address in distributed gateway

For traffic to flow from access to core, there is only one default gateway IP and MAC host address per subnet. The same pair of host addresses is repeated on all PE nodes. Such a pair of IP and MAC addresses is called an anycast address. This IRB topology and configuration is known as a distributed anycast L3 gateway.

EVPN PE connectivity

The EVPN PEs are connected to each other by a spine, so they have IP reachability to each other's loopback interfaces. The IP network and MPLS tunnels existing between these EVPN PEs constitute the IP MPLS underlay fabric.

EVPN control plane

You can configure the MPLS tunnels to tunnel Layer 2 traffic and to overlay VPN on these tunnels. The EVPN control plane distributes both Layer 2 MAC reachability and Layer 3 IP reachability for hosts within the context of the VPN. It overlays a tenant's VPN network on top of the MPLS underlay fabric. Thus, you can have tenant's hosts, which are in the same subnet Layer 2 domain but distributed across the fabric, communicate to each other as if they are in a Layer 2 network.

The Layer 2 VLAN and the corresponding IP subnet are not only a network of physically connected hosts on Layer 2 links but an overlay network on top of underlay IP MPLS fabric, which is spread across the data center.

Routing service in EVPN

A routing service, which enables stretching of the subnet across the fabric, is available. It also provides Layer 3 VPN and performs routing between subnets within the context of the Layer 3 VPN. The EVPN PEs provide Layer 2 bridging service between hosts that are spread across the fabric within a Layer 2 domain that is stretched across the fabric, and Layer 3 VPN service or inter-subnet routing service for hosts in different subnets within Layer 3 VPN.

Symmetric IRB

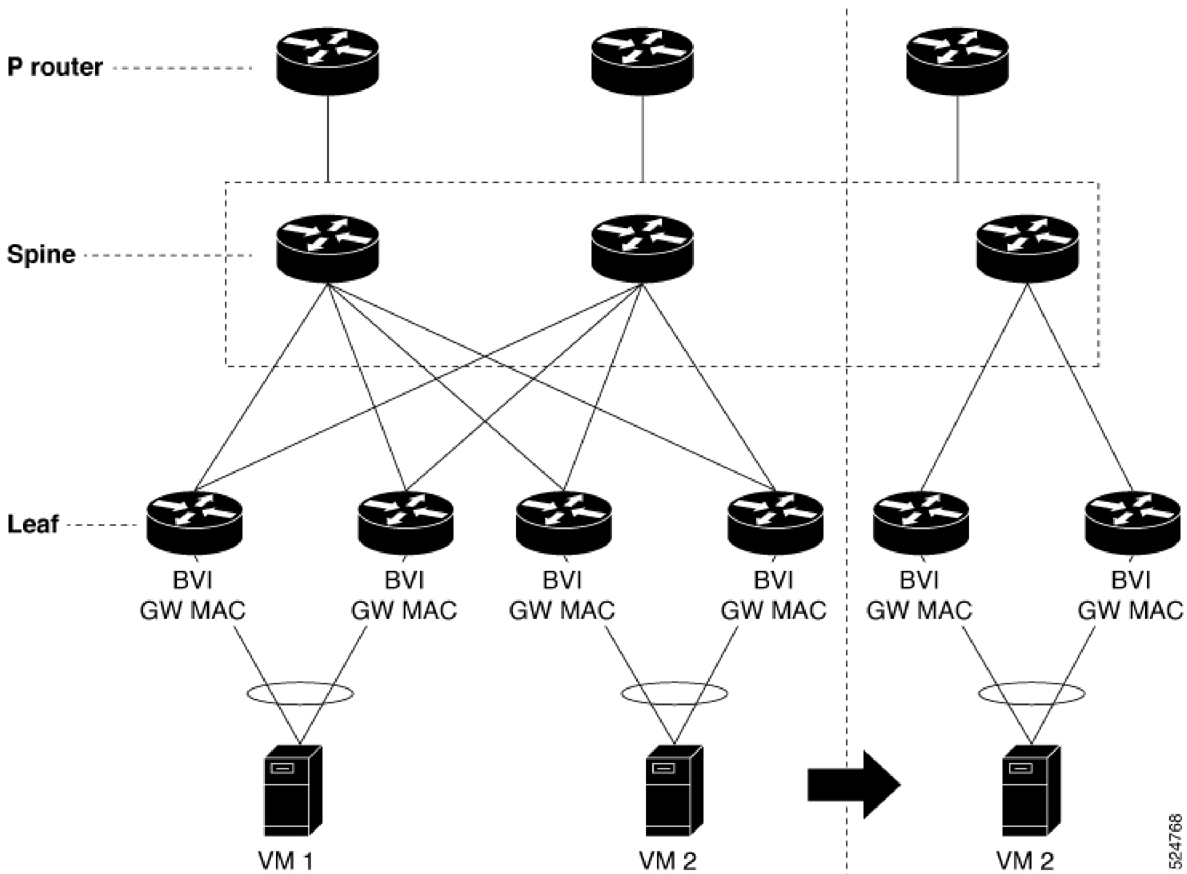
EVPN IRB for a given subnet is configured on all the EVPN PEs that are hosted on that subnet. To facilitate optimal routing while supporting transparent virtual machine mobility, hosts are configured with a single default gateway address for their local subnet. That single (anycast) gateway address is configured with a single (anycast) MAC address on all EVPN PE nodes locally supporting that subnet. This process is repeated for each locally defined subnet requiring anycast gateway support.

The host-to-host Layer 3 traffic, similar to Layer 3 VPN PE-PE forwarding, is routed on the source EVPN PE to the destination EVPN PE next-hop over an IP or MPLS tunnel, where it is routed again to the directly connected host. Such forwarding is also known as Symmetric IRB because the Layer 3 flows are routed at both the source and destination EVPN PEs.

Topology

Let us understand EVPN IRB distributed anycast gateway using a topology diagram.

Figure 3: EVPN IRB distributed anycast gateway



VM connectivity and mobility

In the above topology diagram, the two VMs are in the same subnet, but they are not connected directly through each other via a Layer 2 link. The Layer 2 link is replaced by MPLS tunnels that are connecting them. The whole fabric acts as a single switch and bridges traffic from one VM to the other, enabling VM mobility.

Custom MAC address

In the above topology diagram, the VMs, VM1 and VM2, are connected to each other. When VM2 migrates to a different switch and different server, the VM's current MAC address and IP address are retained. When the subnet is stretched between two EVPN PEs, the same IRB configuration is applied on both devices.

MAC and IP unicast control plane

This use case has following types:

Prefix routing or no subnet stretch

IP reachability across the fabric is established using subnet prefix routes that are advertised using EVPN Route Type 5 with the VPN label and VRF RTs. Host ARP and MAC sync are established across multi-homing EVPN PEs using MAC+IP Route Type 2 based on a shared ESI to support BVI all-active multi-homing load balance on both the multi-homing EVPN PEs.

Host routing or stretched subnet

When a host is discovered through ARP, the MAC and IP Route Type 2 is advertised with both MAC VRF and IP VRF router targets, and with VPN labels for both MAC-VRF and IP-VRF. Particularly, the VRF route targets and Layer 3 VPN label are associated with Route Type 2 to achieve PE-PE IP routing identical to traditional L3VPNs. A remote EVPN PE installs IP/32 entries directly in Layer 3 VRF table through the advertising EVPN PE next-hop with the Layer 3 VPN label encapsulation, much like a Layer 3 VPN imposition PE. This approach avoids the need to install separate adjacency rewrites for each remote host in a stretched subnet. Instead, it inherits a key Layer 3 VPN scale benefit of being able to share a common forwarding rewrite or load-balance resource across all IP host entries reachable through a set of EVPN PEs.

ARP and MAC sync

For hosts that are connected through LAG to more than one EVPN PE, the local host ARP and MAC entries are learnt in data plane on either one or both of the multihoming EVPN PEs. Local ARP and MAC entries are synced across the two multihoming EVPN PEs using MAC and IP Route Type 2 based on a shared ESI to enable local switching through both the multihoming EVPN PEs. Essentially, a MAC and IP Route Type 2 that is received with a local ESI causes the installation of a synced MAC entry that points to the local AC port, and a synced ARP entry that is installed on the local BVI interface.

Note Only one Ethernet Flow Point (EFP) is supported per non-Zero ESI per bridge domain or EVI. This is a limitation of EVPN.

MAC and IP route re-origination

MAC and IP Route Type 2 received with a local ESI, which is used to sync MAC and ARP entries, is also re-originated from the router that installs a SYNC entry, if the host is not locally learnt and advertised based on local learning. This route re-origination is required to establish overlay IP ECMP paths on remote EVPN PEs, and to minimize traffic hit on local AC link failures, that can result in MAC and IP route withdraw in the overlay.

Intra-subnet unicast data plane

The Layer 2 traffic is bridged on the source EVPN PE using ECMP paths to remote EVPN PEs, established through MAC+IP RT2, for every ES and for every EVI, ES and EAD Route Type 2 routes that are advertised from the local EVPN PEs.

Inter-subnet unicast data plane

Inter-subnet traffic is routed on the source EVPN PEs through overlay ECMP to the destination EVPN PEs next-hops. Data packet are encapsulated with the VPN label advertised from the EVPN PE and tunnel label for the BGP next-hop towards the spine. It is then routed again on the destination EVPN PE using a local ARP adjacency towards the host. IP ECMP on the remote EVPN PEs is established through local and re-originated routes advertised from the local EVPN PEs.

EVPN IRB route types

In EVPN IRB, different route types are used to handle various aspects of routing and bridging. Here are the primary EVPN route types related to IRB.

Route type 2: MAC/IP advertisement route

Route Type 2 (RT2) advertises the host IP and MAC addresses to the peers within Network Layer Reachability Information (NLRI). This route type enables the distribution of Layer 2 and Layer 3 reachability information. The control plane learning of MAC addresses reduces unknown unicast flooding.

A MAC/IP Advertisement Route type specific EVPN NLRI consists of the following fields:

```
+---------------------------------------+
|Route Type (1 octet)                   |*
+---------------------------------------+
|Length (1 octet)                       |
+---------------------------------------+
|RD (8 octets)                          |*
+---------------------------------------+
|Ethernet Segment Identifier (10 octets)|
+---------------------------------------+
|Ethernet Tag ID (4 octets)             |*
+---------------------------------------+
|MAC Address Length (1 octet)           |*
+---------------------------------------+
|MAC Address (6 octets)                 |*
+---------------------------------------+
|IP Address Length (1 octet)            |*
+---------------------------------------+
|IP Address (0, 4, or 16 octets)        |*
+---------------------------------------+
|MPLS Label1 (3 octets)                 |
+---------------------------------------+
|MPLS Label2 (0 or 3 octets)            |
+---------------------------------------+
```

366356

NLRI format: RT 2

`[Type][Len][RD][ESI][ETag][MAC Addr Len][MAC Addr][IP Addr Len][IP Addr][MPLS Label1][MPLS Label2]`

Net attributes: `[Type][RD][ETag][MAC Addr Len][MAC Addr][IP Addr Len][IP Addr]`

Path attributes: `[ESI], [MPLS Label1], [MPLS Label2]`

Example

```
route-policy evpn-policy
  if rd in (10.0.0.2:0) [and/or evpn-route-type is 2] [and/or esi in
(0000.0000.0000.0000.0000)] [and/or etag is 0] [and/or macaddress in (0013.aabb.ccdd)]
[and/or destination in (1.2.3.4/32)] then
    set ..
  endif
end-policy
```

Route type 5: IP prefix route

Route Type 5 (RT5) advertises the IP prefixes in the EVPN domain. This route type is crucial for integrating Layer 3 routing with the EVPN infrastructure. An IP prefix route type specific EVPN NLRI consists of the following fields:

```
+---------------------------------------+
|Route Type (1 octet)                   |*
+---------------------------------------+
|Length (1 octet)                       |
+---------------------------------------+
|RD   (8 octets)                        |*
+---------------------------------------+
|Ethernet Segment Identifier (10 octets)|
+---------------------------------------+
|Ethernet Tag ID (4 octets)             |*
+---------------------------------------+
|IP Address Length (1 octet)            |*
+---------------------------------------+
|IP Address (4 or 16 octets)            |*
+---------------------------------------+
|GW IP Address (4 or 16 octets)         |
+---------------------------------------+
|MPLS Label (3 octets)                  |
+---------------------------------------+
```

NLRI format: RT 5

`[Type][Len][RD][ESI][ETag][IP Addr Len][IP Addr][GW IP Addr][Label]`

Net attributes: `[Type][RD][ETag][IP Addr Len][IP Addr]`

Path attributes: `[ESI], [GW IP Addr], [Label]`

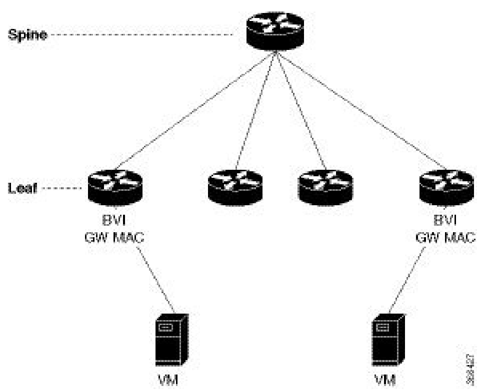
Example

```
route-policy evpn-policy
  if rd in (30.30.30.30:1) [and/or evpn-route-type is 5] [and/or esi in
(0000.0000.0000.0000.0000)] [and/or etag is 0] [and/or destination in (12.2.0.0/16)] [and/or
 evpn-gateway in (0.0.0.0)] then
    set ..
  endif
end-policy
```

EVPN single-homing

The EVPN provider edge (PE) devices learn the MAC address and IP address from the ARP traffic that they receive from the customer edge (CE) devices. The PEs create the MAC+IP routes. The PEs advertise the MAC+IP routes to MPLS core. They inject the host IP routes to IP-VPN gateway. Subnet routes are also advertised from the access EVPN PEs in addition to host routes. All the PE nodes add the host routes in the IP-VRF table. The EVPN PE nodes add MAC route to the MAC-VRF table. The IP-VPN PE advertise the subnet routes to the provider edge devices which add the subnet routes to IP-VRF table. On the PE devices, IRB gateway IP addresses and MAC addresses are not advertised through BGP. IRB gateway IP addresses or MAC addresses are used to send ARP requests towards the datacenter CEs.

Figure 4: EVPN single-homing access gateway



The above topology depicts how EVPN single-homing access gateway enables network connectivity by allowing a CE device to connect to one PE device. The PE device is attached to the Ethernet Segment through bundle or physical interfaces. Null Ethernet Segment Identifier (ESI) is used for single-homing.

EVPN multi-homing all-active

EVPN multi-homing access gateway enables redundant network connectivity by allowing a CE device to connect to more than one PE device. Disruptions to the network connectivity are prevented by allowing a CE device to be connected to a PE device or several PE devices through multi-homing. Ethernet segment is the bunch of Ethernet links through which a CE device is connected to more than one PE devices. The Multi-chassis Link Aggregation Group (MC-LAG) bundle operates as an Ethernet segment.

In EVPN IRB, both EVPN and IP VPN (both VPNv4 and VPNv6) address families are enabled between Data Center Interconnect (DCI) gateways. When Layer 2 (L2) stretch is not available in multiple datacenters (DC), routing is established through VPNv4 or VPNv6 routes. When Layer 2 stretch is available, host routing is applied where IP-MAC routes are learnt by ARP/IPv6 ND and are distributed to EVPN/BGP. In remote peer gateway, these IP-MAC EVPN routes are imported into IP VPN routing table from EVPN route-type 2 routes with secondary label and Layer 3 VRF route-target.

EVPN IRB with all-active multi-homing without subnet stretch or host-routing across the fabric

For those subnets that are local to a set of multi-homing EVPN PEs, EVPN IRB distributed anycast gateway is established through subnet routes that are advertised using EVPN Route Type 5 to VRF-hosting remote leaf nodes. Though there is no need for the /32 routes within the subnet to be advertised, host MAC and ARP entries have to synced across the EVPN PE to which the servers are multi-homed.

Note The subnet stretch feature with EVPN IRB is exclusively available for use within VRF instances and is not applicable to the global VRF.

This type of multi-homing has the following characteristics:

- All-active EV LAG on access
- Layer 3 ECMP for the fabric for dual-homed hosts based on subnet routes
- Absence of layer 2 subnet stretch from remote PE to local EVPN IRB multi-homing PE
- Layer 2 stretch within redundancy group of leaf nodes with orphan ports

Prefix-routing solution for inter-subnet traffic from remote PE to EVPN IRB mutil-homing PE is summarized here.

Across multi-homing EVPN PEs:

- Local ARP cache and MAC addresses are synchronized for dual-homed hosts through EVPN MAC+IP host route advertisements. They are imported as local, and are based on the local ESI match, for optimal forwarding to the access gateway.
- Orphan MAC addresses and host IP addresses are installed as remote addresses over the fabric.
- ES/EAD routes are exchanges for the designated forwarder (DF) election and split-horizon label.

Across remote EVPN PEs:

- Dual-homed MAC+IP EVPN Route Type 2 is exchanged with the ESI, EVI Label, Layer 2-Route Type. It is not imported across the fabric, if there is no subnet stretch or host-routing.
- The subnet IP EVPN Route Type 5 is exchanged with VRF label and Layer 3-Route Type.
- Layer 3 Route Type for the VRFs is imported that are present locally.
- Layer 2 Route Type for locally present BDs is imported. It is only imported from the leaf in the same redundancy group, if BD is not stretched.

EVPN IRB with all-active multi-homing with subnet stretch or host-routing across the fabric

For a bridge domain or subnet that is stretched across remote EVPN PEs, both /32 host routes and MAC routes are distributed in a EVPN overlay control plane to enable Layer 2 and Layer 3 traffic to the end points in a stretched subnet.

This type of multi-homing has the following characteristics:

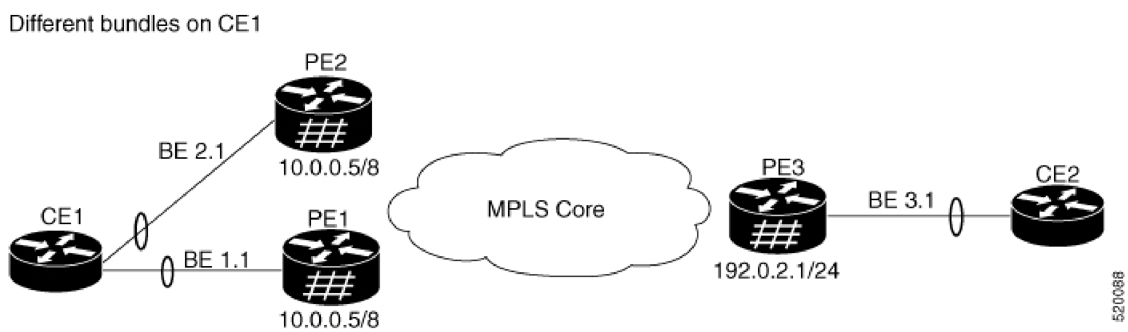
- Layer 2 or Layer 3 ECMP for the fabric for dual-homed hosts based on Route Type 1 and Route Type 2
- Layer 3 unipath over the Fabric for single-homed hosts based on Route Type 2

- Layer 2 subnet stretch over the fabric
- Layer 2 stretch within redundancy group of leaf nodes with orphan ports

EVPN single-active multihoming for anycast gateway IRB

The EVPN single-active multihoming for anycast gateway IRB feature supports single-active redundancy mode. In this mode, the provider edge (PE) nodes locally connected to an Ethernet Segment load balance traffic to and from the Ethernet Segment based on EVPN service instance (EVI). Within an EVPN service instance, only one PE forwards traffic to and from the Ethernet Segment (ES). This feature supports intersubnet scenario only.

Figure 5: EVPN: Single-active multihoming for anycast gateway IRB



Consider a topology where CE1 is multihomed to PE1 and PE2. Bundle Ethernet interfaces BE 1.1, BE 2.1, and the ingress interface must belong to the same switching domain on CE1. Enable host routing and configure anycast gateway IP address on both these peering PEs. PE1 and PE2 are connected to PE3 through MPLS core. PE3 has reachability of subnet 10.0.0.5/8 to both peering PEs. Peering PEs has reachability to PE3 subnet 192.0.2.1/24. CE2 is connected to PE3 through an Ethernet interface bundle. PE1 and PE2 advertise Type four routes, and then performs designated forwarder (DF) election. The non-DF blocks the traffic in both the directions in single-active mode.

Consider traffic flow from CE1 to CE2. CE1 sends an Address Resolution Protocol (ARP) broadcast request to both PE1 and PE2. Peering PEs perform designated forwarder (DF) election for shared EVI. If PE1 is the designated forwarder for the EVI, PE1 replies to the ARP request from CE1. PE2 drops the traffic from CE1. Thereafter, all the unicast traffic is sent through PE1. PE2 is set to standby or blocked state and traffic is not sent over this path. PE1 advertises MAC to PE3. PE3 always sends and receives traffic through PE1. PE3 sends the traffic to CE2 over Ethernet interface bundle. If BE1 fails, PE2 becomes the DF and traffic flows through PE2.

Configure EVPN single-active multihoming

Perform the following tasks on PE1 and PE2 to configure EVPN single-active multihoming.

Procedure

Step 1 Configure EVPN IRB with host routing.

Example:

```
Router# configure
Router(config)# l2vpn
Router(config-l2vpn)# bridge group 6005
Router(config-l2vpn-bg)# bridge-domain 6005
Router(config-l2vpn-bg-bd)# routed interface BVI50
Router(config-l2vpn-bg-bd-bvi)# exit
Router(config-l2vpn-bg-bd-bvi)# interface Bundle-Ether2.1
Router(config-l2vpn-bg-bd-ac)# evi 6005
Router(config-l2vpnbg-bd-evi)# commit
Router(config-l2vpnbg-bd-evi)# exit
Router(config)# interface BVI50
Router(config-if)# host-routing
Router(config-if)# vrf 30
Router(config-if)# ipv4 address 10.0.0.5 255.0.0.0
Router(config-if)# local-proxy-arp
Router(config-if)# mac-address 1.1.1
Router(config-if)# commit
```

This section shows EVPN IRB with host routing running configuration.

```
configure
l2vpn
 bridge group 6005
  bridge-domain 6005
   interface Bundle-Ether2.1
    evi 6005
!
   !
 interface BVI34
 host-routing
 vrf 30
 ipv4 address 10.0.0.5 255.0.0.0
 arp learning local
 local-proxy-arp
 mac-address 1.1.1
```

Step 2 Configure EVPN Ethernet segment.

Example:

```
Router# configure
Router(config)# evpn
Router(config-evpn)# interface Bundle-Ether1
Router(config-evpn-ac)# ethernet-segment
Router(config-evpn-ac-es)# identifier type 0 40.00.00.00.00.00.00.00.01
Router(config-evpn-ac-es)# load-balancing-mode single-active
Router(config-evpn-ac-es)# bgp route-target 4000.0000.0001
Router(config-evpn-ac-es)# commit
```

This section shows EVPN Ethernet segment running configuration.

```
configure
evpn
 interface Bundle-Ether1
   ethernet-segment
    identifier type 0 40.00.00.00.00.00.00.00.01
    load-balancing-mode single-active
    bgp route-target 4000.0000.0001
    !
!
!
```

Step 3 Configure EVPN Service Instance (EVI) parameters.

Example:

```
Router# configure
Router(config)# evpn
Router(config-evpn)# evi 6005
Router(config-evpn-evi)# bgp
Router(config-evpn-evi-bgp)# rd 200:50
Router(config-evpn-evi-bgp)# route-target import 100:6005
Router(config-evpn-evi-bgp)# route-target export 100:6005
Router(config-evpn-evi-bgp)# commit
```

This section shows EVI parameters running configuration.

```
configure
 evpn
  evi 6005
   bgp
    rd 200:50
    route-target import 100:6005
    route-target export 100:6005
!
!
```

Step 4 Configure Layer 2 interface.

Example:

```
Router# configure
Router(config)# interface bundle-ether2.1 l2transport
Router(config-subif-l2)# no shutdown
Router(config-subif-l2)# encapsulation dot1q 1
Router(config-subif-l2)# rewrite ingress tag pop 1 symmetric
Router(config-subif-l2)#commit
Router(config-subif-l2)#exit
```

This section shows the Layer 2 interface running configuration.

```
configure
 interface bundle-ether2.1 l2transport
  no shutdown
  encapsulation dot1q 1
  rewrite ingress tag pop 1 symmetric
!
```

Step 5 Configure a bridge domain on PE1 and PE2.

Example:

```
Router# configure
Router(config)# l2vpn
Router(config-l2vpn)# bridge group 6005
Router(config-l2vpn-bg)# bridge-domain 6005
Router(config-l2vpn-bg-bd)# interface Bundle-Ether2.1
Router(config-l2vpn-bg-bd-ac)# evi 6005
Router(config-l2vpnbg-bd-evi)# commit
Router(config-l2vpnbg-bd-evi)# exit
```

This section shows the bridge domain running configuration.

```
configure
l2vpn
 bridge group 6005
  bridge-domain 6005
   interface Bundle-Ether2.1
    evi 6005
!
```

Step 6 Configure VRF.

Example:

```
Router# configure
Router(config)# vrf 30
Router(config-vrf)# address-family ipv4 unicast
Router(config-l2vpn-vrf-af)# import route-target 100:6005
Router(config-l2vpn-vrf-af)# export route-target 100:6005
Router(config-l2vpn-vrf-af)# commit
```

This section shows the VRF running configuration.

```
configure
 vrf 30
  address-family ipv4 unicast
   route-target import 100:6005
   route-target export 100:6005

!
```

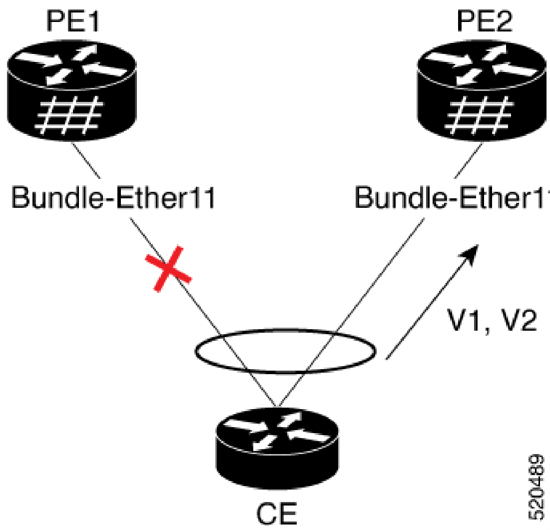
EVPN IRB port-active multihoming

The EVPN IRB port-active multihoming supports single-active redundancy load balancing at the port-level or the interface-level. You can use this functionality when you want to forward the traffic to a specific interface, rather than have a per-flow load balancing across multiple PE routers. The EVPN IRB port-active multihoming provides a faster convergence during a link failure. It enables protocol simplification as only one of the physical ports is active at a given time. You can enable this functionality only on bundle interfaces.

EVPN port-active provides protocol simplification compared to Inter-Chassis Communication Protocol (ICCP), which runs on top of Label Distribution Protocol (LDP). You can use this functionality as an alternative to multi-chassis link aggregation group (MC-LAG) with ICCP. You can also use this functionality when you want certain QoS features to work.

This feature allows one of the PEs to be in active mode and another in the standby mode at the port-level. Only the PE that is in the active mode sends and receives the traffic. The other PE remains in the standby mode. The PEs use the Designated Forwarder (DF) election mechanism to determine which PE must be in the active mode and which must be in the standby mode. You can use either modulo or Highest Random Weight (HRW) algorithm for per port DF election. By default, the modulo algorithm is used for per port DF election.

Figure 6: EVPN IRB port-active multihoming



Consider a topology where the customer edge device (CE) is multihomed to provider edge devices, PE1 and PE2. Use single link aggregation at the CE. Only one of the two interfaces is in the forwarding state, and the other interface is in the standby state. In this topology, PE2 is in the active mode and PE1 is in the standby mode. Hence, PE2 carries traffic from the CE. All services on the PE2 interface operate in the active mode. All services on the PE1 operate in the standby mode.

If the interface is running LACP, then the standby sets the LACP state to Out-of-Service (OOS) instead of bringing the interface state down. This state enables better convergence on standby to active transition.

If you remove the port-active configuration on both PE1 and PE2 and then add back the port-active configuration on both the PEs, PE2 is chosen as an active interface again.

This feature supports both L2 and L3 port-active functionality. L2 and L3 port-active can coexist on the same bundle. For example, if you configure port-active on a bundle, the bundle can have a mix of both L3 subinterfaces and L2 subinterfaces participating in EVPN IRB.

Configure EVPN IRB port-active multihoming

Perform this task to configure EVPN IRB port-active multihoming.

Procedure

Step 1 Configure the same ESI on both the routers. Configure Ethernet-Segment in port-active load-balancing mode on peering PEs for a specific interface.

Example:

```
/* PE1 and PE2 Configuration */

Router#configure
Router(config)#interface Bundle-Ether11
Router(config-if)#lacp system mac 3637.3637.3637
Router(config-if)#exit
```

```
Router(config)#evpn
Router(config-evpn)#interface Bundle-Ether11
Router(config-evpn-ac)#ethernet-segment
Router(config-evpn-ac-es)#identifier type 0 11.11.11.11.11.00.11.11.11
Router(config-evpn-ac-es)#load-balancing-mode port-active
Router(config-evpn-ac-es)#commit

/* If you want enable L3 port-active, configure the IP address */
Router#configure
Router(config)#interface Bundle-Ether11
Router(config-if)#ipv4 address 10.0.0.1 255.0.0.0
Router(config-if)#ipv6 address 10::1/64
Router(config-if)#commit
```

Step 2 Verify the port-active running configuration.

Example:

```
Router:PE2#show running-config
configure
 interface Bundle-Ether11
  lacp system mac 3637.3637.3637
 !

evpn
 interface Bundle-Ether11
  ethernet-segment
   identifier type 0 11.11.11.11.11.00.11.11.11
    load-balancing-mode port-active
   !
  !
interface Bundle-Ether11
 ipv4 address 10.0.0.1 255.0.0.0
 ipv6 address 10::1/64
 !
!
```

Step 3 Verify that you have configured the port-active multihoming feature successfully.

Example:

```
Router:PE2#show bundle bundle-ether 11

Bundle-Ether11
  Status:                                    Up
  Local links <active/standby/configured>:   1 / 0 / 1
  Local bandwidth <effective/available>:     1000000 (1000000) kbps
  MAC address (source):                      02b4.3cb4.a004 (Chassis pool)
  Inter-chassis link:                        No
  Minimum active links / bandwidth:          1 / 1 kbps
  Maximum active links:                      64
  Wait while timer:                          2000 ms
  Load balancing:
    Link order signaling:                    Not configured
    Hash type:                               Default
    Locality threshold:                      None
  LACP:                                      Operational
    Flap suppression timer:                  Off
    Cisco extensions:                        Disabled
    Non-revertive:                           Disabled
  mLACP:                                     Not configured
  IPv4 BFD:                                  Not configured
```

```
  IPv6 BFD:                                  Not configured

  Port                  Device           State        Port ID         B/W, kbps
  --------------------  ---------------  -----------  --------------
  Gi0/2/0/8             Local            Active       0x8000, 0x0006     1000000
      Link is Active

/* PE2 is in the active mode, hence the status shows as Up and the Link as Active. */

Router:PE1#show bundle bundle-ether 11

Bundle-Ether11
  Status:                                    LACP OOS  (out of service)
  Local links <active/standby/configured>:   0 / 1 / 1
  Local bandwidth <effective/available>:     0 (0) kbps
  MAC address (source):                      02cf.94c1.0a04 (Chassis pool)
  Inter-chassis link:                        No
  Minimum active links / bandwidth:          1 / 1 kbps
  Maximum active links:                      64
  Wait while timer:                          2000 ms
  Load balancing:
    Link order signaling:                    Not configured
    Hash type:                               Default
    Locality threshold:                      None
  LACP:                                      Operational
    Flap suppression timer:                  Off
    Cisco extensions:                        Disabled
    Non-revertive:                           Disabled
  mLACP:                                     Not configured
  IPv4 BFD:                                  Not configured
  IPv6 BFD:                                  Not configured

  Port                  Device           State        Port ID         B/W, kbps
  --------------------  ---------------  -----------  --------------
  Gi0/2/0/7             Local         Standby       0x8000, 0x0006     1000000
      Link is in standby due to bundle out of service state

/* PE1 is in the standby mode, hence the status shows as LACP OOS  (out of service) and the Link is
 in standby due to bundle out of service state. */

Router:CE#sshow bundle bundle-ether 11

Bundle-Ether11
  Status:                                    Up
  Local links <active/standby/configured>:   1 / 0 / 2
  Local bandwidth <effective/available>:     1000000 (1000000) kbps
  MAC address (source):                      02ff.566c.be04 (Chassis pool)
  Inter-chassis link:                        No
  Minimum active links / bandwidth:          1 / 1 kbps
  Maximum active links:                      64
  Wait while timer:                          2000 ms
  Load balancing:
    Link order signaling:                    Not configured
    Hash type:                               Default
    Locality threshold:                      None
  LACP:                                      Operational
    Flap suppression timer:                  Off
    Cisco extensions:                        Disabled
    Non-revertive:                           Disabled
  mLACP:                                     Not configured
  IPv4 BFD:                                  Not configured
  IPv6 BFD:                                  Not configured

  Port                  Device           State        Port ID         B/W, kbps
```

```
  --------------------  ---------------  -----------  --------------
  Gi0/0/0/8             Local            Active       0x8000, 0x0006     1000000
      Link is Active
  Gi0/0/0/16            Local            Negotiating  0x8000, 0x000b     1000000
      Partner is not Synchronized (Waiting, Standby, or LAG ID mismatch)

Router:PE2#show evpn ethernet-segment interface BE11 detail

/* The following output shows that the port-active mode is configured and the port is in the UP state.
 */

Ethernet Segment Id              Interface                          Nexthops
------------------------ ---------------------------------- --------------------
0011.1111.1111.0011.1111 BE11                                 192.168.0.2
                                                              192.168.0.3
  ES to BGP Gates   : Ready
  ES to L2FIB Gates : Ready
  Main port         :
     Interface name : Bundle-Ether11
     Interface MAC  : 02b4.3cb4.a004
     IfHandle       : 0x00004170
     State          : Up
     Redundancy     : Not Defined
  ESI type          : 0
     Value          : 11.1111.1111.0011.1111
  ES Import RT      : 1111.1111.1100 (from ESI)
  Source MAC        : 0000.0000.0000 (N/A)
  Topology          :
     Operational    : MH
     Configured     : Port-Active
  Service Carving   : Auto-selection
     Multicast      : Disabled
  Convergence       :
     Mobility-Flush : Count 0, Skip 0, Last n/a
  Peering Details   : 2 Nexthops
     192.168.0.2 [MOD:P:7fff]
     192.168.0.3 [MOD:P:00]
  Service Carving Results:
     Forwarders     : 0
     Elected        : 0
     Not Elected    : 0
  EVPN-VPWS Service Carving Results:
     Primary        : 0
     Backup         : 0
     Non-DF         : 0
  MAC Flushing mode : STP-TCN
  Peering timer     : 3 sec [not running]
  Recovery timer    : 20 sec [not running]
  Carving timer     : 0 sec [not running]
  Local SHG label   : None
  Remote SHG labels : 0
  Access signal mode: Bundle OOS (Default)


Router:PE1#show evpn ethernet-segment interface BE11 detail

/* The following output shows that the por-active mode is configured and the port is in the Standby
 state. */

Ethernet Segment Id              Interface                          Nexthops
------------------------ ---------------------------------- --------------------
0011.1111.1111.0011.1111 BE11                                 192.168.0.2
                                                              192.168.0.3
  ES to BGP Gates   : Ready
```

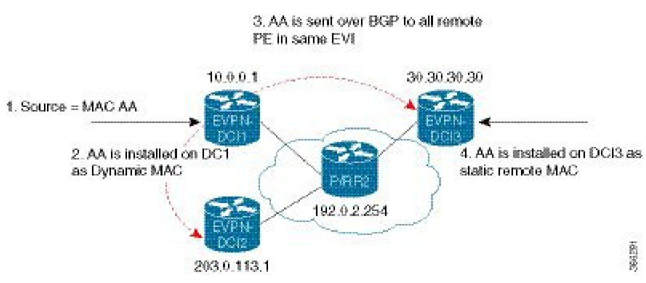
```
ES to L2FIB Gates : Ready
Main port         :
   Interface name : Bundle-Ether11
   Interface MAC  : 02cf.941c.0a04
   IfHandle       : 0x00004170
   State          : Standby
   Redundancy     : Not Defined
ESI type          : 0
   Value          : 11.1111.1111.0011.1111
ES Import RT      : 1111.1111.1100 (from ESI)
Source MAC        : 0000.0000.0000 (N/A)
Topology          :
   Operational    : MH
   Configured     : Port-Active
Service Carving   : Auto-selection
   Multicast      : Disabled
Convergence       :
   Mobility-Flush : Count 0, Skip 0, Last n/a
Peering Details   : 2 Nexthops
   192.168.0.2 [MOD:P:00]
   192.168.0.3 [MOD:P:7fff]
Service Carving Results:
   Forwarders     : 0
   Elected        : 0
   Not Elected    : 0
EVPN-VPWS Service Carving Results:
   Primary        : 0
   Backup         : 0
   Non-DF         : 0
MAC Flushing mode : STP-TCN
Peering timer     : 3 sec [not running]
Recovery timer    : 20 sec [not running]
Carving timer     : 0 sec [not running]
Local SHG label   : None
Remote SHG labels : 0
Access signal mode: Bundle OOS (Default)
```

EVPN IRB software MAC learning

MAC learning is the method of learning the MAC addresses of all devices available in a VLAN.

The MAC addresses learned on one device needs to be learned or distributed on the other devices in a VLAN. Software MAC learning feature enables the distribution of the MAC addresses learned on one device to the other devices connected to a network. The MAC addresses are learnt from the remote devices using BGP.

Figure 7: Software MAC learning



The above figure illustrates the process of software MAC learning. The following are the steps involved in the process:

1. Traffic comes in on one port in the bridge domain.
2. The source MAC address (AA) is learnt on DCI1 and is stored as a dynamic MAC entry.
3. The MAC address (AA) is converted into a type-2 BGP route and is sent over BGP to all the remote PEs in the same EVI.
4. The MAC address (AA) is updated on DCI3 as a static remote MAC address.

EVPN IRB software MAC aging

You can configure MAC aging on a bridge domain to set the maximum aging time for learned MAC addresses. Decrease the aging time when you want to move the hosts to allow the bridge to adapt to the changes quickly. However, in an EVPN network, the data plane and control plane are always synchronized. Furthermore, it is desirable to have a longer aging times for:

- MAC route stability and reliability
- Support for very high scale of MAC routes
- Reliable and consistent accounting without overloading the control plane

For the above-mentioned reasons, when you enable EVPN, maximum MAC aging times are not fully considered for the configured MAC aging values on the bridge domain. Also, it is observed that the aging times can be long, more than 2 hours.

MAC freezing

MAC freezing or duplicate IP address detection feature automatically detects any host with a duplicate IP address and blocks all MAC-IP routes that have a duplicate IP address.

This protects the network from hosts that are assigned duplicate IP addresses unintentionally or by malicious intent in an EVPN fabric. Hosts with duplicate IP address cause unnecessary churn in a network and causes traffic loss to either or both the hosts with the same IP address.

The system handles mobility of EVPN hosts by keeping track of MAC and IP addresses as they move from one host to another. If two hosts are assigned the same IP address, the IOS XR system keeps learning and re-learning MAC-IP routes from both the hosts. Each time it learns the MAC-IP route from one host, it is counted as one move since the newly learnt route supersedes the route previously learnt from the other host. This continues back and forth until the IP address is marked as duplicate based on the configured parameters.

It uses the following parameters to determine when an IP address should be marked as duplicate, and frozen or unfrozen as it moves between different hosts. The configurable parameters are:

- **move-interval**: The period within which a MAC or IP address has to move certain number of times between different hosts to be considered as duplicate and frozen temporarily. This number is specified in the **move-count** parameter.
- **move-count**: The number of times a MAC or IP address has to move within the interval specified for the **move-interval** parameter between different hosts to be considered a duplicate.

- **freeze-time**: The length of time a MAC or IP address is locked after it has been detected as a duplicate. After this period, the IP address is unlocked and it is allowed to learn again.
- **retry-count**: The number of times a MAC or IP address is unlocked after it has been detected as a duplicate before it is frozen permanently.

The system maintains a count of the number of times an IP address has been moved from one host to another host, either to another local host or to a host behind a remote Top of Rack (TOR). If an IP address moves certain number of times specified in the **move-count** parameter within the interval specified in the **move-interval** parameter is considered a duplicate IP address. All MAC-IP routes with that IP address is frozen for the time specified in the **freeze-time** parameter. A syslog notifies the user that the particular IP address is frozen. While an IP address is frozen, any new MAC-IP routes or updates to existing MAC-IP routes with the frozen IP address are ignored.

After **freeze-time** has elapsed, the corresponding MAC-IP routes are unfrozen and the value of the **move-count** is reset to zero. For any unfrozen local MAC-IP routes, an ARP probe and flush are initiated while the remote MAC-IP routes are put in the probe mode. This restarts the duplicate detection process.

The system also maintains the information about the number of times a particular IP address has been frozen and unfrozen. If an IP address is marked as duplicate after it is unfrozen **retry-count** times, it is frozen permanently until user manually unfreezes it. Use the following commands to manually unfreeze frozen MAC, IPv4 and IPV6 addresses respectively:

- **clear l2route evpn mac**{*mac-address*} | **all** [**evi** *evi*] **frozen-flag**
- **clear l2route evpn ipv4**{*ipv4-address*} | **all** [**evi** *evi*] **frozen-flag**
- **clear l2route evpn ipv6**{*ipv6-address*} | **all** [**evi** *evi*] **frozen-flag**

Configure MAC freezing

Perfrom these tasks to configure MAC freezing.

Procedure

Step 1 Configure duplicate detection for IPv4 and IPv6 addresses.

Example:

```
/* IPv4 address duplicate detection configuration */

Router# configure
Router(config)# evpn
Router(config-evpn)# host ipv4-address duplicate-detection
Router(config-evpn-host-ipv4-addr)# move-count 2
Router(config-evpn-host-ipv4-addr)# freeze-time 10
Router(config-evpn-host-ipv4-addr)# retry-count 2
Router(config-evpn-host-ipv4-addr)# commit
Router(config-evpn-host-ipv4-addr)# end

/* IPv6 address duplicate detection configuration */

Router# configure
Router(config)# evpn
Router(config-evpn)# host ipv6-address duplicate-detection
Router(config-evpn-host-ipv6-addr)# move-count 2
```

```
Router(config-evpn-host-ipv6-addr)# freeze-time 10
Router(config-evpn-host-ipv6-addr)# retry-count 2
Router(config-evpn-host-ipv6-addr)# commit
```

Step 2 Check the running configuration.

Example:

```
Router# show running-config
evpn
 host ipv4-address duplicate-detection
  move-count 2
  freeze-time 10
  retry-count 2
 !
!
evpn
 host ipv6-address duplicate-detection
  move-count 2
  freeze-time 10
  retry-count 2
 !
```

Step 3 Verify the details of the duplicate IP address detection and recovery parameters.

Example:

```
Router#show l2route evpn mac-ip all detail

Flags:  (Stt)=Static; (L)=Local; (R)=Remote; (F)=Flood;
        (N)=No Redistribution; (Rtr)=RP/0/RSP0/CPU0:router MAC; (B)=Best Route;
        (S)=Peer Sync; (Spl)=Split; (Rcv)=Recd;
        (D)=Duplicate MAC; (Z)=Frozen MAC;

Topo ID    Mac Address     IP Address  Prod   Next Hop(s)        Seq No  Flags         Opaque Data
Type    Opaque Data Len   Opaque Data Value
-------    -----------     ----------  ----   ----------        ------  -----        ----------------
    ---------------   -----------------
33         0022.6730.0001  10.130.0.2  L2VPN  Bundle-Ether6.1300  0      SB 0 12      0x06000000
```