BGP Control Plane

Feature Information for BGP Control Plane

Table 1: Feature Information for BGP Control Plane

Feature	Releases	Feature Information
PoAP diagnostics	7.2(0)N1(1)	Included a new section on *POAP Diagnostics*. POAP failure can be detected with locator LED.
Default Route Advertisement	7.2(0)N1(1)	Included a new section on *Default Route Advertisement*.
Border Leaf U-shape support	7.2(0)D1(1) 7.2(0)N1(1)	Included a new section on *Border Leaf Deployment Consideration* to support U-shape connectivity.

BGP Control Plane Setup

Multi-Protocol BGP (MP-BGP) is the primary protocol for exchanging host, subnet and default routes for IPv4 and IPv6 address families. MP-BGP based Control-Plane using EVPN NLRI (Network Layer Reachability Information) to transport end host information (IP and MAC) is used to transport the EVPN address family.

The following sections describe the reason for the POAP setting for BGP. The BGP configuration is same on all leaf nodes in the fabric. There are some additional knobs for the leaf node that are in the role of border leaf, we recommend that there be more than one border leaf in the fabric for redundancy reason. There are one or more switches that act as route reflectors (RR) that are configured on the spine, they have configuration related to being route reflectors. The following sections explain the general settings that apply to all leaf nodes, then knobs specific to border leaf and finally the spine knobs to act as a route reflector specific knob.

General BGP Configuration

Route-Target

Auto generated at the leaf and border leaf by combining the fabric ASN and Layer-3 Virtual Network Identifier (VNI).

The BGP route-target extended community is a path attribute shared by one or more routes in an UPDATE Message. Routes can be imported by using route-target as filter. Here, route-target carries a 2-byte ASN and a 4-byte VNI.

FABRIC ASN: VNI

Route Distinguisher

Auto generated at the leaf and border leaf by combining router ID and VRF ID. By making same route originated from different switches have a different Route Distinguisher (RD), the routes become unique. In MP-BGP, each route is uniquely qualified by a 8-byte RD. Here, the RD carries a 4-byte router ID and a 2-byte VRF ID.

Note The router ID is the same as the IP address configured on the backbone VLAN/SVI for BGP peering.

Switch router ID: local vrf id

Following is a sample configuration:

```
vrf context CiscoLive:Part4
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn
  address-family ipv6 unicast
    route-target both auto
    route-target both auto evpn
```

If the router ID is 220.1.1.1, local VRF ID is 4, Fabric AS is 65000, VNI is 65004 then RD = 220.1.1.1:4 and RT = 100:65004.

Add Path Support

The use of 'Add Path' is to allow one or more paths on a leaf and border leaf node to reach a given host. This facilitates Equal Cost Multipath (ECMP), faster convergence and host moves.

For example, spine acting as a router reflector:

```
route-map ALL-PATHS permit 10
  set path-selection all advertise

router bgp 65000

  address-family ipv4 unicast
    maximum-paths ibgp 2
    nexthop trigger-delay critical 250 non-critical 10000
    additional-paths send
    additional-paths selection route-map ALL-PATHS
```

At the leaf and border leaf:

```
route-map ALL-PATHS permit 10
  set path-selection all advertise
```

```
router bgp 65000

  address-family ipv4 unicast
    redistribute hmm route-map FABRIC-RMAP-REDIST-HOST
    maximum-paths ibgp 2
    nexthop trigger-delay critical 250 non-critical 10000
    nexthop route-map bgp_next_hop_filter
    additional-paths receive
    additional-paths selection route-map ALL-PATHS
```

General BGP settings for all Leaf nodes including Border Leaf

1. Feature BGP: Enables the feature on the box, needed on all leaf, border leaf and any spine that acts as route reflector.
2. BGP Router Autonomous System Number: All the nodes belong to one AS, this variable defines the AS value for the whole fabric.
    1. Every leaf node is connected to one or more route reflector neighbors. We recommend that you configure two route reflectors for redundancy purpose. At least one route reflector is needed in the fabric.

```
router bgp 65103

 router-id 192.0.2.1
    address-family ipv4 unicast
    address-family l2vpn evpn
 neighbor 192.0.2.10 remote-as 65103
    update-source loopback0
    address-family l2vpn evpn
      send-community both
 neighbor 192.0.2.20 remote-as 65103
    update-source loopback0
    address-family l2vpn evpn
      send-community both
```

    - Due to simplified topology, typically most nodes except the route reflector have not more than two BGP sessions, thus more aggressive BGP timers can be used to speed up convergence. The following command is used to speed up convergence in the fabric for node failures. The timer depends on the speed of convergence of FabricPath IS-IS and the removal of the BGP next hop IP address leading to withdrawal of the propagation of VRF prefixes. For more information, see examples in the following routing policy section.

Note The convergence time for FabricPath IS-IS is in subseconds as the number of nodes and label switched path (LSP) is less.

Usage example:

```
router bgp 65000

   address-family ipv4 unicast
    redistribute hmm route-map FABRIC-RMAP-REDIST-HOST
    maximum-paths ibgp 2
    nexthop trigger-delay critical 250 non-critical 10000
    nexthop route-map bgp_next_hop_filter
```

```
additional-paths receive
additional-paths selection route-map ALL-PATHS
```

- Default ECMP in unified fabric is set to 2. It is a balance between redundancy and saving hardware resources. This satisfies the common case of a vPC pair generating the same route and multiple border leaf nodes hosting the same VRF. For more information, see examples in the following routing policy section.

BGP Routing Policy

Cisco NX-OS operating system requires, that any route distribution passes through a route-map, to filter the distribution. Here are the policy statements that are configured by default through POAP. These are the common needs of unicast forwarding for hosts. This policy is configured on leaf and border leaf nodes.

1. Match any IPv4 address.

```
ip access-list HOSTS
  10 permit ip any any
```

2. Match any IPv6 address.

```
ipv6 access-list V6HOSTS
  10 permit ipv6 any any
```

3. The following route-maps allow the redistribution of all routes (IPv4 and IPv6 respectively) except for those learned over the control VLAN interface (backbone VLAN is used to set up the BGP topology). These route-maps are generally used for host redistribution via the HMM protocol.

```
route-map FABRIC-RMAP-REDIST-HOST deny 10
  match interface Vlan $$BACKBONE_VLAN$$
route-map FABRIC-RMAP-REDIST-HOST permit 20
  match ip address HOSTS

route-map FABRIC-RMAP-REDIST-V6HOST  deny 10
  match interface Vlan $$BACKBONE_VLAN$$
route-map FABRIC-RMAP-REDIST-V6HOST permit 20
  match ip address V6HOSTS

router bgp 65000

address-family ipv4 unicast
    redistribute hmm route-map FABRIC-RMAP-REDIST-HOST
    maximum-paths ibgp 2
    nexthop trigger-delay critical 250 non-critical 10000
    nexthop route-map bgp_next_hop_filter
    additional-paths receive
    additional-paths selection route-map ALL-PATHS
  address-family ipv6 unicast
    redistribute hmm route-map FABRIC-RMAP-REDIST-V6HOST
    maximum-paths ibgp 2
    nexthop trigger-delay critical 250 non-critical 10000
    additional-paths receive
```

4. The following route-map is used to redistribute server facing subnets. If the subnet route is tagged with the special value of 12345 then it will be redistributed. The same route-map works for IPv6 and IPv4 routes. Generally the default host facing configuration profiles will tag the subnet with this tag. It is highly

recommended that it should not be changed. If there is a subnet that does not require redistribution via BGP, then this tag should not be placed on it. There are several reasons for redistributing subnet address.

1. It enables border leaf to implement filtering policy for extended subnets.
2. It enables fabric to optimize Forwarding Information Base (FIB) usage in certain scenarios.

Multi-tenancy lite version

```
route-map FABRIC-RMAP-REDIST-SUBNET permit 10
  match tag 12345

interface Vlan3509
  no shutdown
  vrf member CiscoLive:Part4
  no ip redirects
  ip address 17.1.0.1/24 tag 12345
  no ipv6 redirects
  fabric forwarding mode proxy-gateway

route-map FABRIC-RMAP-REDIST-SUBNET permit 10
  match tag 12345

router bgp 65000

vrf CiscoLive:Part4
    address-family ipv4 unicast
      redistribute hmm route-map FABRIC-RMAP-REDIST-HOST
      redistribute direct route-map FABRIC-RMAP-REDIST-SUBNET
      maximum-paths ibgp 2
```

Multi-tenancy full version

```
route-map FABRIC-RMAP-REDIST-SUBNET permit 10
  match tag 12345

interface bdi3509
  no shutdown
  vrf member CiscoLive:Part4
  no ip redirects
  ip address 17.1.0.1/24 tag 12345
  no ipv6 redirects
  fabric forwarding mode proxy-gateway

route-map FABRIC-RMAP-REDIST-SUBNET permit 10
  match tag 12345

router bgp 65000

vrf CiscoLive:Part4
    address-family ipv4 unicast
      redistribute hmm route-map FABRIC-RMAP-REDIST-HOST
      redistribute direct route-map FABRIC-RMAP-REDIST-SUBNET
      maximum-paths ibgp 2
```

5. BGP next hop filter: For fast convergence, block the next hop resolution via the control subnet address. Generally all the BGP speakers within the fabric are on the same subnet. FabricPath IS-IS distributes the 32-bit local address for each switch via Link State Packet (LSP). For example, if the control subnet is 10.1.0.0/16, switches may have local IP addresses as 10.1.1.1, 10.1.1.2, 10.1.1.3 and so on. Essentially

each switch has the control backbone VLAN subnet 10.1.0.0 in the global routing table due to local configuration. It also has the 32-bit addresses learnt by FabricPath IS-IS. Suppose due to vPC pair or multiple border leaf there exists an ECMP path to prefix X. Consider if the path is advertised by 10.1.1.1 and 10.1.1.2. In steady state every other leaf in fabric will have an ECMP path to this prefix with the BGP next hop resolved via 10.1.1.1 and 10.1.1.2, now suppose switch 10.1.1.1 gets reloaded for upgrade, FabricPath IS-IS removes 10.1.1.1 immediately (subsecond) from all leaf nodes and withdraws the prefix propagation immediately. If **bgp_next_hop_filter** is not configured then convergence is delayed, as the next hop is resolved via the control subnet and route is not removed till BGP session timeout of route reflector with border leaf with IP address 10.1.1.1.

Note The Cisco NX-OS is an event trigger that uses next hop tracking and does not wait for BGP scan time.

In order to speed up convergence to subsecond, the following route-map is configured. It means if the route's next hop is resolved within control subnet, then it does not allow the subnet route to be used to resolve the next hop. For example, if FabricPath IS-IS removed the 32-bit next hop in subsecond but without the filter the route gets resolved through control subnet. Hence, route is not removed on FabricPath IS-IS event. It gets removed only when BGP session between RR and leaf and border leaf gets removed.

```
ip prefix-list control-subnet seq 100 permit $$BGP_CLIENT_SUBNET$$

route-map bgp_next_hop_filter deny 100
  match ip address prefix-list control-subnet
route-map bgp_next_hop_filter permit 200
  match ip address HOSTS

ip prefix-list control-subnet seq 100 permit 44.2.0.0/22
ip access-list HOSTS
  10 permit ip any any
route-map bgp_next_hop_filter deny 100
  match ip address prefix-list control-subnet
route-map bgp_next_hop_filter permit 200
  match ip address HOSTS
```

Usage example:

```
router bgp 65000

  address-family ipv4 unicast
    redistribute hmm route-map FABRIC-RMAP-REDIST-HOST
    maximum-paths ibgp 2
    nexthop trigger-delay critical 250 non-critical 10000
    nexthop route-map bgp_next_hop_filter
    additional-paths receive
    additional-paths selection route-map ALL-PATHS
```

BGP Settings for Border Leaf

The previous configurations and POAP settings apply to all leaf nodes including border leaf and some specific to route reflector. The following are settings that apply only to the border leaf.

1. **Default route generation from Border Leaf**: There are two options to do this:

- Option 1 is to advertise a default originate for all tenants by using a special route-target (RT) value. All the tenants that wants to use this border leaf will put this RT in the respective RT import statements as shown below, use this option if the number of VRFs is lesser than the maximum VRFs supported by border leaf and the total number of routes is also within the capability of the device, then use this variable to set up a default route for all VRFs. Option 1 is the default setting in the border leaf POAP template.
- Use one default route for all VRFs:

```
address-family vpnv4 unicast
   default-information originate always rd $$BACKBONE_IP$$:$$BGP_AS$$ route-target
 $$BGP_AS$$:$$BGP_RT_VNI$$

router bgp 65000

  address-family vpnv4 unicast
    default-information originate always rd 192.16.1.113:3 route-target 65000:9999
```

At interior leaf nodes, within every tenant:

```
vrf context CiscoLive:Part4
  vni 65004
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target import 65000:9999

vrf context CiscoLive:Part3
  vni 65005
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target import 65000:9999
BGP_RT_VNI, defaultValue=9999;
```

- Option 2 is to advertise VRF specific default originate from the border leaf. With this method, tenants can be load shared by different border leaf nodes by advertising only those tenants that border leaf is hosting. The border leaf scale is achieved by using per VRF default route generation. In that case, omit the default route generation by omitting this variable. The border leaf auto-configuration generates per VRF default route. If not using that feature then configure manually.
- per VRF default originate: This does not come with POAP, either to be entered manually or through border leaf auto-configuration for Layer-3 extension:

```
address-family vpnv4 unicast
          default-information originate rd router ID:Local VRF ID route-target
Fabric ASN:L3 Segment ID (VNI)
```

There is no need for a special route-target import statement at interior leaf. At the border leaf add the default information originate command per tenant under vpnv4 address family. The RD is constructed with border leaf BGP Router ID: Integer

The integer is a two by value, which is unique per tenant. The local VRF ID obtained by show VRF <vrf name> detail command can be used as the integer value 4 in the following example show VRF vpn1 in detail:

```
VRF-Name: vpn1, VRF-ID: 4, State: Up
```

```
    VPNID: unknown
    RD: 220.1.1.1:4
    VNI: 65004
    Max Routes: 0  Mid-Threshold: 0
    Table-ID: 0x80000003, AF: IPv6, Fwd-ID: 0x80000003, State: Up
    Table-ID: 0x00000003, AF: IPv4, Fwd-ID: 0x00000003, State: Up
```

The route-target is made up by Fabric ASN: VNI

At leaf node:

```
vrf context CiscoLive:Part4
  vni 65004
  rd auto
  address-family ipv4 unicast
    route-target both auto

vrf context CiscoLive:Part3
  vni 65005
  rd auto
  address-family ipv4 unicast
    route-target both auto
```

At border leaf:

```
router bgp 65000

  address-family vpnv4 unicast
   default-information originate always rd 192.16.1.113:4 route-target 65000:65004

   default-information originate always rd 192.16.1.113:5 route-target 65000:65005
```

2. **Fabric Site of Origin (SOO)**: Border leaf generates a fabric SOO and attaches it to routing updates going from outside the fabric to inside and inside the fabric to the outside. Fabric SOO is constructed by joining fabric ID and fabric AS. Interior leaf nodes use the fabric forwarding identifier and the local AS value to determine about fabric SOO. Only border leaf inserts SOO attribute in BGP updates.

   ```
   fabric-soo $$BGP_AS$$:$$FABRIC_ID$$
   ```

   POAP.FABRIC_ID: Fabric Identifier is one per fabric, every fabric must be configured with a unique integer as fabric ID. This helps to troubleshoot, which fabric a route is originated from and also to prevent loop. We recommend to start with 1 for first fabric and increment monotonically. The vPC leaf nodes insert SOO for dually homed hosts. They construct it with the following values:

   ```
   vPC Domain ID: Fabric Identifier

   fabric forwarding identifier 1
   router bgp 65000
     router-id 44.2.3.63
     fabric-soo 65000:1
   ```

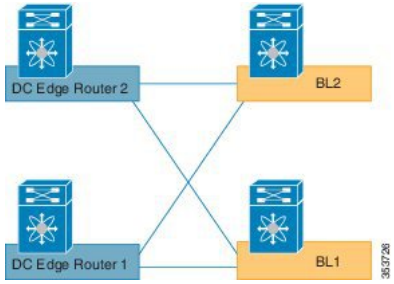
3. (Optional) For release 7.1(0)N1(1), POAP offers port-channels as only interface option between border leaf and DC edge routers. If you do not want to use port-channel, skip this step. Other types of interfaces can be configured manually. This limitation will be fixed in the future releases. Ensure that port admin is up and is not a switch port.

   For border leaf/edge router select the port-channel/interface ID as well as the interface range port-channel(s) towards DC edge router: The border leaf POAP optionally provides user to configure

a port-channel towards each of the DC edge routers it is neighbored with. We recommend that you configure this interface as port-channel even if there is only one member port. This should not be configured as a switch port so that Layer-3 sub interfaces can be configured on this for Layer-3 extension via sub interfaces. This is port-channel on which the border leaf auto-configuration will deploy sub interfaces for Layer-3 extension outside the fabric. It should match the value configured in Cisco Prime DCNM when pairing border leaf with DC edge router.

The following is the recommended topology for full redundancy. This POAP section is to enable it. One or two DC edge routers should be connected to border leaf. Two is the recommended number.

Figure 1: Recommended Topology for Full Redundancy



Note As there are multiple links to two or more edge routers, even if a link to an edge router goes down, it can still advertise the default route into the fabric without blackholing the traffic.

4. For border leaf/edge router select the port-channel/interface for default VRF peering: Global routing table peering with DC edge box: Border leaf provides user with prompt for configuring a sub interface on the port towards DC edge box and also the corresponding BGP session parameters. This is optional depending upon customer topology, need for default table routing and model used for internet access.
5. The border leaf has to be configured with switch role border.

```
fabric forwarding switch-role border
```

6. Set up the LDAP connection to the BL-DCI table. This is the table that enables auto-configuration of border leaf Layer-3 extension to the DC edge router. This is only done at border leaf in addition to the other LDAP tables set up at leaf nodes.

```
fabric database type network
server protocol ldap host rio-dcnm101a.cisco.com vrf management
db-table ou=networks,dc=cisco,dc=com key-type 1
db-security user cn=reader,dc=cisco,dc=com password 7 iwfw1c
fabric database type profile
server protocol ldap host rio-dcnm101a.cisco.com vrf management
db-table ou=profiles,dc=cisco,dc=com
db-security user cn=reader,dc=cisco,dc=com password 7 iwfw1c
fabric database type partition
server protocol ldap host rio-dcnm101a.cisco.com vrf management
db-table ou=partitions,dc=cisco,dc=com
db-security user cn=reader,dc=cisco,dc=com password 7 iwfw1c
fabric database type bl-dci
server protocol ldap host rio-dcnm101a.cisco.com vrf management
db-table ou=bl-dcis,dc=cisco,dc=com
db-security user cn=reader,dc=cisco,dc=com password 7 iwfw1c
```

Usage example:

```
fabric database type bl-dci
    server protocol ldap host ldap-server1.cisco.com vrf management
    db-security user cn=reader,dc=cisco,dc=com password1

db-security user admin password cis
  server protocol ldap host ldap-server2.cisco.com vrf management
    db-table ou=bl-dcis,dc=cisco,dc=com
   db-security user cn=reader,dc=cisco,dc=com password1
```

7. Border leaf specific tenant profile: Border leaf supports border leaf Layer-3 extension auto-configuration. Thus it needs a different profile than what is used by interior leaf nodes. The LDAP only allows one profile per tenant as the lookup key is only tenant name. The following command is used to override this locally at border leaf:

```
fabric database override-vrf-profile vrf-common-universal-bl
```

8. Border leaf should not accept default route from other border leaf nodes in the same fabric. This breaks ASBR function of border leaf and also leaks default route outside the fabric. The following commands are used to filter default route coming from route reflector neighbor.

Note The route-map 'deny-default-route' is required only when you run the previous versions of Cisco NX-OS 7.2(0)N1(1). From Cisco NX-OS 7.2(0)N1(1) or later, the import of default route advertised from the other border leaf node in the same fabric is supported.

```
ip prefix-list default-route seq 5 permit 0.0.0.0/0 le 1
route-map deny-default-route deny 100
  match ip address prefix-list default-route
route-map deny-default-route permit 200
  match ip address HOSTS

router bgp 65000

!Peering to the first RR
neighbor 44.2.0.101 remote-as 65000
    address-family ipv4 unicast
      send-community both
      route-map deny-default-route in
      next-hop-self
    address-family ipv6 unicast
      send-community extended
    address-family vpnv4 unicast
      send-community extended
      route-map deny-default-route in
    address-family vpnv6 unicast
      send-community extended
    address-family ipv4 mvpn
      send-community extended
    address-family ipv6 mvpn
      send-community extended
!Peering to the second RR
  neighbor 44.2.0.144 remote-as 65000
    address-family ipv4 unicast
      send-community both
      route-map deny-default-route in
      next-hop-self
```

```
address-family ipv6 unicast
  send-community extended
address-family vpnv4 unicast
  send-community extended
  route-map deny-default-route in
address-family vpnv6 unicast
  send-community extended
address-family ipv4 mvpn
  send-community extended
address-family ipv6 mvpn
```

9. Host based auto-configuration is disabled at border leaf. The feature evb along with the VDP configuration is missing on border leaf template for the same reason.

```
platform fabric database dot1q disable
```

Note Cisco NX-OS host attachment with auto-config at the border leaf is not supported.

10. For example for extension of tenant towards DC edge, see DC edge router on Appendix.

Default Route Advertisement

Default route advertisement for the default VRF from the border leaf

In case if the interior leaf nodes need to use the default VRF, the border leaf can advertise a default route towards the fabric.

There are several ways to do this, two are explained below.

Default route advertisement using redistribution of static route

Advantage

- Use this approach to withdraw static route from the fabric when external interfaces goes down.

Disadvantage

- The default route points towards external neighbors. Even in the presence of external default route.
- The static route is preferred over the external default route.

Default route advertisement using the 'default-originate' command

Default route advertisement using the **default-originate** command under the peer neighbor configuration context for fabric route reflector.

Advantage

- Simple to configure.

Disadvantage

- Does not withdraw default route even if external connectivity is lost.

Default route advertisement using redistribution of static route details

The recommended way to achieve this is implemented in the border leaf POAP templates. The POAP templates ensure that:

- The default route advertised by the Border Leaf does not leave the fabric by:
  - Attaching the well known community NO_EXPORT_COMMUNITY.
  - A deny route-map for default route on all external neighbors of border leaf.
- The LOCAL_PREFERENCE is set to 50, which is lower than the default preference of 100 when this route is received by other border leaf nodes. This ensures that the border leaf prefers the external default route.
- The admin distance of the static route at the border leaf is set to 254, so that the default routes learnt from external neighbors are always preferred over the locally configured static route.

Route Map and Prefix lists Configuration

```
ip prefix-list default-route seq 5 permit 0.0.0.0/0 le 1
route-map DEFAULT-ROUTE-MODIFY permit 100
  match ip address prefix-list default-route
  set local-preference 50
  set community no-export
route-map DEFAULT-ROUTE-MODIFY permit 1000
route-map DEFAULT-ROUTE-MODIFY-V6 permit 100
  match ipv6 address prefix-list default-route-v6
  set local-preference 50
  set community no-export
route-map DEFAULT-ROUTE-MODIFY-V6 permit 1000

route-map DENY-DEFAULT-ROUTE deny 10
  match ip address prefix-list default-route
route-map DENY-DEFAULT-ROUTE permit 1000

route-map FABRIC-RMAP-REDIST-STATIC permit 10
  match ip address prefix-list default-route

route-map ALL-PATHS permit 10
  set path-selection all advertise
```

Configuration details

Configuration specific to default route origination is given below.

Two box border leaf solution.

```
!Dc Edge facing sub interfaces

interface Ethernet1/35.10
  encapsulation dot1Q 10
  ip address 30.1.1.1/24

interface Ethernet1/36.10
  encapsulation dot1Q 10
  ip address 40.1.1.1/24
```

```
!Static route towards DC Edge

ip route 0.0.0.0/0 30.1.1.2 254
ip route 0.0.0.0/0 40.1.1.2 254


!Relevant BGP configuration

router bgp 65000
  router-id 128.89.0.20
  fabric-soo 65000:1
  address-family ipv4 unicast
    redistribute hmm route-map FABRIC-RMAP-REDIST-HOST
    redistribute static route-map FABRIC-RMAP-REDIST-STATIC
    maximum-paths ibgp 2
    nexthop trigger-delay critical 250 non-critical 10000
    nexthop route-map BGP_NEXT_HOP_FILTER
    default-information originate
    additional-paths receive
    additional-paths selection route-map ALL-PATHS
/**RR neighbor**/
neighbor 128.89.0.100 remote-as 65000
    address-family ipv4 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY in
      next-hop-self

[SNIP]
/*external neighbors */
neighbor 30.1.1.2 remote-as 300
    peer-type fabric-external
    address-family ipv4 unicast
      send-community both
      route-map DENY-DEFAULT-ROUTE out
neighbor 40.1.1.2 remote-as 300
    peer-type fabric-external
    address-family ipv4 unicast
      send-community both
      route-map DENY-DEFAULT-ROUTE out
```

default-information originate

Allows default route to be redistributed. By default, the default route is not redistributed without explicitly allowing the redistribution through this command.

Two Box Border leaf

For the two box solution, point the static route next hop to the DC-EDGE router address for the sub interface. This will ensure that the default route is withdrawn when the interface goes down. This is automatically done if a POAP template is used to configure the interface and sub-interface towards DC-EDGE box.

BorderPe

There are two options for BorderPe based on your preference:

1. Point the default static route to MPLS VPN facing interfaces.
    - Will be withdrawn if external connectivity is lost.
    - POAP template uses this approach for IPv4.

2. Point the default static route to NULL0.
    - Will not be withdrawn if external connectivity is lost.

Default route using default originate commands under RR neighbors Details

Default route can be advertised towards fabric by default originate command as shown below.

Route Map and Prefix lists Configuration.

```
ip prefix-list default-route seq 5 permit 0.0.0.0/0 le 1
ipv6 prefix-list default-route-v6 seq 5 permit 0::/0
route-map DENY-DEFAULT-ROUTE deny 10
  match ip address prefix-list default-route
route-map DENY-DEFAULT-ROUTE permit 1000
```

BGP configuration.

```
router bgp 65000
  router-id 128.89.0.20
  fabric-soo 65000:1
  address-family ipv4 unicast
    redistribute hmm route-map FABRIC-RMAP-REDIST-HOST
    maximum-paths ibgp 2
    nexthop trigger-delay critical 250 non-critical 10000
    nexthop route-map BGP_NEXT_HOP_FILTER
    additional-paths receive
    additional-paths selection route-map ALL-PATHS
/**External neighbors */
neighbor 30.1.1.2 remote-as 300
    peer-type fabric-external
    address-family ipv4 unicast
      send-community both
      route-map DENY-DEFAULT-ROUTE out

  neighbor 40.1.1.2 remote-as 300
    peer-type fabric-external
    address-family ipv4 unicast
      send-community both
      route-map DENY-DEFAULT-ROUTE out
/**RR neighbor**//

  neighbor 128.89.0.100 remote-as 65000
    address-family ipv4 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY in
      default-originate
      next-hop-self
```

Border Leaf Deployment Consideration

The previous configurations and description applies to border leaf topologies with full-meshed peering with the DC edge routers. Full-meshed connectivity between border leaf and DC edge router along with node redundancy is recommended, as this topology supports the highest level of redundancy. However there are certain deployments, where the full-meshed approach cannot be deployed.

For example, consider a case where the physical installation of the border leaf nodes and DC edge routers are in different buildings and you only have limited fiber available. Such topologies, where each border leaf has only a single link to the DC edge router is often called U-shape.

By default, the border leaf when using default route configuration according to option 1 (default route injection for all VRF with default-information originate always configured under the VPNv4/6 address family) injects the default route in the fabric independent if the external link towards the DC edge router is down or up.

In U-shape topologies this could cause blackholing for certain flows, as the traffic sourced on the leaf switch is hashed along the two default routes (ECMP) injected by the two border leaf nodes.

Cisco NX-OS version 7.2(0)D1(1) or 7.2(0)N1(1) or later is required on the border leaf as this software version supports the U-shape topology. The DCNM 7.2(1) with the V3 POAP templates (for example, Fabric_N5600_N6K_BorderLeaf_v3) provides the required configuration.

The border leaf POAP template provides the required configuration with the specific route-maps to avoid blackholing as shown below.

```
ip prefix-list default-route seq 5 permit 0.0.0.0/0 le 1
ipv6 prefix-list default-route-v6 seq 5 permit 0::/0

route-map DEFAULT-ROUTE-MODIFY permit 100
  match ip address prefix-list default-route
  set local-preference 50
route-map DEFAULT-ROUTE-MODIFY permit 1000

router bgp 65000

!Peering to the first RR
neighbor 44.2.0.101 remote-as 65000
    address-family ipv4 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY in
      next-hop-self
    address-family ipv6 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY-V6 in
      next-hop-self
    address-family vpnv4 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY in
    address-family vpnv6 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY-V6 in
    address-family ipv4 mvpn
      send-community both
    address-family ipv6 mvpn
      send-community both

!Peering to the second RR
neighbor 44.2.0.144 remote-as 65000
    address-family ipv4 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY in
      next-hop-self
    address-family ipv6 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY-V6 in
      next-hop-self
    address-family vpnv4 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY in
```

```
    address-family vpnv6 unicast
      send-community both
      route-map DEFAULT-ROUTE-MODIFY-V6 in
    address-family ipv4 mvpn
      send-community both
    address-family ipv6 mvpn
      send-community both
```

Note The route-map 'deny-default-route' as shown in the previous section is replaced by the route-map 'DEFAULT-ROUTE-MODIFY'.

The route-map "DEFAULT-ROUTE-MODIFY" along the additional BGP route-target import statement 65000:9999 (same as on the interior leaf) will re-import the default route advertised by the other border leaf. This default route is imported per specific tenant with a lower local preference and is only installed in the forwarding table when the external learnt default route from the DC edge router is unavailable.

At border leaf, within every tenant:

```
vrf context CiscoLive:Part4
  vni 65004
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target import 65000:9999

vrf context CiscoLive:Part3
  vni 65005
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target import 65000:9999
```

POAP Diagnostics

POAP failure can be detected with locator LED. When the POAP process starts, the locator-LED will flash the pattern 21 (flashing twice, short pause, flashing once, long pause) to indicate that POAP is in progress.

The device has several LEDs such as chassis LED, status LED, port LED, and so on. For PoAP diagnostics, you must follow the chassis (beacon) LED flashing in pattern 21.

Given below are the LED flashing in a pattern that is recognizable and visible to human eyes:

Table 2: Chassis LED Pattern

LED Pattern	Description
Blue LED - two long flashes, pause, one short flash, long pause	PoAP is currently running. If this pattern is flashing and not turned off after a considerable amount of time lapse, it indicates PoAP is in error condition such as DHCP discovery failure or script execution failure.
No light	PoAP is not running.

Router Reflector Configuration

Use subnet for neighbor address so that each neighbor does not have to be explicitly configured.

For example, if control subnet is 192.168.99.0/24 and fabric as is 65101.

```
router bgp 65101

 router-id 192.168.99.1
 address-family ipv4 unicast
   maximum-paths ibgp 2
   nexthop trigger-delay critical 250 non-critical 10000
   additional-paths send
   additional-paths selection route-map ALL-PATHS
 address-family ipv6 unicast
   maximum-paths ibgp 2
   nexthop trigger-delay critical 250 non-critical 10000
   additional-paths send
   additional-paths selection route-map ALL-PATHS
 address-family vpnv4 unicast
   nexthop trigger-delay critical 250 non-critical 10000
   additional-paths send
   additional-paths selection route-map ALL-PATHS
 address-family vpnv6 unicast
   nexthop trigger-delay critical 250 non-critical 10000
   additional-paths send
   additional-paths selection route-map ALL-PATHS
 address-family ipv4 mvpn
   nexthop trigger-delay critical 250 non-critical 10000
   additional-paths send
   additional-paths selection route-map ALL-PATHS
 address-family ipv6 mvpn
   nexthop trigger-delay critical 250 non-critical 10000
   additional-paths send
   additional-paths selection route-map ALL-PATHS
   neighbor 192.168.99.0/24 remote-as 65101
 address-family ipv4 unicast
   send-community both
   route-reflector-client
 address-family ipv6 unicast
   send-community extended
   route-reflector-client
 address-family vpnv4 unicast
   send-community extended
   route-reflector-client
 address-family vpnv6 unicast
   send-community extended
   route-reflector-client
 address-family ipv4 mvpn
   send-community extended
   route-reflector-client
 address-family ipv6 mvpn
   send-community extended
   route-reflector-client
```