Configuring VXLAN EVPN Multi-Site

This chapter contains these sections:

About VXLAN EVPN Multi-Site

The VXLAN EVPN Multi-Site solution interconnects two or more BGP-based Ethernet VPN (EVPN) sites/fabrics (overlay domains) in a scalable fashion over an IP-only network. This solution uses border gateways (BGWs) in anycast or vPC mode to terminate and interconnect two sites. The BGWs provide the network control boundary that is necessary for traffic enforcement and failure containment functionality.

In the BGP control plane for releases prior to Cisco NX-OS Release 9.3(5), BGP sessions between the BGWs rewrite the next hop information of EVPN routes and reoriginate them. Beginning with Cisco NX-OS Release 9.3(5), reorigination is always enabled (with either single or dual route distinguishers), and rewrite is not performed. For more information, see Dual RD Support for Multi-Site, on page 3.

VXLAN Tunnel Endpoints (VTEPs) are only aware of their overlay domain internal neighbors, including the BGWs. All routes external to the fabric have a next hop on the BGWs for Layer 2 and Layer 3 traffic.

The BGW is the node that interacts with nodes within a site and with nodes that are external to the site. For example, in a leaf-spine data center fabric, it can be a leaf, a spine, or a separate device acting as a gateway to interconnect the sites.

The VXLAN EVPN Multi-Site feature can be conceptualized as multiple site-local EVPN control planes and IP forwarding domains interconnected via a single common EVPN control and IP forwarding domain. Every EVPN node is identified with a unique site-scope identifier. A site-local EVPN domain consists of EVPN nodes with the same site identifier. BGWs on one hand are also part of the site-specific EVPN domain and on the other hand a part of a common EVPN domain to interconnect with BGWs from other sites. For a given site, these BGWs facilitate site-specific nodes to visualize all other sites to be reachable only via them. This means:

- Site-local bridging domains are interconnected only via BGWs with bridging domains from other sites.

- Site-local routing domains are interconnected only via BGWs with routing domains from other sites.

- Site-local flood domains are interconnected only via BGWs with flood domains from other sites.

Selective Advertisement is defined as the configuration of the per-tenant information on the BGW. Specifically, this means IP VRF or MAC VRF (EVPN instance). In cases where external connectivity (VRF-lite) and EVPN Multi-Site coexist on the same BGW, the advertisements are always enabled.

Note   The MVPN VRI ID must be configured for TRM on anycast BGW if the site ID is greater than two bytes. The same VRI ID needs to be configured in all anycast BGWs that are part of the same site. However, the VRI ID must be unique within the network. That is, other anycast BGWs or vPC leaves must use different VRI IDs.

About VXLAN EVPN Multi-Site with IPv6 Underlay

Beginning with Cisco NX-OS Release 10.4(3)F, the support is provided for VXLAN EVPN Multi-Site with IPv6 Underlay.

Figure 1: Topology - VXLAN EVPN Multi-Site with IPv6 Underlay
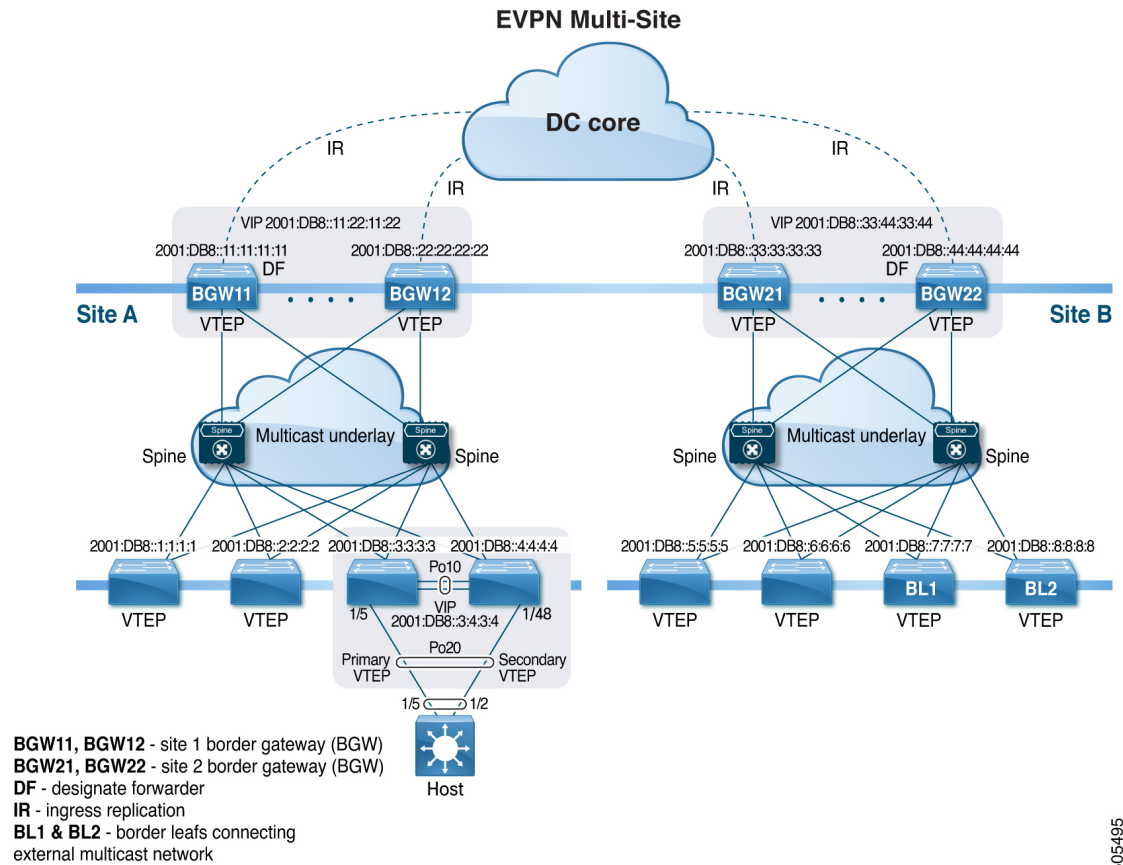
The above topology shows four leafs and two spines in the VXLAN EVPN fabric and two Anycast BGWs. Inside the fabric, the underlay is an IPv6 Multicast running PIMv6. RP is positioned in the spine with anycast RP. BGWs support VXLAN with IPv6 Protocol-Independent Multicast (PIMv6) Any-Source Multicast (ASM) on the fabric side and Ingress Replication (IPv6) on the DCI side.

Beginning with Cisco NX-OS Release 10.5(1)F, the underlay network supports the following combinations for VXLAN EVPN Multi-Site:

- In the data center fabric, both Multicast Underlay (PIMv6) Any-Source Multicast (ASM) and Ingress Replication (IPv6) are supported.

- In the Data Center Interconnect (DCI), only Ingress Replication (IPv6) is supported.

Dual RD Support for Multi-Site

Beginning with Cisco NX-OS Release 9.3(5), VXLAN EVPN Multi-Site supports route reorigination with dual route distinguishers (RDs). This behavior is enabled automatically.

Each VRF or L2VNI tracks two RDs: a primary RD (which is unique) and a secondary RD (which is the same across BGWs). Reoriginated routes are advertised with the secondary type-0 RD (site-id:VNI). All other

routes are advertised with the primary RD. The secondary RD is allocated automatically once the router is in Multi-Site BGW mode.

If the site ID is greater than 2 bytes, the secondary RD can't be generated automatically on the Multi-Site BGW, and the following message appears:

```
%BGP-4-DUAL_RD_GENERATION_FAILED: bgp- [12564] Unable to generate dual RD on EVPN multisite
 border gateway. This may increase memory consumption on other BGP routers receiving
re-originated EVPN routes. Configure router bgp <asn> ; rd dual id <id> to avoid it.
```

In this case, you can either manually configure the secondary RD value or disable dual RDs. For more information, see Configuring Dual RD Support for Multi-Site, on page 14.

RP Placement in the DCI Core

- PIM RP placement for DCI multicast underlay

  - PIM RP and multicast groups for fabric underlay and DCI underlay must be different.

  - DCI underlay group must not overlap with fabric underlay group.

  - Multicast groups and RPs in DCI and fabric networks should be distinct and configured based on specific ranges.

  - RP can be placed on any node in the DCI core, there can be one or more RPs.

- Direct BGW to BGW Peering Deployment

  - PIM RP should be set on BGWs with anycast PIM RP for redundancy.

- BGW to Cloud Model Deployment

  - PIM RP could be placed on the DCI underlay layer, such as core routers or superspines.

Interoperability with EVPN Multi-Homing Using ESI for Multi-Site Anycast BGW

Beginning Cisco NX-OS Release 10.2(2)F, EVPN MAC/IP routes (Type 2) with non-reserved as well as with reserved ESI (0 or MAX-ESI) values are evaluated for forwarding (ESI RX). The definition of the EVPN MAC/IP route resolution is defined in RFC 7432 Section 9.2.2.

EVPN MAC/IP routes (Type 2) -

- with reserved ESI value (0 or MAX-ESI) are resolved solely by the MAC/IP route alone (BGP next-hop within Type 2).

- with non-reserved ESI value are resolved only if an accompanied per-ES Ethernet Auto-Discovery route (Type 1, per-ES EAD) is present.

In addition to the MAC/IP route resolution as mentioned above, the Multi-Site BGW supports the forward, rewrite and re-originate of MAC/IP routes with reserved and non-reserved ESI values. In all these cases, the per-ES EAD route is re-originated by the Multi-Site BGW.

The EVPN MAC/IP route resolution with the different ESI values is supported on Cisco Nexus 9300-EX, -FX, -FX2, -FX3, and -GX Platform Switches in Anycast and vPC Border Gateway mode.

Guidelines and Limitations for VXLAN EVPN Multi-Site

VXLAN EVPN Multi-Site has the following configuration guidelines and limitations:

- The following switches support VXLAN EVPN Multi-Site:

    - Cisco Nexus 9300-EX and 9300-FX platform switches

    - Cisco Nexus 9300-FX2 platform switches

    - Cisco Nexus 9300-FX3 platform switches

    - Cisco Nexus 9300-GX platform switches

    - Cisco Nexus 9300-GX2 platform switches

    - Cisco Nexus 9332D-H2R switches

    - Cisco Nexus 93400LD-H1 switches

    - Cisco Nexus 9364C-H1 switches

    - Cisco Nexus 9800 platform switches with X9836DM-A and X98900CD-A line cards

    - Cisco Nexus 9500 platform switches with -EX or -FX or -GX or -FX3 line cards

> **Note**   Cisco Nexus 9500 platform switches with -R/RX line cards don't support VXLAN EVPN Multi-Site.

    - Beginning with Cisco NX-OS Release 10.2(3)F, the VXLAN EVPN Multi-Site is supported on the Cisco Nexus 9300-GX2 platform switches.

    - Beginning with Cisco NX-OS Release 10.4(1)F, the VXLAN EVPN Multi-Site is supported on the Cisco Nexus 9332D-H2R switches.

    - Beginning with Cisco NX-OS Release 10.4(2)F, the VXLAN EVPN Multi-Site is supported on the Cisco Nexus 93400LD-H1 switches.

    - Beginning with Cisco NX-OS Release 10.4(3)F, the VXLAN EVPN Multi-Site is supported on the Cisco Nexus 9364C-H1 switches.

    - Beginning with Cisco NX-OS Release 10.5(2)F, the following features are supported on Cisco Nexus 9500 Series switches with N9K-X9736C-FX3 line card.

        - Multi-Hop BFD

        - VXLAN and iVXLAN stripping

        - DCI advertise PIP (without cloudsec) on vPC and Anycast BGW

Switch or Port restrictions

- The **evpn multisite dci-tracking** is mandatory for anycast BGWs and vPC BGW DCI links.

- EVPN multisite DCI-tracking and EVPN multisite fabric-tracking are only supported on physical interfaces. Use on SVIs is not supported.

Deployment restrictions

- In a VXLAN EVPN Multi-Site deployment, when you use the ttag feature, make sure that the ttag is stripped (**ttag-strip**) on BGW's DCI interfaces attached to Non-NXOS gear.

- VXLAN EVPN Multi-Site and Tenant Routed Multicast (TRM) are supported between sources and receivers deployed across different sites.

- The Multi-Site BGW allows the coexistence of Multi-Site extensions (Layer 2 unicast/multicast and Layer 3 unicast) as well as Layer 3 unicast and multicast external connectivity.

- In TRM with multi-site deployments, all BGWs receive traffic from fabric. However, only the designated forwarder (DF) BGW forwards the traffic. All other BGWs drop the traffic through a default drop ACL. This ACL is programmed in all DCI tracking ports. Don't remove the **evpn multisite dci-tracking** configuration from the DCI uplink ports. If you do, you remove the ACL, which creates a nondeterministic traffic flow in which packets can be dropped or duplicated instead of deterministically forwarded by only one BGW, the DF.

- Prior to NX-OS 10.2(2)F only ingress replication was supported between DCI peers across the core. Beginning with Cisco NX-OS Release 10.2(2)F both ingress replication and multicast are supported between DCI peers across the core.

- The DCI underlay group and the fabric underlay group must be distinct, ensuring no overlap between DCI multicast and fabric multicast underlay groups.

- Bind NVE to a loopback address that is separate from loopback addresses that are required by Layer 3 protocols. A best practice is to use a dedicated loopback address for the NVE source interface (PIP VTEP) and multi-site source interface (anycast and virtual IP VTEP).

- Beginning with Cisco NX-OS Release 9.3(5), if you disable the **host-reachability protocol bgp** command under the NVE interface in a VXLAN EVPN Multi-Site topology, the NVE interface stays operationally down.

- Beginning with Cisco NX-OS Release 9.3(5), Multi-Site Border Gateways re-originate incoming remote routes when advertising to the site's local spine/leaf switches. These re-originated routes modify the following fields:

    - RD value changes to [Multisite Site ID:L3 VNID].

    - It is mandatory that Route-Targets are defined on all VTEP that are participating in a given VRF, this includes and is explicitly required for the BGW to extend the given VRF. Prior to Cisco NX-OS Release 9.3(5), Route-Targets from intra-site VTEPs were inadvertently kept across the site boundary, even if not defined on the BGW. Starting from Cisco NX-OS Release 9.3(5) the mandatory behavior is enforced. By adding the necessary Route-Targets to the BGW, the change from inadvertent Route-Target advertisement to explicit Route-Target advertisement can be performed.

    - Path type changes from external to local.

- For SVI-related triggers (such as shut/unshut or PIM enable/disable), a 30-second delay was added, allowing the Multicast FIB (MFIB) Distribution module (MFDM) to clear the hardware table before toggling between L2 and L3 modes or vice versa.

- Ensure that the **ip pim sparse-mode** is enabled on the Multi-Site VIP loopback interface.

- To improve the convergence in case of fabric link failure and avoid issues in case of fabric link flapping, ensure to configure multi-hop BFD between loopbacks of spines and BGWs.

  In the specific scenario where a BGW node becomes completely isolated from the fabric due to all its fabric links failing, the use of multi-hop BFD ensures that the BGP sessions between the spines and the isolated BGW can be immediately brought down, without relying on the configured BGP hold-time value.

- In a VXLAN Multi-Site environment, a border gateway device that uses ECMP for routing through both a VXLAN overlay and an L3 prefix to access remote site subnets might encounter adjacency resolution failure for one of these routes. If the switch attempts to use this unresolved prefix, it will result in traffic being dropped.

- To improve convergence during the reload of anycast BGW routers in multi-plane topology, ensure to configure the **multi-hop BFD** and **nexthop trigger-delay** commands.

- Following guidelines and limitations are applied when a multisite Border Gateway is put into Maintenance Mode:

  - BUM Traffic from remote Fabrics will still be attracted to the Border gateway that is in maintenance mode

  - Border Gateway in maintenance mode still participates in Designated Forwarder Election

  - Default Maintenance mode profile applies the command "ip pim isolate" and so the Border gateway is isolated from S,G tree towards the fabric direction. This leads to BUM traffic loss and hence an appropriate maintenance mode profile should be used for Border Gateways than the default.

vPC BGW restrictions

- BGWs in a vPC topology are supported.

- vPC mode can support only two BGWs.

- vPC mode can support both Layer 2 hosts and Layer 3 services on local interfaces.

- In vPC mode, BUM is replicated to either of the BGWs for traffic coming from the external site. Hence, both BGWs are forwarders for site external to site internal (DCI to fabric) direction.

- In vPC mode, BUM is replicated to either of the BGWs for traffic coming from the local site leaf for a VLAN using Ingress Replication (IR) underlay. Both BGWs are forwarders for site internal to site external (fabric to DCI) direction for VLANs using the IR underlay.

- In vPC mode, BUM is replicated to both BGWs for traffic coming from the local site leaf for a VLAN using the multicast underlay. Therefore, a decapper/forwarder election happens, and the decapsulation winner/forwarder only forwards the site-local traffic to external site BGWs for VLANs using the multicast underlay.

- In vPC mode, all Layer 3 services/attachments are advertised in BGP via EVPN Type-5 routes with their virtual IP as next hop. If the VIP/PIP feature is configured, they are advertised with PIP as the next hop.

Unsupported features

- Multicast Flood Domain between inter-site/fabric BGWs isn't supported.

- iBGP EVPN Peering between BGWs of different fabrics/sites isn't supported.

- PIM BiDir is not supported for fabric underlay multicast replication with VXLAN Multi-Site.

- FEX is not supported on a vPC BGW and Anycast BGW.

Anycast BGW restrictions

- Anycast mode can support up to six BGWs per site.

- Anycast mode can support only Layer 3 services that are attached to local interfaces.

- In Anycast mode, BUM is replicated to each border leaf. DF election between the border leafs for a particular site determines which border leaf forwards the inter-site traffic (fabric to DCI and conversely) for that site.

- In Anycast mode, all Layer 3 services are advertised in BGP via EVPN Type-5 routes with their physical IP as the next hop.

- If different Anycast Gateway MAC addresses are configured across sites, enable ARP suppression and ND suppression for all VLANs that have been extended.

Supported features

- Beginning with Cisco NX-OS Release 9.3(5), VTEPs support VXLAN-encapsulated traffic over parent interfaces if subinterfaces are configured. This feature is supported for VXLAN EVPN Multi-Site and DCI. DCI tracking can be enabled only on the parent interface.

- Beginning with Cisco NX-OS Release 9.3(5), VXLAN EVPN Multi-Site supports asymmetric VNIs. For more information, see Multi-Site with Asymmetric VNIs and .

- **Dual RD**

  The following guidelines and limitations apply to dual RD support for Multi-Site:

  - Dual RD are supported beginning with Cisco NX-OS Release 9.3(5).

  - Dual RD is enabled automatically for Cisco Nexus 9332C, 9364C, 9300-EX, and 9300-FX/FX2 platform switches and Cisco Nexus 9500 platform switches with -EX/FX/FX3 line cards that have VXLAN EVPN Multi-Site enabled.

  - Beginning with Cisco NX-OS Release 10.2(3)F, the dual RD support for Multi-Site is supported on the Cisco Nexus 9300-FX3 platform switches.

  - To use CloudSec or other features that require PIP advertisement for multi-site reoriginated routes, configure BGP additional paths on the route server if dual RD are enabled on the BGW, or disable dual RD.

  - Sending secondary RD additional paths at the BGW node isn't supported.

  - During an ISSU, the number of paths for the leaf nodes might double temporarily while all BGWs are being upgraded.

Guidelines and Limitations for VXLAN Multi-Site Anycast BGW Support on Cisco Nexus 9800 Series Switches

- Beginning with Cisco NX-OS Release 10.4(3)F, the VXLAN Multi-Site Anycast BGW is supported on the Cisco Nexus 9808/9804 switches with X9836DM-A and X98900CD-A line cards.

  - VXLAN Multi-Site Anycast BGW supports the following features:

    - VXLAN BGP EVPN fabric and multi-site interconnect

    - VXLAN Layer-2 VNI and new Layer-3 VNI which is not VLAN based

    - IPv4 underlay

    - IPv6 underlay

    - Ingress Replication on fabric and DCI side

    - Multicast underlay in Fabric

    - Bud node

    - TRMv4

    - TRMv6

    - NGOAM

    - VXLAN Counters

      - Per VXLAN peer based total packet/byte counters are supported.

      - Per VNI based total packet/byte counters are supported

  - VXLAN Multi-Site Anycast BGW does not support the following features:

    - Downstream VNI and route leak

    - L3 Port channel as a fabric or DCI link

    - Multicast underlay on DCI side

    - VXLAN access features

    - IGMP snooping

    - Separate VXLAN counters for broadcast, multicast, and unicast traffic

    - Data MDT

    - EVPN storm control

Guidelines and Limitations for VXLAN EVPN Multi-Site with IPv6 Underlay

VXLAN EVPN Multi-Site with IPv6 Underlay has the following configuration guidelines and limitations:

- Cisco Nexus 9300-FX, FX2, FX3, GX, GX2, H2R and H1 ToR switches are supported as the leaf VTEP or BGW.

- Cisco Nexus N9K-X9716D-GX and N9K-X9736C-FX, N9K-X9736C-FX3 line cards are supported only on the spine (EoR).

- When an EoR is deployed as a spine node with Multicast Underlay (PIMv6) Any-Source Multicast (ASM), it is mandatory to configure non-default template using one of the following commands in global configuration mode:

  - **system routing template-multicast-heavy**

  - **system routing template-multicast-ext-heavy**

- vPC BGWs are not supported with IPv6 multicast underlay.

- Dual stack configuration is not supported for NVE source interface loopback and multi-site interface loopback.

- Beginning with Cisco NX-OS Release 10.5(1)F, VXLAN EVPN Multi-Site in the data center fabric supports both Multicast Underlay (PIMv6) Any-Source Multicast (ASM) and Ingress Replication (IPv6) in the underlay. This support is available on the following switches and line cards:

  - Cisco Nexus 9300-FX, FX2, FX3, GX, GX2, H2R, and H1 ToR switches as the leaf VTEPs.

  - Cisco Nexus N9K-X9716D-GX and N9K-X9736C-FX line cards as spines if the underlay is configured for Multicast Underlay (PIMv6) Any-Source Multicast (ASM).

  - Cisco Nexus N9K-X9716D-GX and N9K-X9736C-FX line cards as VTEPs if the underlay uses Ingress Replication (IPv6).

- Beginning with Cisco NX-OS Release 10.5(2)F, VXLAN EVPN Multi-Site with IPv6 Underlay support is extended on Cisco Nexus 9500 Series switches with N9K-X9736C-FX3 line cards as VTEPs if the underlay uses Ingress Replication (IPv6).

Enabling VXLAN EVPN Multi-Site

This procedure enables the VXLAN EVPN Multi-Site feature. Multi-Site is enabled on the BGWs only. The site-id must be the same on all BGWs in the fabric/site.

Procedure

	Command or Action	**Purpose**
Step 1	**configure terminal** **Example:** switch# **configure terminal**	Enters global configuration mode.
Step 2	**evpn multisite border-gateway** *ms-id* **Example:** switch(config)# **evpn multisite border-gateway 100**	Configures the site ID for a site/fabric. The range of values for *ms-id* is 1 to 2,814,749,767,110,655. The *ms-id* must be the same in all BGWs within the same fabric/site.

	Command or Action	Purpose
Step 3	**split-horizon per-site** **Example:** switch(config-evpn-msite-bgw)# **split-horizon per-site**	Enables to receive packets encapsulated with DCI group from another border gateway on the same site and avoids duplication of packets. **Note** Use this command when DCI multicast underlay is configured on a site with anycast border gateway.
Step 4	**interface nve 1** **Example:** switch(config-evpn-msite-bgw)# **interface nve 1**	Creates a VXLAN overlay interface that terminates VXLAN tunnels. **Note** Only one NVE interface is allowed on the switch.
Step 5	**source-interface loopback** *src-if* **Example:** switch(config-if-nve)# **source-interface loopback 0**	The source interface must be a loopback interface that is configured on the switch with a valid /32 IP address. This /32 IP address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising it through a dynamic routing protocol in the transport network.
Step 6	**host-reachability protocol bgp** **Example:** switch(config-if-nve)# **host-reachability protocol bgp**	Defines BGP as the mechanism for host reachability advertisement.
Step 7	**multisite border-gateway interface loopback** *vi-num* **Example:** switch(config-if-nve)# **multisite border-gateway interface loopback 100**	Defines the loopback interface used for the BGW virtual IP address (VIP). The border-gateway interface must be a loopback interface that is configured on the switch with a valid /32 IP address. This /32 IP address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising it through a dynamic routing protocol in the transport network. This loopback must be different than the source interface loopback. The range of *vi-num* is from 0 to 1023.
Step 8	**no shutdown** **Example:** switch(config-if-nve)# **no shutdown**	Negates the **shutdown** command.
Step 9	**exit** **Example:** switch(config-if-nve)# **exit**	Exits the NVE configuration mode.
Step 10	**interface loopback** *loopback-number* **Example:** switch(config)# **interface loopback 0**	Configures the loopback interface.

	Command or Action	**Purpose**
Step 11	**ip address** *ip-address* **Example:** `switch(config-if)# ip address 198.0.2.0/32`	Configures the IP address for the loopback interface.

Enabling VXLAN EVPN Multi-Site with IPv6 Multicast Underlay

This procedure enables the VXLAN EVPN Multi-Site feature with IPv6 multicast underlay. Multi-Site is enabled on the BGWs only. The site-id must be the same on all BGWs in the fabric/site.

SUMMARY STEPS

1. **configure terminal**
2. **evpn multisite border-gateway** *ms-id*
3. **interface nve 1**
4. **source-interface loopback** *src-if*
5. **host-reachability protocol bgp**
6. **multisite border-gateway interface loopback** *vi-num*
7. (Optional) **multisite virtual-rmac** *mac-address*
8. **member vni** *vni-range*
9. **multisite ingress-replication**
10. **mcast-group** *ipv6-address*
11. **no shutdown**
12. **exit**
13. **interface loopback** *loopback-number*
14. **ipv6 address** *ipv6-address*

DETAILED STEPS

Procedure

	Command or Action	**Purpose**
Step 1	**configure terminal** **Example:** `switch# configure terminal`	Enters global configuration mode.
Step 2	**evpn multisite border-gateway** *ms-id* **Example:** `switch(config)# evpn multisite border-gateway 100`	Configures the site ID for a site/fabric. The range of values for *ms-id* is 1 to 2,814,749,767,110,655. The *ms-id* must be the same in all BGWs within the same fabric/site. **Note** The **mvpn vri id** *id* command is required on BGWs if site-id value is greater than 2 bytes, and this value has to be same across all same site BGWs and unique in TRM

	Command or Action	Purpose
		domain. Also this value must not collide with any site-id value.
Step 3	**interface nve 1** **Example:** switch(config-evpn-msite-bgw)# **interface nve 1**	Creates a VXLAN overlay interface that terminates VXLAN tunnels. **Note** Only one NVE interface is allowed on the switch.
Step 4	**source-interface loopback** *src-if* **Example:** switch(config-if-nve)# **source-interface loopback 0**	The source interface must be a loopback interface that is configured on the switch with a valid /128 IPv6 address. This /128 IPv6 address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising it through a dynamic routing protocol in the transport network.
Step 5	**host-reachability protocol bgp** **Example:** switch(config-if-nve)# **host-reachability protocol bgp**	Defines BGP as the mechanism for host reachability advertisement.
Step 6	**multisite border-gateway interface loopback** *vi-num* **Example:** switch(config-if-nve)# **multisite border-gateway interface loopback 100**	Defines the loopback interface used for the BGW virtual IPv6 address (VIP). The border-gateway interface must be a loopback interface that is configured on the switch with a valid /128 IPv6 address. This /128 IPv6 address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising it through a dynamic routing protocol in the transport network. This loopback must be different than the source interface loopback. The range of *vi-num* is from 0 to 1023.
Step 7	(Optional) **multisite virtual-rmac** *mac-address* **Example:** **switch(config-if-nve)# multisite virtual-rmac 0600.0000.abcd**	For interoperability with other switches, user have to manually configure vMAC on Nexus 9000 switches to override the auto generated vMAC. The default behavior is to auto generate. If manual VMAC is configured, manual vMAC will take precedence. **Note** Only unicast MAC address range is supported for vMAC address configuration.
Step 8	**member vni** *vni-range* **Example:** switch(config-if-nve)# **member vni 50101**	Configures the virtual network identifier (VNI). The range for vni-range is from 1 to 16,777,214. The value of vni-range can be a single value like 5000 or a range like 5001-5008.
Step 9	**multisite ingress-replication** **Example:**	Defines the Multi-Site replication method for extending TRM functionality across sites.

	Command or Action	Purpose
	`switch(config-if-nve-vni)# multisite ingress-replication`	
Step 10	**mcast-group** *ipv6-address* **Example:** `switch(config-if-nve-vni)# mcast-group ff03::101`	Configures the IPv6 Multicast group within the fabric
Step 11	**no shutdown** **Example:** `switch(config-if-nve)# no shutdown`	Negates the **shutdown** command.
Step 12	**exit** **Example:** `switch(config-if-nve)# exit`	Exits the NVE configuration mode.
Step 13	**interface loopback** *loopback-number* **Example:** `switch(config)# interface loopback 0`	Configures the loopback interface.
Step 14	**ipv6 address** *ipv6-address* **Example:** `switch(config-if)# ipv6 address 2001:DB8::11:11:11:11/128`	Configures the IPv6 address for the loopback interface.

Configuring Dual RD Support for Multi-Site

Follow these steps if you need to manually configure the secondary RD value or disable dual RDs.

Before you begin

Enable VXLAN EVPN Multi-Site.

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	
	 Example:
	
	 `switch# configure terminal`
	 `switch(config)#`
		Enters global configuration mode.

Step 2	**router bgp** *as-num*
	
	 Example:
	
	 `switch(config)# router bgp 100`
	 `switch(config-router)#`
		Configures the autonomous system number. The range for *as-num* is from 1 to 4,294,967,295.

	Command or Action	Purpose
Step 3	[**no**] **rd dual id** [*2-bytes*] **Example:** `switch(config-router)# rd dual id 1`	Defines the first 2 bytes of the secondary RD. The ID must be the same across the Multi-Site BGWs. The range is from 1 to 65535. **Note** If necessary, you can use the **no rd dual** command to disable dual RDs and fall back to a single RD.
Step 4	(Optional) **show bgp evi** *evi-id* **Example:** `switch(config-router)# show bgp evi 100`	Displays the secondary RD configured as part of the **rd dual id** [*2-bytes*] command for the specified EVI.

Example

The following example shows sample output for the **show bgp evi** *evi-id* command:

```
switch# show bgp evi 100
----------------------------------------------
  L2VNI ID                  : 100 (L2-100)
  RD                        : 3.3.3.3:32867
  Secondary RD              : 1:100
  Prefixes (local/total)    : 1/6
  Created                   : Jun 23 22:35:13.368170
  Last Oper Up/Down         : Jun 23 22:35:13.369005 / never
  Enabled                   : Yes

  Active Export RT list     :
        100:100
  Active Import RT list     :
        100:100
```

Configuring VNI Dual Mode

This procedure describes the configuration of the BUM traffic domain for a given VLAN. Support exists for using multicast or ingress replication inside the fabric/site and ingress replication across different fabrics/sites.

Note If you have multiple VRFs and only one is extended to ALL leaf switches, you can add a dummy loopback to that one extended VRF and advertise through BGP. Otherwise, you'll need to check how many VRFs are extended and to which switches, and then add a dummy loopback to the respective VRFs and advertise them as well. Therefore, use the **advertise-pip** command to prevent potential user errors in the future.

For more information about configuring multicast or ingress replication for a large number of VNIs, see Example of VXLAN BGP EVPN (eBGP).

Procedure

	Command or Action	**Purpose**
Step 1	**configure terminal** **Example:** switch# **configure terminal**	Enters global configuration mode.
Step 2	**interface nve 1** **Example:** switch(config)# **interface nve 1**	Creates a VXLAN overlay interface that terminates VXLAN tunnels. **Note** Only one NVE interface is allowed on the switch.
Step 3	**member vni** *vni-range* **Example:** switch(config-if-nve)# **member vni 200**	Configures the virtual network identifier (VNI). The range for *vni-range* is from 1 to 16,777,214. The value of *vni-range* can be a single value like 5000 or a range like 5001-5008. **Note** Enter one of the Step 4 or Step 5 commands.
Step 4	**mcast-group** *ip-addr* **Example:** switch(config-if-nve-vni)# **mcast-group 255.0.4.1**	Configures the NVE Multicast group IP prefix within the fabric.
Step 5	**ingress-replication protocol bgp** **Example:** switch(config-if-nve-vni)# **ingress-replication protocol bgp**	Enables BGP EVPN with ingress replication for the VNI within the fabric.
Step 6	**multisite ingress-replication** **Example:** switch(config-if-nve-vni)# **multisite ingress-replication**	Defines the Multi-Site BUM replication method for extending the Layer 2 VNI.

Configuring Fabric/DCI Link Tracking

This procedure describes the configuration to track all DCI-facing interfaces and site internal/fabric facing interfaces. Tracking is mandatory and is used to disable reorigination of EVPN routes either from or to a site if all the DCI/fabric links go down.

Procedure

	Command or Action	Purpose
Step 1	**configure terminal** **Example:** switch# **configure terminal**	Enters global configuration mode.
Step 2	**interface ethernet** *port* **Example:** switch(config)# **interface ethernet1/1**	Enters interface configuration mode for the DCI or fabric interface. **Note** Enter one of the following commands in Step 3 or Step 4.
Step 3	**evpn multisite dci-tracking** **Example:** switch(config-if)# **evpn multisite dci-tracking**	Configures DCI interface tracking.
Step 4	(Optional) **evpn multisite fabric-tracking** **Example:** switch(config-if)# **evpn multisite fabric-tracking**	Configures EVPN Multi-Site fabric tracking. The **evpn multisite fabric-tracking** is mandatory for anycast BGWs and vPC BGW fabric links.
Step 5	**ip address** *ip-addr* | **ipv6 address** *ipv6-addr* **Example:** For IPv4 switch(config-if)# ip address 192.1.1.1 **Example:** For IPv6 switch(config-if)# ipv6 address 2001:DB8::192:1:1:1	Configures the IP or IPv6 address.
Step 6	**no shutdown** **Example:** switch(config-if)# no shutdown	Negates the **shutdown** command.

Configuring Fabric External Neighbors

This procedure describes the configuration of fabric external/DCI neighbors for communication to other site/fabric BGWs.

Procedure

	Command or Action	**Purpose**
Step 1	**configure terminal** **Example:** switch# **configure terminal**	Enters global configuration mode.
Step 2	**router bgp** *as-num* **Example:** switch(config)# **router bgp 100**	Configures the autonomous system number. The range for *as-num* is from 1 to 4,294,967,295.
Step 3	**neighbor** [*ip-addr* | *ipv6-addr*] **Example:** For IPv4 switch(config-router)# **neighbor 100.0.0.1** **Example:** For IPv6 switch(config-router)# **neighbor 2001:DB8::100:0:0:1**	Configures a BGP neighbor.
Step 4	**remote-as** *value* **Example:** switch(config-router-neighbor)# **remote-as 69000**	Configures remote peer's autonomous system number.
Step 5	**peer-type fabric-external** **Example:** switch(config-router-neighbor)# **peer-type fabric-external**	Enables the next hop rewrite for Multi-Site. Defines site external BGP neighbors for EVPN exchange. The default for **peer-type** is **fabric-internal**. **Note** The **peer-type fabric-external** command is required only for VXLAN Multi-Site BGWs. It is not required for pseudo BGWs.
Step 6	**address-family l2vpn evpn** **Example:** switch(config-router-neighbor)# **address-family l2vpn evpn**	Configures the address family Layer 2 VPN EVPN under the BGP neighbor.
Step 7	**rewrite-evpn-rt-asn** **Example:** switch(config-router-neighbor)# **rewrite-evpn-rt-asn**	Rewrites the route target (RT) information to simplify the MAC-VRF and IP-VRF configuration. BGP receives a route, and as it processes the RT attributes, it checks if the AS value matches the peer AS that is sending that route and replaces it. Specifically, this command changes the incoming route target's AS number to match the BGP-configured neighbor's remote AS number. You can see the modified RT value in the receiver router.

Configuring VXLAN EVPN Multi-Site Storm Control

VXLAN EVPN Multi-Site Storm Control allows rate limiting of multidestination (BUM) traffic on Multi-Site BGWs. You can control BUM traffic sent over the DCI link using a policer on fabric links in the ingress direction.

Remote peer reachability must be only through DCI links. Appropriate routing configuration must ensure that remote site routes are not advertised over Fabric links.

Multicast traffic is policed only on DCI interfaces, while unknown unicast and broadcast traffic is policed on both DCI and fabric interfaces.

Cisco NX-OS Release 9.3(6) and later releases optimize rate granularity and accuracy. Bandwidth is calculated based on the accumulated DCI uplink bandwidth, and only interfaces tagged with DCI tracking are considered. (Prior releases also include fabric-tagged interfaces.) In addition, granularity is enhanced by supporting two digits after the decimal point. These enhancements apply to the Cisco Nexus 9300-EX, 9300-FX/FX2/FX3, and 9300-GX platform switches.

Beginning with Cisco NX-OS Release 10.5(2)F, VXLAN EVPN Multi-Site Storm Control is supported on Cisco Nexus 9500 Series switches with N9K-X9736C-FX3 line card.

Note For information on access port storm control, see the Cisco Nexus 9000 Series NX-OS Layer 2 Configuration Guide.

SUMMARY STEPS

1. **configure terminal**
2. [**no**] **evpn storm-control** {**broadcast** | **multicast** | **unicast**} {**level** *level*}

DETAILED STEPS

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	
	Example:
	
	```switch# configure terminal```
	```switch(config)#```
		Enters global configuration mode.

Step 2	[**no**] **evpn storm-control** {**broadcast**	**multicast**	**unicast**} {**level** *level*}
			
			Example:
			
			```switch(config)# evpn storm-control unicast```
			``` level 10```
			
			Example:
				Configures the storm suppression level as a number from 0–100.
				
				0 means that all traffic is dropped, and 100 means that all traffic is allowed. For any value in between, the unknown unicast traffic rate is restricted to a percentage of available bandwidth. For example, a value of 10 means that the traffic rate is restricted to 10% of the available bandwidth, and anything above that rate is dropped.

Command or Action	Purpose
`switch(config)#` **evpn storm-control unicast level 10.20**	Beginning with Cisco NX-OS Release 9.3(6), you can configure the level as a fractional value by adding two digits after the decimal point. For example, you can enter a value of 10.20.

Verifying VXLAN EVPN Multi-Site Storm Control

To display EVPN storm control setting information, enter the following command:

Command	Purpose
slot 1 show hardware vxlan storm-control	Displays the status of EVPN storm control setting.

Note Once the Storm control hits the threshold, a message is logged as stated below:

```
BGWY-1 %ETHPORT-5-STORM_CONTROL_ABOVE_THRESHOLD: Traffic in port Ethernet1/32 exceeds the
configured threshold , action - Trap (message repeated 38 times)
```

Multi-Site with vPC Support

About Multi-Site with vPC Support

The BGWs can be in a vPC complex. In this case, it is possible to support dually-attached directly-connected hosts that might be bridged or routed as well as dually-attached firewalls or service attachments. The vPC BGWs have vPC-specific multihoming techniques and do not rely on EVPN Type 4 routes for DF election or split horizon.

Guidelines and Limitations for Multi-Site with vPC Support

Multi-Site with vPC support has the following configuration guidelines and limitations:

- 4000 VNIs for vPC are not supported.

- For BUM with continued VIP use, the MCT link is used as transport upon core isolation or fabric isolation, and for unicast traffic in fabric isolation.

- Beginning with Cisco NX-OS Release 10.1(2), TRM Multisite with vPC BGW is supported.

- The routes to remote Multisite BGW loopback addresses must always prioritize the DCI link path over the iBGP protocol between vPC Border Gateway switches configured using the backup SVI. The backup SVI should be used strictly in the event of a DCI link failure.

- vPC BGWs are not supported with IPv6 multicast underlay.

Configuring Multi-Site with vPC Support

This procedure describes the configuration of Multi-Site with vPC support:

- Configure vPC domain.

- Configure port channels.

- Configuring vPC Peer Link.

Procedure

	Command or Action	**Purpose**
Step 1	**configure terminal** **Example:** switch# **configure terminal**	Enters global configuration mode.
Step 2	**feature vpc** **Example:** switch(config)# **feature vpc**	Enables vPCs on the device.
Step 3	**feature interface-vlan** **Example:** switch(config)# **feature interface-vlan**	Enables the interface VLAN feature on the device.
Step 4	**feature lacp** **Example:** switch(config)# **feature lacp**	Enables the LACP feature on the device.
Step 5	**feature pim** **Example:** switch(config)# **feature pim**	Enables the PIM feature on the device.
Step 6	**feature ospf** **Example:** switch(config)# **feature ospf**	Enables the OSPF feature on the device.
Step 7	**ip pim rp-address** *address* **group-list** *range* **Example:** switch(config)# **ip pim rp-address 100.100.100.1 group-list 224.0.0/4**	Defines a PIM RP address for the underlay multicast group range.
Step 8	**vpc domain** *domain-id* **Example:** switch(config)# **vpc domain 1**	Creates a vPC domain on the device and enters vpn-domain configuration mode for configuration purposes. There is no default. The range is from 1 to 1000.

	Command or Action	Purpose
Step 9	**peer switch** **Example:** switch(config-vpc-domain)# **peer switch**	Defines the peer switch.
Step 10	**peer gateway** **Example:** switch(config-vpc-domain)# **peer gateway**	Enables Layer 3 forwarding for packets destined to the gateway MAC address of the vPC.
Step 11	**peer-keepalive destination** *ip-address* **Example:** switch(config-vpc-domain)# **peer-keepalive destination 172.28.230.85**	Configures the IPv4 address for the remote end of the vPC peer-keepalive link. **Note** The system does not form the vPC peer link until you configure a vPC peer-keepalive link. The management ports and VRF are the defaults.
Step 12	**ip arp synchronize** **Example:** switch(config-vpc-domain)# **ip arp synchronize**	Enables IP ARP synchronize under the vPC domain to facilitate faster ARP table population following device reload.
Step 13	**ipv6 nd synchronize** **Example:** switch(config-vpc-domain)# **ipv6 nd synchronize**	Enables IPv6 ND synchronization under the vPC domain to facilitate faster ND table population following device reload.
Step 14	Create the vPC peer-link. **Example:** switch(config)# **interface port-channel 1** switch(config)# **switchport** switch(config)# **switchport mode trunk** switch(config)# **switchport trunk allowed vlan 1,10,100-200** switch(config)# **mtu 9216** switch(config)# **vpc peer-link** switch(config)# **no shut** switch(config)# **interface Ethernet 1/1, 1/21** switch(config)# **switchport** switch(config)# **mtu 9216** switch(config)# **channel-group 1 mode active** switch(config)# **no shutdown**	Creates the vPC peer-link port-channel interface and adds two member interfaces to it.
Step 15	**system nve infra-vlans** *range* **Example:** switch(config)# **system nve infra-vlans 10**	Defines a non-VXLAN-enabled VLAN as a backup routed path.
Step 16	**vlan** *number* **Example:** switch(config)# **vlan 10**	Creates the VLAN to be used as an infra-VLAN.

	Command or Action	**Purpose**
Step 17	Create the SVI. **Example:** `switch(config)# `**`interface vlan 10`** `switch(config)# `**`ip address 10.10.10.1/30`** `switch(config)# `**`ip router ospf process UNDERLAY area 0`** `switch(config)# `**`ip pim sparse-mode`** `switch(config)# `**`no ip redirects`** `switch(config)# `**`mtu 9216`** `switch(config)# `**`no shutdown`**	Creates the SVI used for the backup routed path over the vPC peer-link.
Step 18	(Optional) **delay restore interface-vlan** *seconds* **Example:** `switch(config-vpc-domain)# `**`delay restore interface-vlan 45`**	Enables the delay restore timer for SVIs. We recommend tuning this value when the SVI/VNI scale is high. For example, when the SCI count is 1000, we recommend that you set the delay restore to 45 seconds.
Step 19	**evpn multisite border-gateway** *ms-id* **Example:** `switch(config)# `**`evpn multisite border-gateway 100`**	Configures the site ID for a site/fabric. The range of values for *ms-id* is 1 to 281474976710655. The *ms-id* must be the same in all BGWs within the same fabric/site.
Step 20	**interface nve 1** **Example:** `switch(config-evpn-msite-bgw)# `**`interface nve 1`**	Creates a VXLAN overlay interface that terminates VXLAN tunnels. **Note** Only one NVE interface is allowed on the switch.
Step 21	**source-interface loopback** *src-if* **Example:** `switch(config-if-nve)# `**`source-interface loopback 0`**	Defines the source interface, which must be a loopback interface with a valid /32 IP address. This /32 IP address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising the address through a dynamic routing protocol in the transport network.
Step 22	**host-reachability protocol bgp** **Example:** `switch(config-if-nve)# `**`host-reachability protocol bgp`**	Defines BGP as the mechanism for host reachability advertisement.
Step 23	**multisite border-gateway interface loopback** *vi-num* **Example:** `switch(config-if-nve)# `**`multisite border-gateway interface loopback 100`**	Defines the loopback interface used for the BGW virtual IP address (VIP). The BGW interface must be a loopback interface that is configured on the switch with a valid /32 IP address. This /32 IP address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising the address through a dynamic routing protocol in the transport network. This loopback must be different than the source interface loopback. The range of *vi-num* is from 0 to 1023.

	Command or Action	**Purpose**
Step 24	**no shutdown** **Example:** switch(config-if-nve)# **no shutdown**	Negates the **shutdown** command.
Step 25	**exit** **Example:** switch(config-if-nve)# **exit**	Exits the NVE configuration mode.
Step 26	**interface loopback** *loopback-number* **Example:** switch(config)# **interface loopback 0**	Configures the loopback interface.
Step 27	**ip address** *ip-address* **Example:** switch(config-if)# **ip address 198.0.2.0/32**	Configures the primary IP address for the loopback interface.
Step 28	**ip address** *ip-address* **secondary** **Example:** switch(config-if)# **ip address 198.0.2.1/32 secondary**	Configures the secondary IP address for the loopback interface.
Step 29	**ip pim sparse-mode** **Example:** switch(config-if)# **ip pim sparse-mode**	Configures PIM sparse mode on the loopback interface.

Verifying the Multi-Site with vPC Support Configuration

To display Multi-Site with vPC support information, enter one of the following commands:

show vpc brief	Displays general vPC and CC status.
show vpc consistency-parameters global	Displays the status of those parameters that must be consistent across all vPC interfaces.
show vpc consistency-parameters vni	Displays configuration information for VNIs under the NVE interface that must be consistent across both vPC peers.

Output example for the **show vpc brief** command:

```
switch# show vpc brief
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link

vPC domain id                    : 1
Peer status                      : peer adjacency formed ok     (<--- peer up)
vPC keep-alive status            : peer is alive
Configuration consistency status : success (<----- CC passed)
```

```
Per-vlan consistency status        : success                         (<---- per-VNI CCpassed)
Type-2 consistency status          : success
vPC role                           : secondary
Number of vPCs configured          : 1
Peer Gateway                       : Enabled
Dual-active excluded VLANs         : -
Graceful Consistency Check         : Enabled
Auto-recovery status               : Enabled, timer is off.(timeout = 240s)
Delay-restore status               : Timer is off.(timeout = 30s)
Delay-restore SVI status           : Timer is off.(timeout = 10s)
Operational Layer3 Peer-router     : Disabled
[...]
```

Output example for the **show vpc consistency-parameters global** command:

```
switch# show vpc consistency-parameters global

    Legend:
        Type 1 : vPC will be suspended in case of mismatch

Name                        Type  Local Value           Peer Value
-------------               ----  --------------------  ----------------------
[...]
Nve1 Adm St, Src Adm St,    1     Up, Up, 2.1.44.5, CP, Up, Up, 2.1.44.5, CP,
Sec IP, Host Reach, VMAC          TRUE, Disabled,       TRUE, Disabled,
Adv, SA,mcast l2, mcast           0.0.0.0, 0.0.0.0,     0.0.0.0, 0.0.0.0,
l3, IR BGP,MS Adm St, Reo          Disabled, Up,         Disabled, Up,
                                  200.200.200.200       200.200.200.200

[...]
```

Output example for the **show vpc consistency-parameters vni** command:

```
switch(config-if-nve-vni)# show vpc consistency-parameters vni

    Legend:
        Type 1 : vPC will be suspended in case of mismatch

Name                        Type  Local Value           Peer Value
-------------               ----  --------------------  ----------------------
Nve1 Vni, Mcast, Mode,      1     11577, 234.1.1.1,     11577, 234.1.1.1,
Type, Flags                       Mcast, L2, MS IR      Mcast, L2, MS IR
Nve1 Vni, Mcast, Mode,      1     11576, 234.1.1.1,     11576, 234.1.1.1,
Type, Flags                       Mcast, L2, MS IR      Mcast, L2, MS IR
[...]
```

Configuration Example for Multi-Site with Asymmetric VNIs

The following example shows how two sites with different sets of VNIs can connect to the same MAC VRF or IP VRF. One site uses VNI 200 internally, and the other site uses VNI 300 internally. Route-target auto no longer matches because the VNI values are different. Therefore, the route-target values must be manually configured. In this example, the value 222:333 stitches together the two VNIs from different sites.

The BGW of site 1 has L2VNI 200 and L3VNI 201.

The BGW of site 2 has L2VNI 300 and L3VNI 301.

Note   This configuration example assumes that basic Multi-Site configurations are already in place.

Note You must have VLAN-to-VRF mapping on the BGW. This requirement is necessary to maintain L2VNI-to-L3VNI mapping, which is needed for reorigination of MAC-IP routes at BGWs.

Layer 3 Configuration

In the BGW node of site 1, configure the common RT 201:301 for stitching the two sites using L3VNI 201 and L3VNI 301:

```
vrf context vni201
  vni 201
  address-family ipv4 unicast
    route-target both auto evpn
    route-target import 201:301 evpn
    route-target export 201:301 evpn
```

In the BGW node of site 2, configure the common RT 201:301 for stitching the two sites using L3VNI 201 and L3VNI 301:

```
vrf context vni301
  vni 301
  address-family ipv4 unicast
    route-target both auto evpn
    route-target import 201:301 evpn
    route-target export 201:301 evpn
```

Layer 2 Configuration

In the BGW node of site 1, configure the common RT 222:333 for stitching the two sites using L2VNI 200 and L2VNI 300:

```
evpn
  vni 200 l2
    rd auto
    route-target import auto
    route-target import 222:333
    route-target export auto
    route-target export 222:333
```

For proper reorigination of L3 labels of MAC-IP routes, associate the VRF (L3VNI) to the L2VNI:

```
interface Vlan 200
  vrf member vni201
```

In the BGW node of site 2, configure the common RT 222:333 for stitching the two sites using L2VNI 200 and L2VNI 300:

```
evpn
  vni 300 l2
    rd auto
    route-target import auto
    route-target import 222:333
    route-target export auto
    route-target export 222:333
```

For proper reorigination of L3 labels of MAC-IP routes, associate the VRF (L3VNI) to the L2VNI:

```
interface vlan 300
  vrf member vni301
```

Advertise Using PIP Towards Fabric

Beginning with Cisco NX-OS Release 10.5(1)F, you can configure BGW to advertise external EVPN type-5 routes with PIP as next-hop and PIP's RMAC towards fabric side. With this configuration, BGW uses PIP instead of VIP for route advertisement.

Guidelines and Limitations

Following guidelines and limitations are applicable when you advertise route using PIP:

- Only L3 support is added in Cisco NX-OS Release 10.5(1)F.

- This feature is not applicable on VPC BGW.

- With this solution, traffic loss is expected until routes are updated on BGW and Leaf after remote BGW goes down.

- User must configure **maximum-paths** under EVPN and VRF's address-family on the leaf. This enables BGP to select all the paths as best-path or multi-paths and download all next-hops to forwarding plane to achieve load-balancing.

- In topologies with separate BGW and Spine, user must do one of the following 2 options:

    - Disable dual RD on BGWs.

    - configure add-path command in Spine to advertise all the EVPN paths to Leaf switches if dual RD is enabled on BGWs.

- The **fabric-advertise-pip l3** command has to be configured on all the BGWs on the same site.

- This solution is applicable to multiplane topology with only one BGW per plane per site. If more than one BGW per site is connected to a single plane, this solution is not required.

- When **fabric-advertise-pip l3** is enabled, BGWs accept remote type-5 routes from other BGWs on the same site with their PIP addresses. This causes increase in the number of paths per route on BGWs and is directly proportional to the number of BGWs on same site.

Configuring BGW to Advertise using PIP Towards Fabric

This section provides the configuration procedure for enabling advertise of remote routes with PIP as next-hop towards fabric on an Anycast BGW.

SUMMARY STEPS

1. **configure terminal**
2. **evpn multisite border-gateway** *ms-id*
3. **fabric-advertise-pip l3**

DETAILED STEPS

Procedure

	Command or Action	**Purpose**
Step 1	**configure terminal** **Example:** `switch# configure terminal`	Enters global configuration mode.
Step 2	**evpn multisite border-gateway** *ms-id* **Example:** `switch(config)# evpn multisite border-gateway 100`	Configures the site ID for a site/fabric. The range of values for ms-id is 1 to 2,814,749,767,110,655. The ms-id must be the same in all BGWs within the same fabric/site.
Step 3	**fabric-advertise-pip l3** **Example:** `switch(config-evpn-msite-bgw)# fabric-advertise-pip l3`	Enables advertisement of remote EVPN type-5 routes with PIP next-hop towards fabric.

Verifying the Configuration

Use the **show nve interface nve 1 detail** command to verify the configuration:

```
switch(config)# show nve interface nve 1 detail
Interface: nve1, State: Up, encapsulation: VXLAN
VPC Capability: VPC-VIP-Only [not-notified]
Local Router MAC: 4464.3c31.802f
Host Learning Mode: Control-Plane
Source-Interface: loopback1 (primary: 20:1::21, secondary: 0.0.0.0)
Source Interface State: Up
Virtual RMAC Advertisement: No
NVE Flags:
Interface Handle: 0x49000001
Source Interface hold-down-time: 180
Source Interface hold-up-time: 30
Remaining hold-down time: 0 seconds
Virtual Router MAC: N/A
Virtual Router MAC Re-origination: 0022.3344.5566
Interface state: nve-intf-add-complete
Fabric convergence time: 37 seconds
Fabric convergence time left: 0 seconds
Multisite delay-restore time: 50 seconds
Multisite delay-restore time left: 0 seconds
Multisite dci-advertise-pip configured: False
Multisite fabric-advertise-pip l3 configured: True
```

TRM with Multi-Site

This section contains the following topics:

Information About Configuring TRM with Multi-Site

Tenant Routed Multicast (TRM) with Multi-Site enables multicast forwarding across multiple VXLAN EVPN fabrics that are connected via Multi-Site. This feature provides Layer 3 multicast services across sites for sources and receivers across different sites. It addresses the requirement of East-West multicast traffic between sites.

Each TRM site is operating independently. Border gateways on each site allow stitching across the sites. There can be multiple border gateways for each site. Multicast source and receiver information across sites is propagated by BGP on the border gateways that are configured with TRM. The border gateway on each site receives the multicast packet and re-encapsulates the packet before sending it to the local site. Beginning with Cisco NX-OS Release 10.1(2), TRM with Multi-Site supports both Anycast Border Gateway and vPC Border Gateway.

The border gateway that is elected as Designated Forwarder (DF) for the L3VNI forwards the traffic from fabric toward the core side. In the TRM Multicast-Anycast Gateway model, we use the VIP-R based model to send traffic toward remote sites. The IR destination IP is the VIP-R of the remote site. Each site that has the receiver gets only one copy from the source site. DF forwarding is applicable only on Anycast Border Gateways.

Note Only the DF sends the traffic toward remote sites.

On the remote site, the BGW that receives the inter-site multicast traffic from the core forwards the traffic toward the fabric side. The DF check is not done from the core to fabric direction because non-DF can also receive the VIP-R copy from the source site.

Figure 2: TRM with Multi-Site Topology, BL External Multicast Connectivity
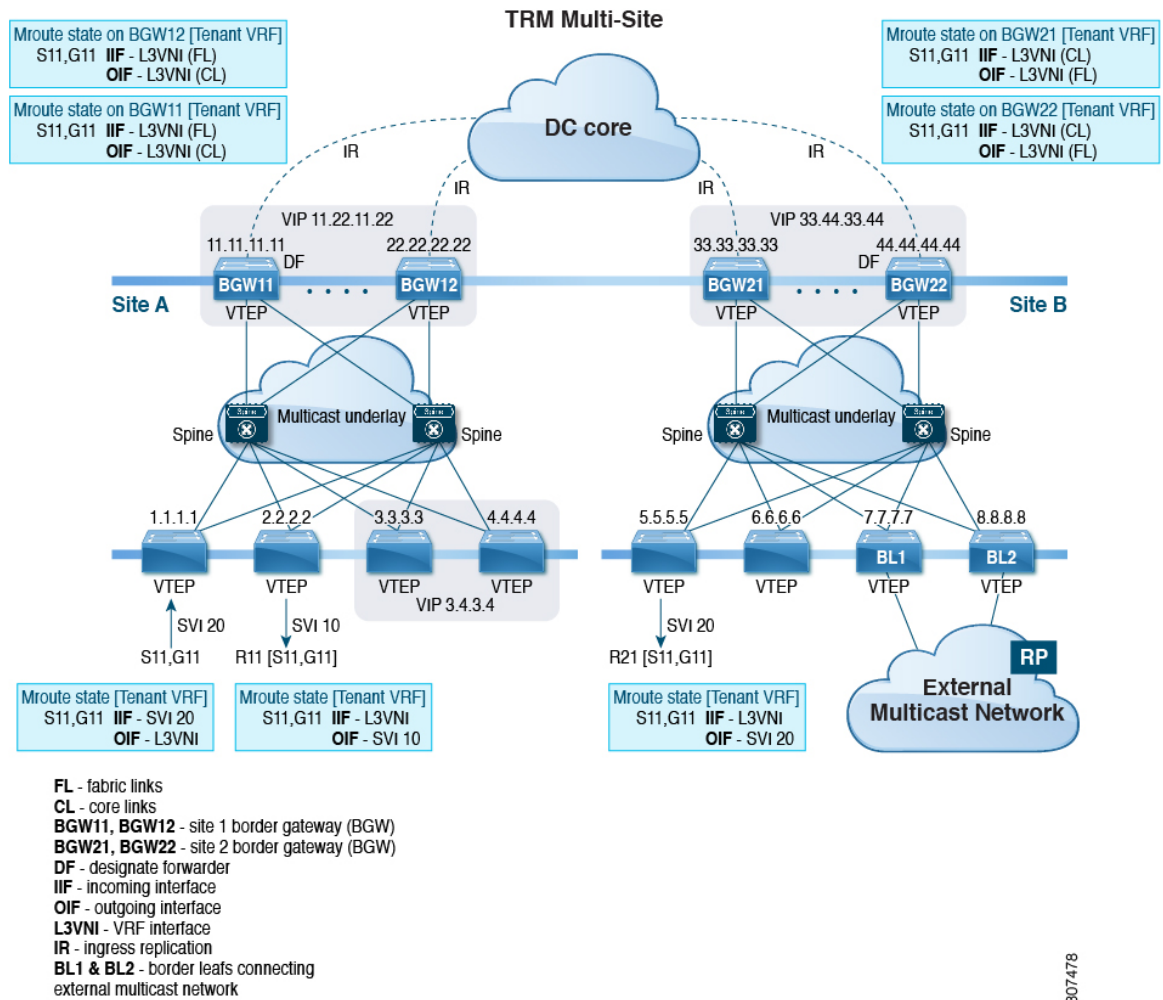
Beginning with Cisco NX-OS Release 9.3(3), TRM with Multi-Site supports BGW connections to the external multicast network in addition to the BL connectivity, which is supported in previous releases. Forwarding occurs as documented in the previous example, except the exit point to the external multicast network can optionally be provided through the BGW.

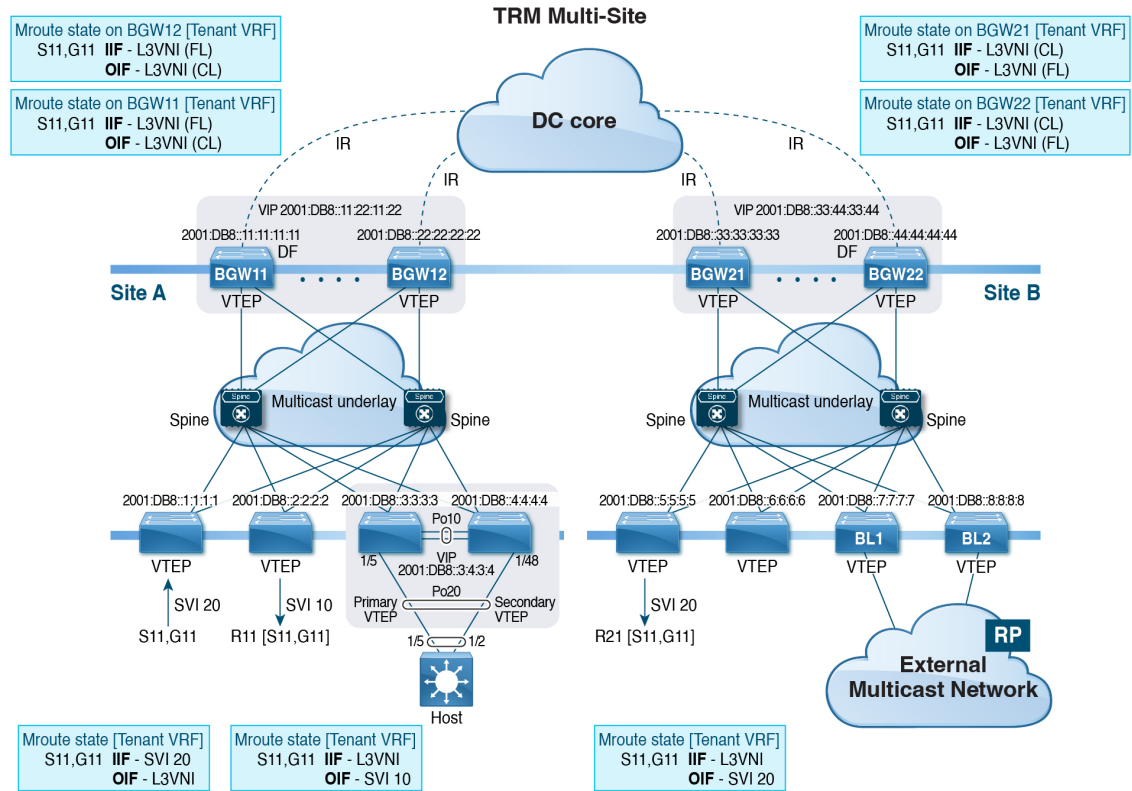*Figure 3: TRM with Multi-Site Topology, BGW External Multicast Connectivity*

Mroute state on BGW12 [Tenant VRF]
S11,G11 IIF - L3VNI (FL)
          OIF - L3VNI (CL)

Mroute state on BGW11 [Tenant VRF]
S11,G11 IIF - L3VNI (FL)
          OIF - L3VNI (CL)

Mroute state on BGW21 [Tenant VRF]
S11,G11 IIF - L3VNI (CL)
          OIF - L3VNI (FL)

Mroute state on BGW22 [Tenant VRF]
S11,G11 IIF - L3VNI (CL)
          OIF - L3VNI (FL)

TRM Multi-Site

DC core

IR

External Multicast Network

RP

VIP 11.22.11.22
11.11.11.11
DF
BGW11
VTEP
Site A

22.22.22.22
BGW12
VTEP

VIP 33.44.33.44
33.33.33.33
BGW+BL
VTEP

44.44.44.44
DF
BGW+BL
VTEP
Site B

Spine Multicast underlay Spine

Spine Multicast underlay Spine

1.1.1.1 2.2.2.2 3.3.3.3 4.4.4.4
VTEP VTEP VTEP VTEP
VIP 3.4.3.4

5.5.5.5 6.6.6.6 7.7.7.7 8.8.8.8
VTEP VTEP VTEP VTEP

SVI 20 SVI 10
S11,G11 R11 [S11,G11]

SVI 20
R21 [S11,G11]

Mroute state [Tenant VRF]
S11,G11 IIF - SVI 20
          OIF - L3VNI

Mroute state [Tenant VRF]
S11,G11 IIF - L3VNI
          OIF - SVI 10

Mroute state [Tenant VRF]
S11,G11 IIF - L3VNI
          OIF - SVI 20

FL - fabric links
CL - core links
BGW11, BGW12 - site 1 border gateway (BGW)
BGW21, BGW22 - site 2 border gateways
  connecting external multicast network
DF - designate forwarder
IIF - incoming interface
OIF - outgoing interface
L3VNI - VRF interface
IR - ingress replication
BL1 & BL2 - border leafs

307659

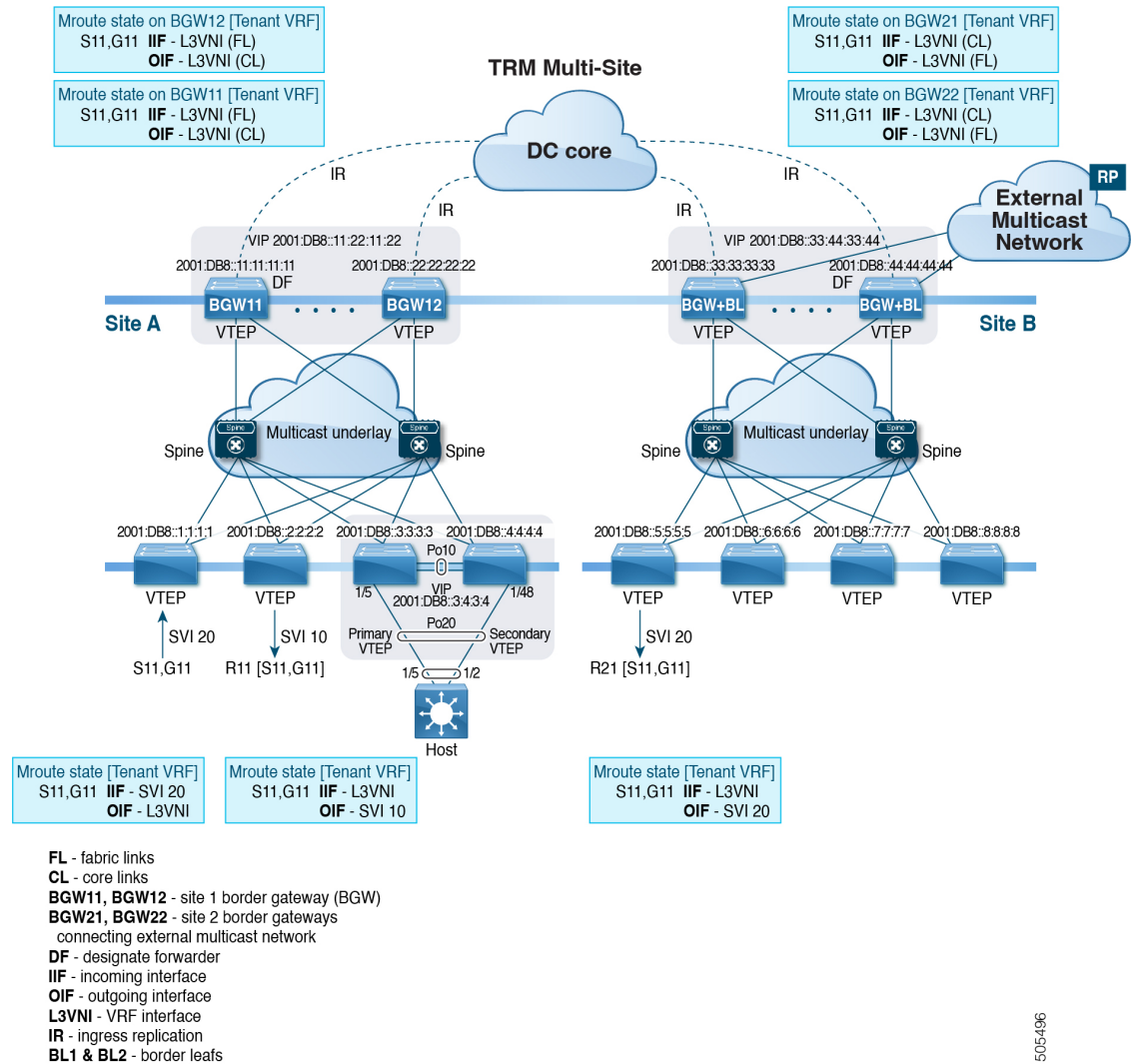# Information About Configuring TRM Multi-Site with IPv6 Underlay

Beginning with Cisco NX-OS Release 10.4(3)F, the support is provided for TRM Multi-Site with IPv6 Underlay.

Figure 4: TRM Multi-Site with IPv6 Underlay Topology, BL External Multicast Connectivity

FL - fabric links
CL - core links
BGW11, BGW12 - site 1 border gateway (BGW)
BGW21, BGW22 - site 2 border gateway (BGW)
DF - designate forwarder
IIF - incoming interface
OIF - outgoing interface
L3VNI - VRF interface
IR - ingress replication
BL1 & BL2 - border leafs connecting
external multicast network

505494

Figure 5: TRM Multi-Site with IPv6 Underlay Topology, BGW External Multicast Connectivity

FL - fabric links
CL - core links
BGW11, BGW12 - site 1 border gateway (BGW)
BGW21, BGW22 - site 2 border gateways
  connecting external multicast network
DF - designate forwarder
IIF - incoming interface
OIF - outgoing interface
L3VNI - VRF interface
IR - ingress replication
BL1 & BL2 - border leafs

The above topology shows four leafs and two spines in the VXLAN EVPN fabric and two Anycast BGWs. Inside the fabric, the underlay is an IPv6 Multicast running PIMv6. RP is positioned in the spine with anycast RP. BGWs support VXLAN with IPv6 Protocol-Independent Multicast (PIMv6) Any-Source Multicast (ASM) on the fabric side and Ingress Replication (IPv6) on the DCI side.

Beginning with Cisco NX-OS Release 10.5(1)F, the underlay network supports the following combinations for TRM Multi-Site:

- In the data center fabric, both Multicast Underlay (PIMv6) Any-Source Multicast (ASM) and Ingress Replication (IPv6) are supported.

- In the Data Center Interconnect (DCI), only Ingress Replication (IPv6) is supported.

Guidelines and Limitations for TRM with Multi-Site

TRM with Multi-Site has the following guidelines and limitations:

- The following platforms support TRM with Multi-Site:
    - Cisco Nexus 9300-EX platform switches
    - Cisco Nexus 9300-FX/FX2/FX3 platform switches
    - Cisco Nexus 9300-GX platform switches
    - Cisco Nexus 9300-GX2 platform switches
    - Cisco Nexus 9332D-H2R switches
    - Cisco Nexus 93400LD-H1 switches
    - Cisco Nexus 9364C-H1 switches
    - Cisco Nexus 9500 platform switches with -EX/FX/FX3 line cards

- Beginning with Cisco NX-OS Release 9.3(3), a border leaf and Multi-Site border gateway can coexist on the same node for multicast traffic.

- Beginning with Cisco NX-OS Release 9.3(3), all border gateways for a given site must run the same Cisco NX-OS 9.3(x) image.

- Cisco NX-OS Release 10.1(2) has the following guidelines and limitations:
    - You need to add a VRF lite link (per Tenant VRF) between the vPC peers in order to support the L3 hosts attached to the vPC primary and secondary peers.
    - Backup SVI is needed between the two vPC peers.
    - Orphan ports attached with L2 and L3 are supported with vPC BGW.
   - TRM multi-site with vPC BGW is not supported with vMCT.

    For details on TRM and Configuring TRM with vPC Support, see Configuring Tenant Routed Multicast.

- TRM multi-site with vPC BGW and with Anycast BGW are supported on Cisco Nexus 9300-EX, FX, FX2, and FX3 family switches. Beginning with Cisco NX-OS Release 10.2(1)F, TRM with vPC BGW and with Anycast BGW are supported on Cisco Nexus 9300-GX family switches.

- Beginning with Cisco NX-OS Release 10.2(1q)F, TRM with Multi-Site is supported on the Cisco Nexus N9K-C9332D-GX2B platform switches.

- Beginning with Cisco NX-OS Release 10.2(1q)F, the TRM multi-site with vPC BGW and with Anycast BGW are supported on the Cisco Nexus C9332D-GX2B platform switches.

- Beginning with Cisco NX-OS Release 10.4(1)F, the TRM multi-site with vPC BGW and with Anycast BGW are supported on the Cisco Nexus 9332D-H2R switches.

- Beginning with Cisco NX-OS Release 10.4(2)F, the TRM multi-site with vPC BGW and with Anycast BGW are supported on the Cisco Nexus 93400LD-H1 switches.

- Beginning with Cisco NX-OS Release 10.4(3)F, the TRM multi-site with vPC BGW and with Anycast BGW are supported on the Cisco Nexus 9364C-H1 switches.

- Beginning with Cisco NX-OS Release 10.5(2)F, the TRM multi-site with vPC BGW and with Anycast BGW are supported on Cisco Nexus 9500 Series switches with N9K-X9736C-FX3 line card.

- Beginning with Cisco NX-OS Release 10.2(2)F, multicast group configuration is used to encapsulate TRM and L2 BUM packets in the DCI core using the **multisite mcast-group** *dci-core-group* command.

- Beginning with Cisco NX-OS Release 10.2(3)F, the TRM multi-site is supported on the Cisco Nexus 9364D-GX2A and 9348D-GX2A switches.

- Beginning with Cisco NX-OS Release 10.4(1)F, the TRM multi-site is supported on the Cisco Nexus 9332D-H2R switches.

- Beginning with Cisco NX-OS Release 10.4(2)F, the TRM multi-site is supported on the Cisco Nexus 93400LD-H1 switches.

- Beginning with Cisco NX-OS Release 10.4(3)F, the TRM multi-site is supported on the Cisco Nexus 9364C-H1 switches.

- TRM with Multi-Site supports the following features:

    - TRM Multi-Site with vPC Border Gateway.

    - PIM ASM multicast underlay in the VXLAN fabric

    - TRM with Multi-Site Layer 3 mode only

    - TRM with Multi-Site with Anycast Gateway

    - Terminating VRF-lite at the border leaf

    - The following RP models with TRM Multi-Site:

        - External RP

        - RP Everywhere

        - Internal RP

- Only one pair of vPC BGW can be configured on one site.

- A pair of vPC BGW and Anycast BGW cannot co-exist on the same site.

- Prior to NX-OS 10.2(2)F only ingress replication was supported between DCI peers across the core. Beginning with Cisco NX-OS Release 10.2(2)F both ingress replication and multicast are supported between DCI peers across the core.

- Border routers reoriginate MVPN routes from fabric to core and from core to fabric.

- Only eBGP peering between border gateways of different sites is supported.

- Each site must have a local RP for the TRM underlay.

- Keep each site's underlay unicast routing isolated from another site's underlay unicast routing. This requirement also applies to Multi-Site.

- MVPN address family must be enabled between BGWs.

- When configuring BGW connections to the external multicast fabric, be aware of the following:

    - The multicast underlay must be configured between all BGWs on the fabric side even if the site doesn't have any leafs in the fabric site.

- Sources and receivers that are Layer-3 attached through VRF-Lite links to the BGW of a single site acting therefore also as Border Leaf (BL) node need to have reachability through the external Layer-3 network. If there's a Layer-3 attached source on BGW BL Node-1 and a Layer-3 attached receiver on BGW BL Node-2 for the same site, the traffic between these two endpoints flows through the external Layer-3 network and not through the fabric.

- External multicast networks should be connected only through the BGW or BL. If a deployment requires external multicast network connectivity from both the BGW and BL at the same site, make sure that external routes that are learned from the BGW are preferred over the BL. To do so, the BGW must have a lower MED and a higher OSPF cost (on the external links) than the BL.

The following figure shows a site with external network connectivity through BGW-BLs and an internal leaf (BL1). The path to the external source should be through BGW-1 (rather than through BL1) to avoid duplication on the remote site receiver.
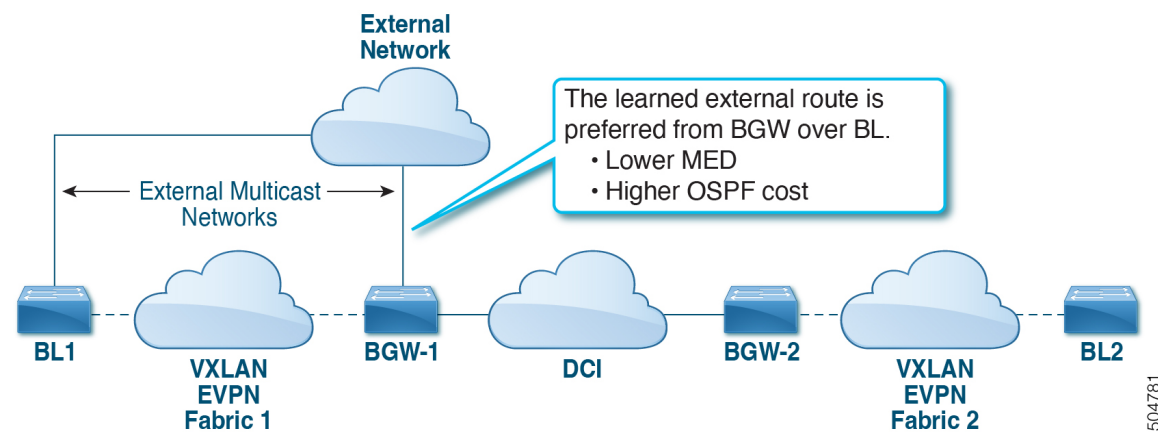
- The BGW supports VRF-lite hand-off and Multi-site configuration on the same physical interface as shown in the diagram.
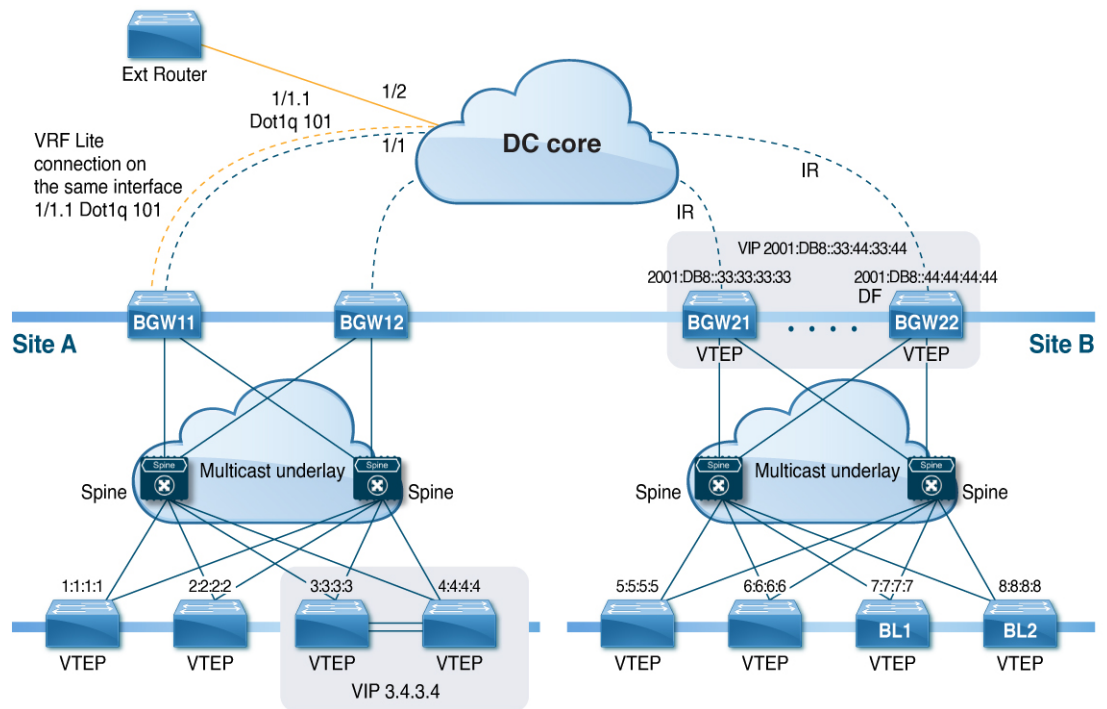
- MED is supported for iBGP only.

Guidelines and Limitations for TRM Multi-Site with IPv6 Underlay

TRM Multi-Site with IPv6 Underlay has the following configuration guidelines and limitations:

- BGWs will support VXLAN with Protocol-Independent Multicast (PIMv6) Any-Source Multicast (ASM) on the fabric side and Ingress Replication (IPv6) on the DCI side.

- Cisco Nexus 9300-FX, FX2, FX3, GX, GX2, H2R and H1 ToR switches are supported as the leaf VTEP.

- Cisco Nexus N9K-X9716D-GX and N9K-X9736C-FX, N9K-X9736C-FX3 line cards are supported only on the spine (EoR).

- When an EoR is deployed as a spine node with Multicast Underlay (PIMv6) Any-Source Multicast (ASM), it is mandatory to configure non-default template using one of the following commands in global configuration mode:

  - **system routing template-multicast-heavy**

  - **system routing template-multicast-ext-heavy**

- Beginning with Cisco NX-OS Release 10.5(1)F, TRM Multi-Site in the data center fabric supports both Multicast Underlay (PIMv6) Any-Source Multicast (ASM) and Ingress Replication (IPv6) in the underlay. This support is available on the following switches and line cards:

  - Cisco Nexus 9300-FX, FX2, FX3, GX, GX2, H2R, and H1 ToR switches as the leaf VTEPs.

  - Cisco Nexus N9K-X9716D-GX and N9K-X9736C-FX line cards as spines if the underlay is configured for Multicast Underlay (PIMv6) Any-Source Multicast (ASM).

> • Cisco Nexus N9K-X9716D-GX and N9K-X9736C-FX line cards as VTEPs if the underlay uses Ingress Replication (IPv6).

> • Beginning with Cisco NX-OS Release 10.5(2)F, TRM Multi-Site with IPv6 Underlay support is extended on Cisco Nexus 9500 Series switches with N9K-X9736C-FX3 line cards as VTEPs if the underlay uses Ingress Replication (IPv6).

Configuring TRM with Multi-Site

Before you begin

The following must be configured:

> • VXLAN TRM

> • VXLAN Multi-Site

This section provides the configuration procedure for Anycast BGW with TRM. For vPC BGW with TRM, vPC must be configured along with VxLAN TRM and VxLAN Multi-site.

Procedure

	Command or Action	**Purpose**
Step 1	**configure terminal** **Example:** `switch# ` **`configure terminal`**	Enters global configuration mode.
Step 2	**interface nve1** **Example:** `switch(config)# ` **`interface nve1`**	Configures the NVE interface.
Step 3	**no shutdown** **Example:** `switch(config-if-nve)# ` **`no shutdown`**	Brings up the NVE interface.
Step 4	**host-reachability protocol bgp** **Example:** `switch(config-if-nve)# ` **`host-reachability protocol`** **`bgp`**	Defines BGP as the mechanism for host reachability advertisement.
Step 5	**source-interface loopback** *src-if* **Example:** `switch(config-if-nve)# ` **`source-interface loopback`** **`0`**	Defines the source interface, which must be a loopback interface with a valid /32 IP address. This /32 IP address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising the address through a dynamic routing protocol in the transport network.

	Command or Action	**Purpose**
Step 6	**multisite border-gateway interface loopback** *vi-num* **Example:** switch(config-if-nve)# **multisite border-gateway interface loopback 1**	Defines the loopback interface used for the border gateway virtual IP address (VIP). The border-gateway interface must be a loopback interface that is configured on the switch with a valid /32 IP address. This /32 IP address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising the address through a dynamic routing protocol in the transport network. This loopback must be different than the source interface loopback. The range of *vi-num* is from 0 to 1023.
Step 7	**member vni** *vni-range* **associate-vrf** **Example:** switch(config-if-nve)# **member vni 10010 associate-vrf**	Configures the virtual network identifier (VNI). The range for *vni-range* is from 1 to 16,777,214 The value of *vni-range* can be a single value like 5000 or a range like 5001-5008.
Step 8	**mcast-group** *ip-addr* **Example:** switch(config-if-nve-vni)# **mcast-group 225.0.0.1**	Configures the NVE multicast group IP prefix within the fabric.
Step 9	**multisite mcast-group** *dci-core-group address* **Example:** switch(config-if-nve-vni)# **multisite mcast-group 226.1.1.1**	Configures the multicast group which is used to encapsulate TRM and L2 BUM packets in the DCI core.
Step 10	**multisite ingress-replication optimized** **Example:** switch(config-if-nve-vni)# **multisite ingress-replication optimized**	Defines the Multi-Site BUM replication method for extending the Layer 2 VNI.

Configuring TRM Multi-Site with IPv6 Underlay

This section provides the configuration procedure on Anycast BGW for TRM with IPv6 Multicast Underlay with Protocol-Independent Multicast (PIMv6) Any-Source Multicast (ASM) on the fabric side and Ingress Replication (IPv6) on the DCI side.

Before you begin

The following must be configured:

- VXLAN TRM
- VXLAN Multi-Site

SUMMARY STEPS

1. **configure terminal**
2. **interface nve1**

3. **no shutdown**
4. **host-reachability protocol bgp**
5. **source-interface loopback** *src-if*
6. **multisite border-gateway interface loopback** *vi-num*
7. **member vni** *vni-range* **associate-vrf**
8. **mcast-group** *ipv6-addr*
9. **multisite ingress-replication optimized**

DETAILED STEPS

Procedure

	Command or Action	Purpose
Step 1	**configure terminal** **Example:** switch# **configure terminal**	Enters global configuration mode.
Step 2	**interface nve1** **Example:** switch(config)# **interface nve1**	Configures the NVE interface.
Step 3	**no shutdown** **Example:** switch(config-if-nve)# **no shutdown**	Brings up the NVE interface.
Step 4	**host-reachability protocol bgp** **Example:** switch(config-if-nve)# **host-reachability protocol bgp**	Defines BGP as the mechanism for host reachability advertisement.
Step 5	**source-interface loopback** *src-if* **Example:** switch(config-if-nve)# **source-interface loopback 0**	Defines the source interface, which must be a loopback interface with a valid /128 IPv6 address. This /128 IPv6 address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising the address through a dynamic routing protocol in the transport network.
Step 6	**multisite border-gateway interface loopback** *vi-num* **Example:** switch(config-if-nve)# **multisite border-gateway interface loopback 1**	Defines the loopback interface used for the border gateway virtual IP address (VIP). The border-gateway interface must be a loopback interface that is configured on the switch with a valid /128 IPv6 address. This /128 IPv6 address must be known by the transient devices in the transport network and the remote VTEPs. This requirement is accomplished by advertising the address through a dynamic routing protocol in the transport network. This loopback must be different than the source interface loopback. The range of *vi-num* is from 0 to 1023.

	Command or Action	Purpose
Step 7	**member vni** *vni-range* **associate-vrf** **Example:** `switch(config-if-nve)# member vni 90001 associate-vrf`	Configures the virtual network identifier (VNI). The range for *vni-range* is from 1 to 16,777,214. The value of *vni-range* can be a single value like 5000 or a range like 5001-5008.
Step 8	**mcast-group** *ipv6-addr* **Example:** `switch(config-if-nve-vni)# mcast-group ff03:ff03::101:1`	Configures the NVE multicast group IPv6 prefix within the fabric.
Step 9	**multisite ingress-replication optimized** **Example:** `switch(config-if-nve-vni)# multisite ingress-replication optimized`	Defines the Multi-Site replication method for extending TRM functionality across sites.

Verifying TRM with Multi-Site Configuration

To display the status for the TRM with Multi-Site configuration, enter the following command:

Command	Purpose
show nve vni *virtual-network-identifier*	Displays the L3VNI. **Note** For this feature, optimized IR is the default setting for the Multi-Site extended L3VNI. MS-IR flag inherently means that it's MS-IR optimized.

Example of the **show nve vni** command:

For IPv4

```
switch(config)# show nve vni 51001
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication


Interface VNI      Multicast-group   State Mode Type [BD/VRF]     Flags
--------- -------- ---------------- ----- ---- ----------------- -----
nve1      51001    226.0.0.1         Up    CP   L3 [cust_1]       MS-IR
```

For IPv6

```
switch(config)# show nve vni 90001
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       S-ND - Suppress ND
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication
```

```
       HYB - Hybrid IRB mode

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ----------------- -----
nve1      90001    ff03:ff03::101:1  Up    CP   L3 [v1]           MS-IR

switch(config)#
```

VXLAN EVPN Multi-Site with RFC 5549 Underlay

VXLAN EVPN Multi-Site with RFC 5549 Underlay Overview

The VXLAN EVPN Multi-Site with RFC 5549 underlay feature is introduced for deployments where you already have a VXLAN IPv4 network infrastructure and must integrate IPv6 functionality.

The RFC 5549 enables the advertisement of underlay IPv4 prefixes using BGP with an IPv6 address as the next hop, effectively allowing IPv4 connectivity to be established over an IPv6 underlay network.

Benefits of VXLAN EVPN Multi-Site with RFC 5549 Underlay

- Helps your network against the exhaustion of IPv4 addresses.

- Enhances scalability and efficiency by integrating IPv6 into the VXLAN IPv4 network infrastructure.

- Offers simpler, more direct communication between devices.

- Provides sub-second convergence.

- Supports the following topologies:

  - Border spine gateway to leaf.

  - vPC BGWs to leaf.

  - Anycast BGWs to leaf.

Supported Platform and Release for VXLAN EVPN Multi-Site with RFC 5549 Underlay

Feature	Release	Platform
VXLAN BGP EVPN with RFC 5549 Underlay on multi-site	10.5(1)F	Cisco Nexus 9000 Series switches

Functionalities of VXLAN EVPN Multi-Site with RFC 5549 Underlay

This section describes how the VXLAN BGP EVPN with RFC 5549 underlay functions on a multi-site environments.

VTEP IPv4 Address Advertisement

The VTEP addresses are advertised in BGP with RFC 5549 underlay such that the IPv4 VTEP addresses are routed using the following IPv6 enabled interfaces.

- **Link-local address (LLA)**: In this type of BGP peering configuration, BGP uses the interface's link-local address to establish IPv6 sessions, allowing for the advertisement of IPv4 addresses through them. This approach eliminates the necessity for configuring global addresses on interfaces.

- **Global IPv6 address**: In this type of BGP peering configuration, BGP creates a standard IPv6 neighbor relationship using the global IPv6 address of the directly connected interface. Therefore, it is required to configure a global IPv6 address on the directly connected interface for the peering to be established.

VXLAN EVPN Services Advertisement over IPv4 Session

When VTEPs IPv4 reachability is advertised with RFC 5549 underlay, an EVPN IPv4 BGP session must be created to advertise VXLANv4 services on a multi-site environment.

VXLAN BGP EVPN with RFC 5549 Underlay on Multi-Site

On Multi-Site, the VXLAN BGP EVPN with RFC 5549 Underlay supports node to node communication as shown in the diagram.

The BGW's VTEP IPv4 address is advertised using IPv6 address (LLA or Global address) as the next-hop address in RFC 5549 underlay between the sites.
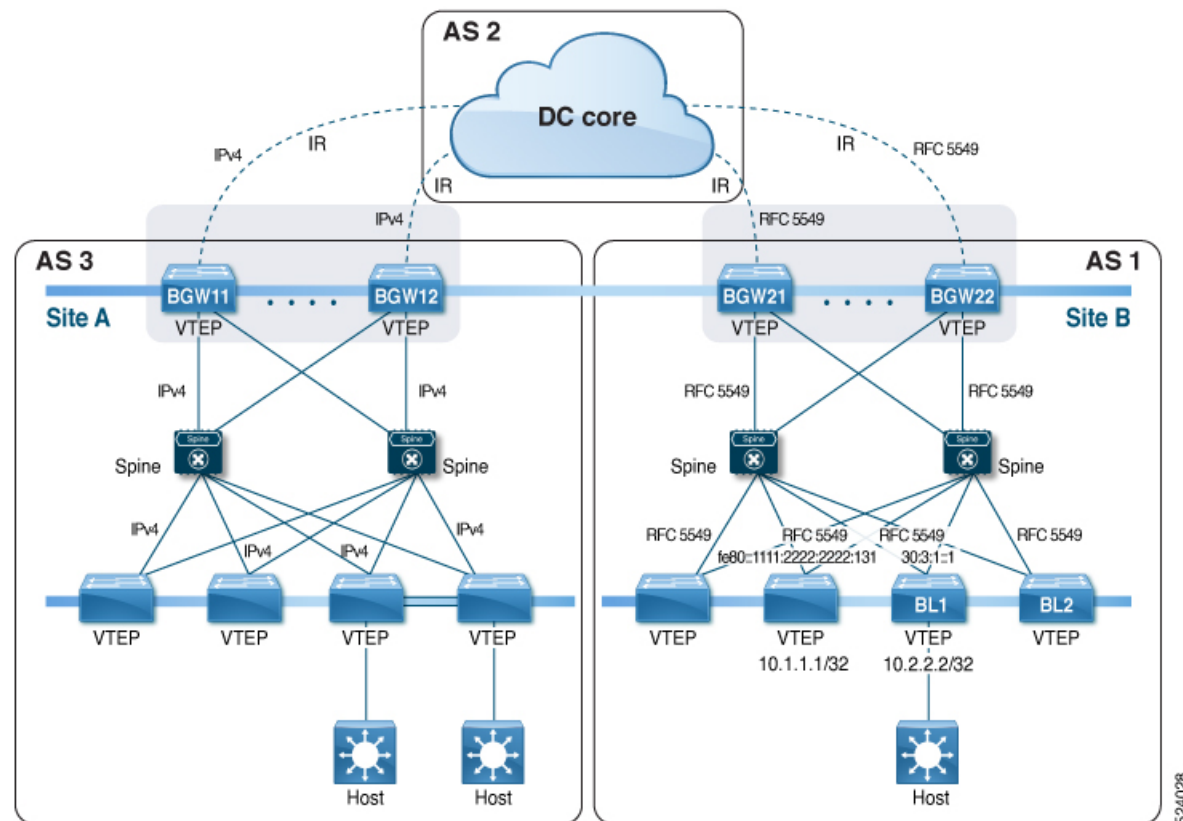
Guidelines for VXLAN EVPN Multi-Site with RFC 5549 Underlay

- VTEP must be from the same address family of IPv4.

- VTEP loop-back address must be an IPv4 address.

- During upgrading, ensure that the supported release versions are used for nodes as mentioned below:

Nodes	Supported Release
Core	10.2(3)F or higher
Spine	10.2(3)F or higher
Leaf	10.2(3)F or higher
Border Gateway	10.5(1)F or higher

- The routes to remote multi-site BGW loopback addresses must always prioritize the DCI link path over the iBGP protocol between vPC Border Gateway switches configured using the backup SVI. The backup SVI must be used strictly in the event of a DCI link failure.

- Configuring BGP peering on multiple interfaces with the same Link-Local Address (LLA) is not supported. If such a configuration is implemented, the resulting behavior will be unpredictable.

Supported and Unsupported Features of VXLAN EVPN Multi-Site with RFC 5549 Underlay

- Following are the supported and unsupported features for VXLAN BGP EVPN with RFC 5549 Underlay on multi-site environment:

Features	Supported Release	Supported or Not Supported
Ingress Replication	10.5(1)F	Supported
Overlay TCP Session	10.5(1)F	Supported
DSVNI	10.5(1)F	Supported
Anycast BGW	10.5(1)F	Supported
vPC BGW	10.5(1)F	Supported
vMCT	10.5(1)F	Supported
Route Leaking: Host addresses and between non default customer VRFs	10.5(1)F	Supported
IPv4 over RFC 5549 with default VRF	10.5(1)F	Supported
IPv4 over RFC 5549 with tenant VRF	-	Not Supported
Policy Based Routing	-	Not Supported
NGOAM	-	Not Supported

Features	Supported Release	Supported or Not Supported
ND Suppression	-	Not Supported
ARP Suppression	-	Not Supported
IPv4 VNF: IPv4 services over IPv4 PE-CE BGP session	-	Not Supported
RFC 5549 VNF: IPv4 services over IPv6 PE-CE BGP session	-	Not Supported
IPv6 VNF: IPv6 services over IPv6 PE-CE BGP session	-	Not Supported
vPC	-	Not Supported
BFD	-	Not Supported
Underlay multicast	-	Not Supported
MultiCast/TRM	-	Not Supported
Firewall Cluster	-	Not Supported
New L3VNI config	-	Not Supported
Group Policy Option	-	Not Supported
First Hop Security	-	Not Supported
PVLAN over VXLAN	-	Not Supported
VXLAN-TE	-	Not Supported
CloudSec	-	Not Supported

Configuring VXLAN EVPN Multi-Site with RFC 5549 Underlay on Multi-Site

You can perform the configuration of VXLAN EVPN Multi-Site with RFC 5549 underlay using LLA or global IPv6 address on a multi-site environment.

To configure BGW VXLAN EVPN Multi-Site with RFC 5549 underlay, follow these steps:

Procedure

Step 1 **configure terminal**

Example:

```
leaf# config terminal
leaf(config)#
```

Enters global configuration mode.

Step 2   **interface ethernet** *port*

Example:

```
leaf(config)# interface ethernet1/2
leaf(config-if)#
```

Enters Link local interface configuration mode for RFC 5549.

Step 3   **ip forward**

Example:

```
leaf(config-if)# ip forward
```

Enables IPv4 based lookup even when the interface VLAN has no IP address defined.

Step 4   To generate link-local address, use any of the following options:

> • Link-local IPv6 address
>
>   OR
> • Global IPv6 address

a)   To configure LLA, use any of the following options:

> • **ipv6 link-local** *LL_ipv6_address* (Manual option)
>
>   OR
>
> • **ipv6 address use-link-local-only** (Auto option)
>
>   OR
>
> • **ipv6 link-local use-bia** (Auto option)

Example:

```
leaf(config-if)# ipv6 link-local fe80::1111:2222:2222:3101
```

OR

```
leaf(config-if)# ipv6 address use-link-local-only
```

OR

```
leaf(config-if)# ipv6 link-local use-bia
```

Configures the link-local IPv6 address for the interface based on the specified choice.

OR

a)   Global IPv6 address: By default generates the LLA

> **Example:**
>
> **ipv6 address** *ipv6_address*
>
> ```
> leaf(config-if)# ipv6 address 2000:1:1::1/64
> ```
>
> Configures the Global IPv6 address for the interface.

Step 5   **exit**

Example:

```
leaf(config-if)# exit
leaf(config)#
```

Exits the interface configuration mode.

Step 6   **router bgp** *as-number*

Example:

```
leaf(config)# router bgp 1
leaf(config-router)#
```

Enters BGP router configuration mode.

Step 7   **neighbor** [*ipv6_address* | *ethernet port*]

Example:

```
leaf(config-router)# neighbor ethernet1/2
```

Configures a BGP neighbor for an interface.

Step 8   **remote-as** *value*

Example:

```
leaf(config-router)# remote-as 2
leaf(config-router-neighbor)#
```

Configures remote peer's autonomous system number for BGP neighbor.

Step 9   **peer-type fabric-external**

Example:

```
leaf(config-router-neighbor)# peer-type fabric-external
```

Enables the next hop rewrite for Multi-Site. Defines site external BGP neighbors for EVPN exchange. The default for **peer-type** is **fabric-internal**.

Note
The **peer-type fabric-external** command is required only for VXLAN Multi-Site BGWs.

Step 10   **address-family ipv4 unicast**

Example:

```
leaf(config-router-neighbor)# address-family ipv4 unicast
leaf(config-router-af)#
```

Configures the IPv4 unicast address family.

Step 11   **disable-peer-as-check**

Example:

```
leaf(config-router-af)# disable-peer-as-check
```

Disables checking the peer AS number during route advertisement. Configure this parameter on the spine for eBGP when all leafs are using the same AS but the spines have a different AS than leafs.

Note
This command is required for eBGP. For more information on eBGP configuration, see eBGP Underlay IP Network.

Examples of VXLAN EVPN Multi-Site with RFC 5549 Underlay for Multi-Site

- The following example shows how the IPv4 address is advertised over the IPv6 interface from spine to leaf using LLA address:

```
spine# show ip route 10.1.1.1
IP Route Table for VRF "default"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

fe80::1111:2222:2222:131, ubest/mbest: 1/0
    *via 10.1.1.1/32%default, Eth1/2, [200/0], 6d09h, bgp-2, external, tag 2
```

- The following example shows how the IPv4 address is advertised over the IPv6 interface from spine to leaf using global IPv6 address:

```
spine# show ip route 10.2.2.2
IP Route Table for VRF "default"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.2.2.2/32, ubest/mbest: 1/0
    *via 30:3:1::1%default, Eth1/2, [200/0], 6d09h, bgp-2, external, tag 2
```