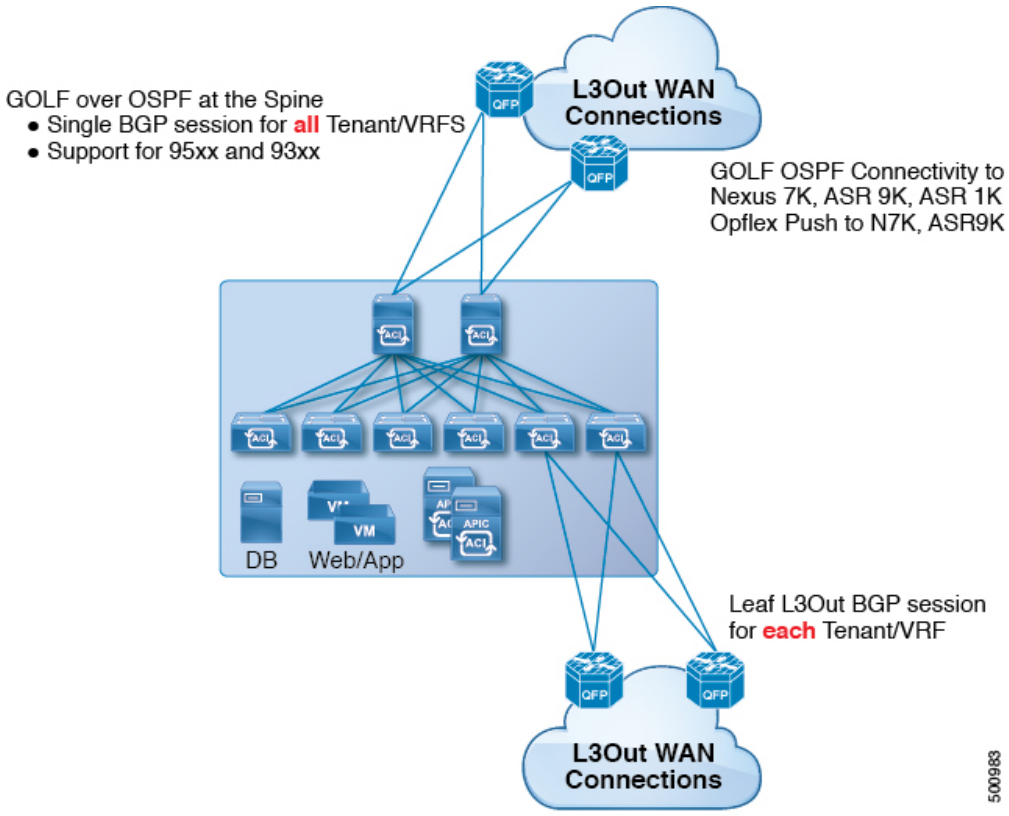
Cisco ACI GOLF

Cisco ACI GOLF

The Cisco ACI GOLF feature (also known as Layer 3 EVPN Services for Fabric WAN) enables much more efficient and scalable ACI fabric WAN connectivity. It uses the BGP EVPN protocol over OSPF for WAN routers that are connected to spine switches.

Cisco ACI GOLF

Figure 1: Cisco ACI GOLF Topology



All tenant WAN connections use a single session on the spine switches where the WAN routers are connected. This aggregation of tenant BGP sessions towards the Data Center Interconnect Gateway (DCIG) improves control plane scale by reducing the number of tenant BGP sessions and the amount of configuration required for all of them. The network is extended out using Layer 3 subinterfaces configured on spine fabric ports. Transit routing with shared services using GOLF is not supported.

A Layer 3 external outside network (`L3extOut`) for GOLF physical connectivity for a spine switch is specified under the `infra` tenant, and includes the following:

- `LNodeP` (`l3extInstP` is not required within the L3Out in the infra tenant.)
- A provider label for the `L3extOut` for GOLF in the infra tenant.
- OSPF protocol policies
- BGP protocol policies

All regular tenants use the above-defined physical connectivity. The `L3extOut` defined in regular tenants requires the following:

- An `l3extInstP` (EPG) with subnets and contracts. The scope of the subnet is used to control import/export route control and security policies. The bridge domain subnet must be set to advertise externally and it must be in the same VRF as the application EPG and the GOLF L3Out EPG.
- Communication between the application EPG and the GOLF L3Out EPG is governed by explicit contracts (not Contract Preferred Groups).

- An `l3extConsLbl` consumer label that must be matched with the same provider label of an `L3Out` for GOLF in the `infra` tenant. Label matching enables application EPGs in other tenants to consume the `LNodeP` external `L3Out` EPG.
- The BGP EVPN session in the matching provider `L3extOut` in the `infra` tenant advertises the tenant routes defined in this `L3Out`.

Guidelines and Limitations for Cisco ACI GOLF

Observe the following Cisco ACI GOLF guidelines and limitations:

- GOLF does not support shared services.
- GOLF does not support transit routing.
- GOLF routers must advertise at least one route to Cisco Application Centric Infrastructure (ACI) to accept traffic. No tunnel is created between leaf switches and the external routers until Cisco ACI receives a route from the external routers.
- All Cisco Nexus 9000 Series Cisco ACI-mode switches and all of the Cisco Nexus 9500 platform Cisco ACI-mode switch line cards and fabric modules support GOLF. With Cisco APIC, release 3.1(x) and higher, this includes the N9K-C9364C switch.
- At this time, only a single GOLF provider policy can be deployed on spine switch interfaces for the whole fabric.
- Up to Cisco APIC release 2.0(2), GOLF is not supported with Cisco ACI Multi-Pod. In release 2.0 (2), the two features are supported in the same fabric only over Cisco Nexus 9000 switches without "EX" on the end of the switch name; for example, N9K-9312TX. Since the 2.1(1) release, the two features can be deployed together over all the switches used in the Cisco ACI Multi-Pod and EVPN topologies.
- When configuring GOLF on a spine switch, wait for the control plane to converge before configuring GOLF on another spine switch.
- A spine switch can be added to multiple provider GOLF outside networks (GOLF L3Outs), but the provider labels have to be different for each GOLF L3Out. Also, in this case, the OSPF Area has to be different on each of the `L3extOut`s and use different loopback addresses.
- The BGP EVPN session in the matching provider `L3Out` in the `infra` tenant advertises the tenant routes defined in this `L3extOut`.
- When deploying three GOLF Outs, if only 1 has a provider/consumer label for GOLF, and 0/0 export aggregation, Cisco APIC will export all routes. This is the same as existing `L3extOut` on leaf switches for tenants.
- If you have an ERSPAN session that has a SPAN destination in a VRF instance, the VRF instance has GOLF enabled, and the ERSPAN source has interfaces on a spine switch, the transit prefix gets sent from a non-GOLF L3Out to the GOLF router with the wrong BGP next-hop.
- If there is direct peering between a spine switch and a data center interconnect (DCI) router, the transit routes from leaf switches to the ASR have the next hop as the PTEP of the leaf switch. In this case, define a static route on the ASR for the TEP range of that Cisco ACI pod. Also, if the DCI is dual-homed to the same pod, then the precedence (administrative distance) of the static route should be the same as the route received through the other link.

- The default `bgpPeerPfxPol` policy restricts routes to 20,000. For Cisco ACI WAN Interconnect peers, increase this as needed.
- In a deployment scenario where there are two `L3extOut`s on one spine switch, and one of them has the provider label `prov1` and peers with the DCI 1, the second `L3extOut` peers with DCI 2 with provider label `prov2`. If the tenant VRF instance has a consumer label pointing to any 1 of the provider labels (either prov1 or prov2), the tenant route will be sent out both DCI 1 and DCI 2.
- When aggregating GOLF OpFlex VRF instances, the leaking of routes cannot occur in the Cisco ACI fabric or on the GOLF device between the GOLF OpFlex VRF instance and any other VRF instance in the system. An external device (not the GOLF router) must be used for the VRF leaking.

Note Cisco ACI does not support IP fragmentation. Therefore, when you configure Layer 3 Outside (L3Out) connections to external routers, or Multi-Pod connections through an Inter-Pod Network (IPN), it is recommended that the interface MTU is set appropriately on both ends of a link. On some platforms, such as Cisco ACI, Cisco NX-OS, and Cisco IOS, the configurable MTU value does not take into account the Ethernet headers (matching IP MTU, and excluding the 14-18 Ethernet header size), while other platforms, such as IOS-XR, include the Ethernet header in the configured MTU value. A configured value of 9000 results in a max IP packet size of 9000 bytes in Cisco ACI, Cisco NX-OS, and Cisco IOS, but results in a max IP packet size of 8986 bytes for an IOS-XR untagged interface.

For the appropriate MTU values for each platform, see the relevant configuration guides.

We highly recommend that you test the MTU using CLI-based commands. For example, on the Cisco NX-OS CLI, use a command such as `ping 1.1.1.1 df-bit packet-size 9000 source-interface ethernet 1/1`.

APIC GOLF Connections Shared by Multi-Site Sites

For APIC Sites in a Multi-Site topology, if stretched VRFs share GOLF connections, follow these guidelines to avoid the risk of cross-VRF traffic issues.

Route Target Configuration between the Spine Switches and the DCI

There are two ways to configure EVPN route targets (RTs) for the GOLF VRFs: Manual RT and Auto RT. The route target is synchronized between ACI spines and DCIs through OpFlex. Auto RT for GOLF VRFs has the Fabric ID embedded in the format: – *ASN*: [*FabricID*] *VNID*

If two sites have VRFs deployed as in the following diagram, traffic between the VRFs can be mixed.

Site 1	Site 2
ASN: 100, Fabric ID: 1	ASN: 100, Fabric ID: 1
VRF A: VNID 1000 Import/Export Route Target: 100: [1] 1000	VRF A: VNID 2000 Import/Export Route Target: 100: [1] 2000
VRF B: VNID 2000 Import/Export Route Target: 100: [1] 2000	VRF B: VNID 1000 Import/Export Route Target: 100: [1] 1000

Route Maps Required on the DCI

Since tunnels are not created across sites when transit routes are leaked through the DCI, the churn in the control plane must be reduced as well. EVPN type-5 and type-2 routes sent from GOLF spine in one site towards the DCI should not be sent to GOLF spine in another site. This can happen when the DCI to spine switches have the following types of BGP sessions:

Site1 — IBGP ---- DCI ---- EBGP ---- Site2

Site1 — EBGP ---- DCI ---- IBGP ---- Site2

Site1 — EBGP ---- DCI ---- EBGP ---- Site2

Site1 — IBGP RR client ---- DCI (RR)---- IBGP ---- Site2

To avoid this happening on the DCI, route maps are used with different BGP communities on the inbound and outbound peer policies.

When routes are received from the GOLF spine at one site, the outbound peer policy towards the GOLF spine at another site filters the routes based on the community in the inbound peer policy. A different outbound peer policy strips off the community towards the WAN. All the route-maps are at peer level.

Configuring ACI GOLF Using the GUI

The following steps describe how to configure infra GOLF services that any tenant network can consume.

Procedure

Step 1 On the menu bar, click **Tenants**, then click **infra** to select the infra tenant.

Step 2 In the **Navigation** pane, expand the **Networking** option and perform the following actions:

a) Right-click **L3Outs** and click **Create L3Out** to open the **Create L3Out** wizard.

b) Enter the necessary information in the **Name**, **VRF** and **L3 Domain** fields.

c) In the **Use For:** field, select **Golf**.

The **Provider Label** and **Route Target** fields appear.

d) In the **Provider Label** field, enter a provider label (for example, `golf`).

e) In the **Route Target** field, choose whether to use automatic or explicit policy-governed BGP route target filtering policy:

- **Automatic** - Implements automatic BGP route-target filtering on VRFs associated with this routed outside configuration.
- **Explicit** - Implements route-target filtering through use of explicitly configured BGP route-target policies on VRFs associated with this routed outside configuration.

Note
Explicit route target policies are configured in the **BGP Route Target Profiles** table on the **BGP Page** of the **Create VRF Wizard**. If you select the **Automatic** option the in **Route Target** field, configuring explicit route target policies in the **Create VRF Wizard** might cause BGP routing disruptions.

f) Leave the remaining fields as-is (BGP selected, and so on), and click **Next**.

The **Nodes and Interfaces** window appears.

Step 3 Enter the necessary information in the **Nodes and Interfaces** window of the **Create L3Out** wizard.

a) In the **Node ID** drop-down list, choose a spine switch node ID.
b) In the **Router ID** field, enter the router ID.
c) (Optional) You can configure another IP address for a loopback address, if necessary.

The **Loopback Address** field is automatically populated with the same entry that you provide in the **Router ID** field. This is the equivalent of the **Use Router ID for Loopback Address** option in previous builds. Enter a different IP address for a loopback address, if you don't want to use route ID for the loopback address. Leave this field empty if you do not want to use the router ID for the loopback address.

d) Leave the **External Control Peering** field checked.
e) Enter necessary additional information in the **Nodes and Interfaces** window.

The fields that are shown in this window vary, depending on the options that you select in the **Layer 3** and **Layer 2** areas.

f) When you have entered the remaining additional information in the **Nodes and Interfaces** window, click **Next**.

The **Protocols** window appears.

Step 4 Enter the necessary information in the **Protocols** window of the **Create L3Out** wizard.

a) In the **BGP Loopback Policies** and **BGP Interface Policies** areas, enter the following information:

- **Peer Address**: Enter the peer IP address
- **EBGP Multihop TTL**: Enter the connection Time To Live (TTL). The range is 1–255 hops; if zero, no TTL is specified. The default is zero.
- **Remote ASN**: Enter a number that uniquely identifies the neighbor autonomous system. The Autonomous System Number can be in 4 byte as plain format 1–4294967295.

Note
ACI does not support asdot or asdot+ format autonomous system numbers.

b) In the **OSPF** area, choose the default OSPF policy, a previously created OSPF policy, or **Create OSPF Interface Policy**.
c) Click **Next**.

The **External EPG** window appears.

Step 5 Enter the necessary information in the **External EPG** window of the **Create L3Out** wizard.

a) In the **Name** field, enter a name for the external network.
b) In the **Provided Contract** field, enter the name of a provided contract.
c) In the **Consumed Contract** field, enter the name of a consumed contract.
d) In the **Allow All Subnet** field, uncheck if you don’t want to advertise all the transit routes out of this L3Out connection.

The Subnets area appears if you uncheck this box. Specify the desired subnets and controls as described in the following steps.

e) Click **Finish** to complete the necessary configurations in the **Create L3Out** wizard.

Step 6 In the **Navigation** pane for any tenant, expand the ***tenant_name*** > **Networking** > **L3Outs** and perform the following actions:

a) Right-click **L3Outs** and click **Create L3Out** to open the wizard.
b) Enter the necessary information in the **Name**, **VRF** and **L3 Domain** fields.

c) Check the box next to the **Use for GOLF** field.
d) In the **Label** field, select **Consumer**.
e) Assign a **Consumer Label**. In this example, use *golf* (which was created earlier).
f) Click **Next**, then click **Finish**.

Distributing BGP EVPN Type-2 Host Routes to a DCIG

Distributing BGP EVPN Type-2 Host Routes to a DCIG

In APIC up to release 2.0(1f), the fabric control plane did not send EVPN host routes directly, but advertised public bridge domain (BD) subnets in the form of BGP EVPN type-5 (IP Prefix) routes to a Data Center Interconnect Gateway (DCIG). This could result in suboptimal traffic forwarding. To improve forwarding, in APIC release 2.1x, you can enable fabric spines to also advertise host routes using EVPN type-2 (MAC-IP) host routes to the DCIG along with the public BD subnets.

To do so, you must perform the following steps:

1. When you configure the BGP Address Family Context Policy, enable Host Route Leak.
2. When you leak the host route to BGP EVPN in a GOLF setup:
   a. To enable host routes when GOLF is enabled, the BGP Address Family Context Policy must be configured under the application tenant (the application tenant is the consumer tenant that leaks the endpoint to BGP EVPN) rather than under the infrastructure tenant.
   b. For a single-pod fabric, the host route feature is not required. The host route feature is required to avoid sub-optimal forwarding in a multi-pod fabric setup. However, if a single-pod fabric is setup, then in order to leak the endpoint to BGP EVPN, a Fabric External Connection Policy must be configured to provide the ETEP IP address. Otherwise, the host route will not leak to BGP EVPN.
3. When you configure VRF properties:
   a. Add the BGP Address Family Context Policy to the BGP Context Per Address Families for IPv4 and IPv6.
   b. Configure BGP Route Target Profiles that identify routes that can be imported or exported from the VRF.

Distributing BGP EVPN Type-2 Host Routes to a DCIG Using the GUI

Enable distributing BGP EVPN type-2 host routes with the following steps:

Before you begin

You must have already configured ACI WAN Interconnect services in the infra tenant, and configured the tenant that will consume the services.

Procedure

Step 1 On the menu bar, click **Tenants** > **infra**.

Step 2 In the Navigation pane, navigate to **Policies** > **Protocol** > **BGP**.

Step 3 Right-click **BGP Address Family Context**, select **Create BGP Address Family Context Policy** and perform the following steps:

a) Type a name for the policy and optionally add a description.
b) Click the **Enable Host Route Leak** check box.
c) Click **Submit**.

Step 4 Click **Tenants** > **tenant-name** (for a tenant that will consume the BGP Address Family Context Policy) and expand **Networking**.

Step 5 Expand **VRFs** and click the VRF that will include the host routes you want to distribute.

Step 6 When you configure the VRF properties, add the **BGP Address Family Context Policy** to the **BGP Context Per Address Families** for IPv4 and IPv6.

Step 7 Click **Submit**.

Troubleshooting EVPN Type-2 Route Advertisement

Troubleshooting EVPN Type-2 Route Advertisement

Troubleshooting EVPN Type-2 Route Distribution to a DCIG

For optimal traffic forwarding in an EVPN topology, you can enable fabric spines to distribute host routes to a Data Center Interconnect Gateway (DCIG) using EVPN type-2 (MAC-IP) routes along with the public BD subnets in the form of BGP EVPN type-5 (IP Prefix) routes. This is enabled using the HostLeak object. If you encounter problems with route distribution, use the steps in this topic to troubleshoot.

SUMMARY STEPS

1. Verify that HostLeak object is enabled under the VRF-AF in question, by entering a command such as the following in the spine-switch CLI:
2. Verify that the config-MO has been successfully processed by BGP, by entering a command such as the following in the spine-switch CLI:
3. Verify that the public BD-subnet has been advertised to DCIG as an EVPN type-5 route:
4. Verify whether the host route advertised to the EVPN peer was an EVPN type-2 MAC-IP route:
5. Verify that the EVPN peer (a DCIG) received the correct type-2 MAC-IP route and the host route was successfully imported into the given VRF, by entering a command such as the following on the DCIG device (assuming that the DCIG is a Cisco ASR 9000 switch in the example below):

DETAILED STEPS

Procedure

Step 1 Verify that HostLeak object is enabled under the VRF-AF in question, by entering a command such as the following in the spine-switch CLI:

Example:

```
spine1# ls /mit/sys/bgp/inst/dom-apple/af-ipv4-ucast/
ctrl-l2vpn-evpn  ctrl-vpnv4-ucast  hostleak  summary
```

Step 2 Verify that the config-MO has been successfully processed by BGP, by entering a command such as the following in the spine-switch CLI:

Example:

```
spine1# show bgp process vrf apple
```

Look for output similar to the following:

```
 Information for address family IPv4 Unicast in VRF apple
    Table Id                  : 0
    Table state               : UP
    Table refcount            : 3
    Peers      Active-peers    Routes     Paths      Networks   Aggregates
    0          0               0          0          0          0

    Redistribution
        None

    Wait for IGP convergence is not configured
    GOLF EVPN MAC-IP route is enabled
    EVPN network next-hop 192.41.1.1
    EVPN network route-map map_pfxleakctrl_v4
    Import route-map rtctrlmap-apple-v4
    EVPN import route-map rtctrlmap-evpn-apple-v4
```

Step 3 Verify that the public BD-subnet has been advertised to DCIG as an EVPN type-5 route:

Example:

```
spine1# show bgp l2vpn evpn 10.6.0.0 vrf overlay-1
Route Distinguisher: 192.41.1.5:4123    (L3VNI 2097154)
BGP routing table entry for [5]:[0]:[0]:[16]:[10.6.0.0]:[0.0.0.0]/224, version 2088
Paths: (1 available, best #1)
Flags: (0x000002 00000000) on xmit-list, is not in rib/evpn
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type: local 0x4000008c 0x0 ref 1, path is valid, is best path
  AS-Path: NONE, path locally originated
    192.41.1.1 (metric 0) from 0.0.0.0 (192.41.1.5)
      Origin IGP, MED not set, localpref 100, weight 32768
      Received label 2097154
      Community: 1234:444
      Extcommunity:
          RT:1234:5101
          4BYTEAS-GENERIC:T:1234:444

  Path-id 1 advertised to peers:
    50.41.50.1
```

In the **Path type** entry, **ref 1** indicates that one route was sent.

Step 4 Verify whether the host route advertised to the EVPN peer was an EVPN type-2 MAC-IP route:

Example:

```
spine1# show bgp l2vpn evpn 10.6.41.1 vrf overlay-1
Route Distinguisher: 10.10.41.2:100    (L2VNI 100)
BGP routing table entry for [2]:[0]:[2097154]:[48]:[0200.0000.0002]:[32]:[10.6.41
.1]/272, version 1146
Shared RD: 192.41.1.5:4123    (L3VNI 2097154)
Paths: (1 available, best #1)
Flags: (0x00010a 00000000) on xmit-list, is not in rib/evpn
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type: local 0x4000008c 0x0 ref 0, path is valid, is best path
  AS-Path: NONE, path locally originated
  EVPN network: [5]:[0]:[0]:[16]:[10.6.0.0]:[0.0.0.0] (VRF apple)
    10.10.41.2 (metric 0) from 0.0.0.0 (192.41.1.5)
      Origin IGP, MED not set, localpref 100, weight 32768
      Received label 2097154 2097154
      Extcommunity:
          RT:1234:16777216

 Path-id 1 advertised to peers:
    50.41.50.1
```

The **Shared RD** line indicates the RD/VNI shared by the EVPN type-2 route and the BD subnet.

The **EVPN Network** line shows the EVPN type-5 route of the BD-Subnet.

The **Path-id advertised to peers** indicates the path advertised to EVPN peers.

Step 5 Verify that the EVPN peer (a DCIG) received the correct type-2 MAC-IP route and the host route was successfully imported into the given VRF, by entering a command such as the following on the DCIG device (assuming that the DCIG is a Cisco ASR 9000 switch in the example below):

Example:

```
RP/0/RSP0/CPU0:asr9k#show bgp vrf apple-2887482362-8-1 10.6.41.1
Tue Sep  6 23:38:50.034 UTC
BGP routing table entry for 10.6.41.1/32, Route Distinguisher: 44.55.66.77:51
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker              2088        2088
Last Modified: Feb 21 08:30:36.850 for 28w2d
Paths: (1 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  Local
    192.41.1.1 (metric 42) from 10.10.41.1 (192.41.1.5)
      Received Label 2097154
      Origin IGP, localpref 100, valid, internal, best, group-best, import-candidate, imported
      Received Path ID 0, Local Path ID 1, version 2088
      Community: 1234:444
      Extended community: 0x0204:1234:444 Encapsulation Type:8 Router
MAC:0200.c029.0101 RT:1234:5101
      RIB RNH: table_id 0xe0000190, Encap 8, VNI 2097154, MAC Address: 0200.c029.0101,
IP Address: 192.41.1.1, IP table_id 0x00000000
      Source AFI: L2VPN EVPN, Source VRF: default,
Source Route Distinguisher: 192.41.1.5:4123
```

In this output, the received RD, next hop, and attributes are the same for the type-2 route and the BD subnet.