Guidelines for Layer 3 Networking

Guidelines for Layer 3 Networking

Use the following guidelines when creating and maintaining Layer 3 outside connections.

Topic	Caution or Guideline
Issue where a border leaf switch in a vPC pair forwards a BGP packet with an incorrect VNID to an on-peer learned endpoint	If the following conditions exist in your configuration: • Two leaf switches are part of a vPC pair • For the two leaf switches connected behind the L3Out, the destination endpoint is connected to the second (peer) border leaf switch, and the endpoint is on-peer learned on that leaf switch If the endpoint is on-peer learned on the ingress leaf switch that receives a BGP packet that is destined to the on-peer learned endpoint, an issue might arise where the transit BGP connection fails to establish between the first layer 3 switch behind the L3Out and the on-peer learned endpoint on the second leaf switch in the vPC pair. This might happen in this situation because the transit BGP packet with port 179 is forwarded incorrectly using the bridge domain VNID instead of the VRF VNID. To resolve this issue, move the endpoint to any other non-peer leaf switch in the fabric so that it is not learned on the leaf switch.
Border leaf switches and GIR (maintenance) mode	If a border leaf switch has a static route and is placed in Graceful Insertion and Removal (GIR) mode, or maintenance mode, the route from the border leaf switch might not be removed from the routing table of switches in the ACI fabric, which causes routing issues. To work around this issue, either: • Configure the same static route with the same administrative distance on the other border leaf switch, or • Use IP SLA or BFD for track reachability to the next hop of the static route

Topic	Caution or Guideline
L3Out aggregate stats do not support egress drop counters	When accessing the **Select Stats** window through **Tenants** > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **Stats**, you will see that L3Out aggregate stats do not support egress drop counters. This is because there is currently no hardware table in the ASICs that record egress drops from the EPG VLAN, so stats do not populate these counters. There are only ingress drops for the EPG VLAN.
Updates through CLI	For Layer 3 external networks created through the API or GUI and updated through the CLI, protocols need to be enabled globally on the external network through the API or GUI, and the node profile for all the participating nodes needs to be added through the API or GUI before doing any further updates through the CLI.
Loopbacks for Layer 3 networks on same node	When configuring two Layer 3 external networks on the same node, the loopbacks need to be configured separately for both Layer 3 networks.
Ingress-based policy enforcement	Starting with Cisco APIC release 1.2(1), ingress-based policy enforcement enables defining policy enforcement for Layer 3 Outside (L3Out) traffic for both egress and ingress directions. The default is ingress. During an upgrade to release 1.2(1) or higher, existing L3Out configurations are set to egress so that the behavior is consistent with the existing configuration. You do not need any special upgrade sequence. After the upgrade, you change the global property value to ingress. When it has been changed, the system reprograms the rules and prefix entries. Rules are removed from the egress leaf and installed on the ingress leaf, if not already present. If not already configured, an `Actrl` prefix entry is installed on the ingress leaf. Direct server return (DSR), and attribute EPGs require ingress based policy enforcement. vzAny and taboo contracts ignore ingress based policy enforcement. Transit rules are applied at ingress.
Bridge Domains with L3Outs	A bridge domain in a tenant can contain a public subnet that is advertised through an `l3extOut` provisioned in the common tenant.
Bridge domain route advertisement For OSPF and EIGRP	When both OSPF and EIGRP are enabled on the same VRF on a node and if the bridge domain subnets are advertised out of one of the L3Outs, it will also get advertised out of the protocol enabled on the other L3Out. For OSPF and EIGRP, the bridge domain route advertisement is per VRF and not per L3Out. The same behavior is expected when multiple OSPF L3Outs (for multiple areas) are enabled on the same VRF and node. In this case, the bridge domain route will be advertised out of all the areas, if it is enabled on one of them.
BGP Maximum Prefix Limit	Starting with Cisco APIC release 1.2(1x), tenant policies for BGP `l3extOut` connections can be configured with a maximum prefix limit, that enables monitoring and restricting the number of route prefixes received from a peer. Once the maximum prefix limit has been exceeded, a log entry is recorded, and further prefixes are rejected. The connection can be restarted if the count drops below the threshold in a fixed interval, or the connection is shut down. Only one option can be used at a time. The default setting is a limit of 20,000 prefixes, after which new prefixes are rejected. When the reject option is deployed, BGP accepts one more prefix beyond the configured limit, before the APIC raises a fault.

Topic	Caution or Guideline
MTU	• 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. • The MTU settings for the Cisco ACI physical interfaces vary: • For sub-interfaces, the physical interface MTU is fixed and is set to 9216 for the front panel ports on the leaf switches. • For SVI, the physical interface MTU is set based on the fabric MTU policy. For example, if the fabric MTU policy is set to 9000, then the physical interface for the SVI is set to 9000.
QoS for L3Outs	To configure QoS policies for an L3Out and enable the policies to be enforced on the BL switch where the L3Out is located, use the following guidelines: • The VRF Policy Control Enforcement Direction must be set to**Egress**. • The VRF Policy Control Enforcement Preference must be set to **Enabled**. • When configuring the contract that controls communication between the EPGs using the L3Out, include the QoS class or Target DSCP in the contract or subject of the contract.
ICMP settings	ICMP redirect and ICMP unreachable are disabled by default in Cisco ACI to protect the switch CPU from generating these packets.