Cisco ACI Best Practices Quick Summary

July 2021

Contents

Introduction

This paper lists configuration options in Cisco ACI that the majority (if not all) of users should leverage. For configurations that should be used depending on use cases, refer to the Cisco ACI Design Guide, other whitepapers, or the configuration guides.

White papers: https://www.cisco.com/c/en/us/solutions/data-center-virtualization/application-centric-infrastructure/white-paper-listing.html

Configuration guides: https://www.cisco.com/c/en/us/support/cloud-systems-management/application-policy-infrastructure-controller-apic/tsd-products-support-series-home.html

Note: The documentation set for this product strives to use bias-free language. For the purposes of this documentation set, bias-free is defined as language that does not imply discrimination based on age, disability, gender, racial identity, ethnic identity, sexual orientation, socioeconomic status, and intersectionality. Exceptions may be present in the documentation due to language that is hardcoded in the user interfaces of the product software, language used based on RFP documentation, or language that is used by a referenced third-party product.

Endpoint Learning Settings

Enforce Subnet Check

Enforce Subnet Check prevents unnecessary or unintended endpoint learnings (both local and remote) based on the subnets configured in the bridge domains under each VRF instance. This is an enhanced version of the **Limit IP Learning to Subnet** option under the bridge domains, which only prevents learning of local endpoints.

The best practice is to enable this option.

Where

- 2.2(2q): Fabric > Access Policies > Global Policies > Fabric Wide Setting Policy
- 2.3(1) – 3.0(1): Not supported
- 3.0(2) – latest: System > System Settings > Fabric Wide Setting

Options/Notes

- This will only work on -EX, -FX, or later leaf nodes. When running older leaf node models, use **Limit IP Learning to Subnet** instead on all bridge domains.
- See the Cisco ACI Endpoint Learning document for details.

Endpoint IP Aging

By default, each endpoint (one MAC address and one or more IP addresses) has only one aging timer, which is called the endpoint retention timer. **IP Aging** enables each IP address to maintain its own timer so that it can age out individually. Without this feature enabled, as long as a MAC address remains active in the fabric, all associated IP addresses that were learned, even if those IP addresses are no longer originating traffic, will remain learned in the fabric associated to that MAC address.

The best practice is to enable this option.

Where

- 2.1(1) – 2.3(1): Fabric > Access Policies > Global Policies > IP Aging Policy
- 3.0(1) – latest: System > System Settings > Endpoint Controls > IP Aging

Options/Notes

- See Cisco ACI Endpoint Learning for details.

Loop Mitigation Settings

Enable MCP (per VLAN)

MisCabling Protocol (MCP) detects loops from external sources (such as misbehaving servers and external networking equipment running STP) and will err-disable the interface on which Cisco ACI receives its own packet.

The best practice is to enable this option (potentially also with "Enable MCP PDU per VLAN") on leaf node ports that are connected to external Layer 2 networks that may introduce loops.

Where

- 1.1(1) – 3.1(2): Fabric > Access Policies > Global Policies > MCP Instance Policy default
- 3.2(1) – latest: Fabric > Access Policies > Policies > Global > MCP Instance Policy default

Options/Notes

- The "Enable MCP PDU per VLAN" option (available after 2.0(2)) enables MCP to send packets on a per-VLAN basis. Otherwise, these packets will only be sent on untagged VLANs and loops will be detected only on those VLANs. Per VLAN MCP has a scalability limit of 256 VLANs per interface.
- Cisco ACI has a per leaf node scalability limit of 2,000 logical ports (VLANs x ports).
- If your system's scale might exceed these limits, make sure to be cautious when enabling MCP, especially with per VLAN, because handling the MCP PDUs per VLAN can be CPU intensive.

  See the Verified Scalability Guide for up-to-date scalability numbers for each firmware version.

EP Loop Protection or Rogue Endpoint Control

EP Loop Protection detects a loop by detecting an endpoint being learned on the same set of two interfaces back and forth. When a loop is detected, the Cisco ACI fabric shuts down one of the interfaces between which the endpoint was moving (***Port Disable***) or disables endpoint learning in the bridge domain that has the loop (***BD Learn Disable***). When neither actions are enabled, **EP Loop Protection** logs the loop event in the internal process (EPMC – Endpoint Manger Client) log, but does not take any actions nor raise a fault.

Rogue Endpoint Control identifies an endpoint (MAC/IP address) as rogue when the same endpoint is learned on different interfaces multiple times within the configured interval. The misbehaving rogue endpoint is pinned down to the interface on which it was last learned to prevent the further move and will be deleted after the configured hold interval. This protects the Cisco ACI fabric from constantly having to update the devices in the fabric regarding the new endpoint location, allowing for a more stable Cisco ACI environment.

A fault is also raised for both options, which can then be sent to your syslog/SNMP trap, if configured.

The best practice is to enable **Rogue Endpoint Control**, which acts per endpoint instead of per port or bridge domain as with **EP Loop Protection**.

Where (EP Loop Protection)

- 1.1(1) – 2.3(1): Fabric > Access Policies > Global Policies > EP Loop Protection Policy
- 3.0(1) – latest: System > System Settings > Endpoint Controls > Ep Loop Protection

Where (Rogue Endpoint Control)

- 3.2(1) – latest: System > System Settings > Endpoint Controls > Rogue EP Control

Options/Notes

- When **Rogue Endpoint Control** is enabled, **EP Loop Protection** does not take effect. Choose one or the other after understanding the pros and cons of each option to mitigate the impact of loops.
- See the Cisco ACI Design Guide and Cisco ACI Endpoint Learning for details.
- When enabling **Rogue Endpoint Control** or **EP Loop Protection** in the existing fabric, ensure that there are no loops or flaps currently happening in the fabric. Otherwise, the error actions will take place immediately.

Bridge Domain Settings

Unicast Routing

Unicast Routing enables the bridge domain to route traffic and learn endpoint IP addresses.

The best practice is not to enable this option when the default gateway for endpoints is not the bridge domain SVI.

Where

- Tenant > Networking > Bridge Domains > Policy > L3 Configurations

Options/Notes

- When the default gateway for endpoints is not the bridge domain switch virtual interface (SVI), the bridge domain only does switching. If **Unicast Routing** is enabled in this case and IP addresses are learned on the bridge domain, this configuration may lead to a packet forwarding issues. See Cisco ACI Endpoint Learning whitepaper for details.

L2 Unknown Unicast

L2 Unknown Unicast decides whether the bridge domain should flood packets that are destined to an unknown MAC address (***Flood***) or should send it to a spine node for COOP database lookup (***Hardware Proxy***).

The best practice is to set this option to ***Flood*** in either of the following scenarios:

- Unicast Routing is disabled.

- Layer 2 is extended using non-Cisco ACI switches, such as during migration.

Where

- Tenant > Networking > Bridge Domains > Policy > General

Options/Notes

- See the Cisco ACI Design Guide for details on the two scenarios mentioned above.

ARP Flooding

ARP Flooding decides whether the bridge domain should flood ARP requests all the time (***Enabled***) or should look up the target IP address in the ARP header and perform unicast routing (***Disabled***).

The best practice is set this option to ***Enabled*** when there are clustered servers, firewalls, or load balancers so that GARP is flooded.

Where

- Tenant > Networking > Bridge Domains > Policy > General

Options/Notes

- See the Cisco ACI Design Guide for details.

QoS Settings

DSCP Translation

DSCP Translation translates Cisco ACI QoS classes into DSCP in the outer IP address header of VXLAN packets to ensure that the classes are preserved when traffic is traversing across pods or sites. Without this option, Cisco ACI QoS classes are carried as CoS by way of the outer Dot1Q header, which has a higher risk of being changed or removed in IPN/ISN.

The best practice is to enable **DSCP Translation** and assign DSCP classes that are not used in IPN/ISN to Cisco ACI QoS classes, which ensures that those DSCP values are not overwritten by IPN/ISN.

Where

- Tenant > infra > Policies > Protocol > DSCP class-CoS translation policy for L3 traffic

Options/Notes

- **DSCP Translation** and **Preserve CoS** cannot be used at the same time.
- **Preserve CoS** also translates Cisco ACI QoS classes along with the original CoS from the ingress leaf node into DSCP. However, **Preserve CoS** uses non-configurable internal DSCP mappings, which means that users do not have the flexibility of choosing which DSCP values to trust and to be untouched in IPN/ISN, while **DSCP Translation** enables you to map DSCP values of your choice to Cisco ACI QoS classes with a trade-off of not being able to perverse the original CoS.
- If your Cisco ACI fabric is not using neither Cisco ACI Multi-Pod and Cisco ACI Multi-Site, you may use **Preserve CoS**.

Other Settings

Fabric Port Tracking

Fabric Port Tracking monitors the number of operational fabric ports (uplinks) on a given leaf node and if it is decreased to the configured threshold or lower, the downlink ports of the leaf node will be brought down so that external devices can switch over to other healthy leaf nodes.

The best practice is to enable this option with zero active fabric ports as the threshold.

Where

- 1.2(2) – 3.2(1): Fabric > Access Policies > Policies > Global > Port Tracking
- 3.2(2) – latest: System > System Settings > Port Tracking

Options/Notes

- If all of your non-Cisco ACI devices are connected to two or more leaf nodes for redundancy with an appropriate failover mechanism, such as vPC, you may configure more than zero as the threshold.
- If all of your APICs are connected to two leaf nodes for redundancy as recommended (see APIC Connectivity), you may enable the ***Include APIC ports*** option.

Global AES Encryption

AES Encryption enables Cisco APICs to encrypt passwords and include them in the configuration export (backup). When importing such a configuration backup, you are required to provide the same passphrase that was used to enable **AES Encryption**. Without configuring this option, the configuration backup does not have passwords and configurations that need passwords, such as VMM domains, will stop working after the import.

The best practice is to enable this option.

Where

- Prior to 4.0(1): Admin > AAA > AES Encryption Passphrase and Keys for Config Export (and Import)
- From 4.0(1): System > System Settings > Global AES Passphrase Encryption Settings

Options/Notes

- If you forget the passphrase, reconfigure **AES Encryption** with a new passphrase and export the configuration again.

VLAN Pool

A **VLAN Pool** decides which VXLAN ID (VNID) is assigned to each VLAN. For example, VLAN 10 from VLAN pool A and VLAN pool B will be assigned different VNID. AEPs represent a group of interfaces on Cisco ACI switches. The Cisco APICs decide which VLAN pool to use for which VLAN on which interface based on domains such as physical domain that tie a VLAN pool and AEPs.

The best practice is to configure minimum number of VLAN pools to avoid overlapping VLAN ranges.

Where

- Fabric > Access Policiess > Pools > VLAN

Options/Notes

- When there are multiple VLAN pools with overlapping VLAN ID ranges tied to the same AEP, VNID assignments may be indeterministic and cause various issues to endpoint learnings, STP BPDU flooding, and so on.
- Ultimately, one or two VLAN pools for the entire fabric may be enough if you do not need features that require different VLAN pools, such as per-port-VLAN.
- Consider **Enforce EPG VLAN Validation** under System > System Settings > Fabric Wide Setting (available starting in the 3.2(6) release), which prevents two domains containing overlapping VLAN pools from being associated to the same EPG. If you are familiar with the VNID assignment logic and need to use overlapping VLAN pools on purpose, you do not need this validation. Otherwise, we recommend that you enable this option.

ISIS Redistribution Metric

In a Cisco ACI Multi-Pod deployment, the **ISIS Redistribution Metric** is the metric set for Cisco ACI infra TEP routes when spine nodes redistribute these routes from a routing protocol (such as OSPF) into ISIS. These redistributed ISIS routes are advertised to leaf nodes in the same pod so that those can reach to the other pod through the spine nodes.

The best practice is set this metric to 62 or lower as opposed to the maximum 63, which is the default.

Where

- Prior to 5.0(1): Fabric > Fabric Policies > Policies > Pod > ISIS Policy Default > ISIS metric for redistributed routes
- From 5.0(1): System > System Settings > ISIS Policy > ISIS metric for redistributed routes

Options/Notes

- When a spine node reboots or newly joins a fabric, until the spine node stabilizes and completes the policy download from the Cisco APIC, the node tries to advertise ISIS redistributed routes with the higher metric. This is known as " overload mode." If the **ISIS Redistribution Metric** is kept at the default value of 63, which is the maximum, the overload functionality is ineffective, since the metric for overload and non-overload is the same. This results in potential longer convergence times after a spine node reboots in a Cisco ACI Multi-Pod setup. By lowering the value, leaf nodes can prefer other stable spine nodes to reach the other pods.
- Configuring the **ISIS Redistribution Metric** value is non-disruptive.

COOP Group

Setting **COOP Group** to ***Strict*** enables the Cisco ACI switch nodes to use MD5 authentication for all COOP communication to ensure that Cisco ACI switch nodes will exchange COOP database information only between the switches in the same fabric.

The best practice is to set this option to ***Strict***.

Where

- System > System Settings > COOP Group

Options/Notes

- The MD5 token is automatically updated every hour by the Cisco APICs and is sent to the switches managed by the Cisco APICs.

Topology

APIC Connectivity

APICs have 2 physical interfaces that you connect to ACI leaf nodes. The APICs use these interfaces to communicate with and manage all ACI switch nodes and other APICs in the same ACI fabric. These interfaces work as active/standby on each APIC.

The best practice is to connect two different leaf nodes for each APIC to ensure each APIC has reachability to the fabric even when one of the leaf nodes is down, such as during upgrades. When possible, distribute all APICs in the APIC cluster to different sets of leaf nodes.

Options/Notes

- See the Cisco ACI Design Guide for details of APIC connectivity and APIC cluster designs.
- The same principal, that is to connect interfaces on each APIC to different switches, applies also when APICs are connected to the ACI fabric remotely over the IPN. See Deploying APIC Cluster Connectivity to the Fabric Over a Layer 3 Network for details on this option.

Switch Connectivity

ACI switches are always in a spine-leaf topology. This fundamental topology is fixed, but can be expanded with advanced features such as multi-tier where you can have a second tier/layer of leaf nodes, or remote leaf where some leaf nodes are connected remotely through IPN.

The best practice is to have a full-mesh cable connectivity between the spine nodes and leaf nodes.

Options/Notes

- See the Cisco ACI Design Guide for more details.

Printed in USA

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam
The Netherlands

Cisco has more than 200 offices worldwide. Address, phone numbers, and fax numbers are listed on the Cisco Website at https://www.cisco.com/go/offices.