Cisco APIC System Management Configuration Guide, Release 5.3(x)

First Published: 2023-11-24

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Trademarks

THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS REFERENCED IN THIS DOCUMENTATION ARE SUBJECT TO CHANGE WITHOUT NOTICE. EXCEPT AS MAY OTHERWISE BE AGREED BY CISCO IN WRITING, ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS DOCUMENTATION ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED.

The Cisco End User License Agreement and any supplemental license terms govern your use of any Cisco software, including this product documentation, and are located at: https://www.cisco.com/c/en/us/about/legal/cloud-and-software/software-terms.html. Cisco product warranty information is available at https://www.cisco.com/c/en/us/products/warranty-listing.html. US Federal Communications Commission Notices are found here https://www.cisco.com/c/en/us/products/us-fcc-notice.html.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Any products and features described herein as in development or available at a future date remain in varying stages of development and will be offered on a when-and if-available basis. Any such product or feature roadmaps are subject to change at the sole discretion of Cisco and Cisco will have no liability for delay in the delivery or failure to deliver any products or feature roadmap items that may be set forth in this document.

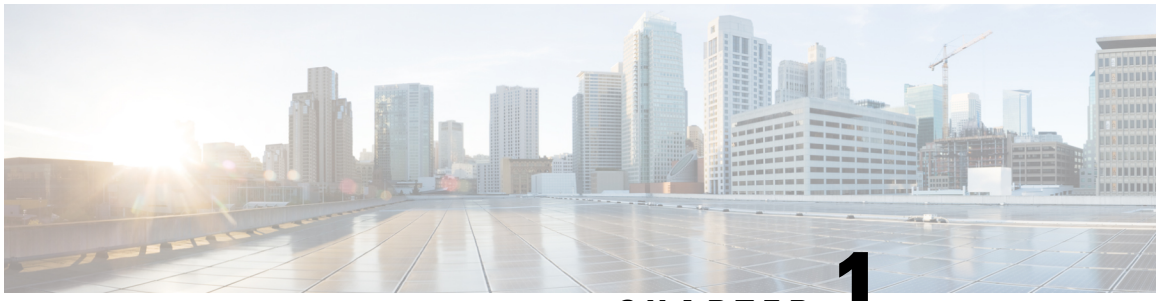
Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

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.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: https://www.cisco.com/c/en/us/about/legal/trademarks.html. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1721R)

Trademarks

CONTENTS



CHAPTER 1

New and Changed Information

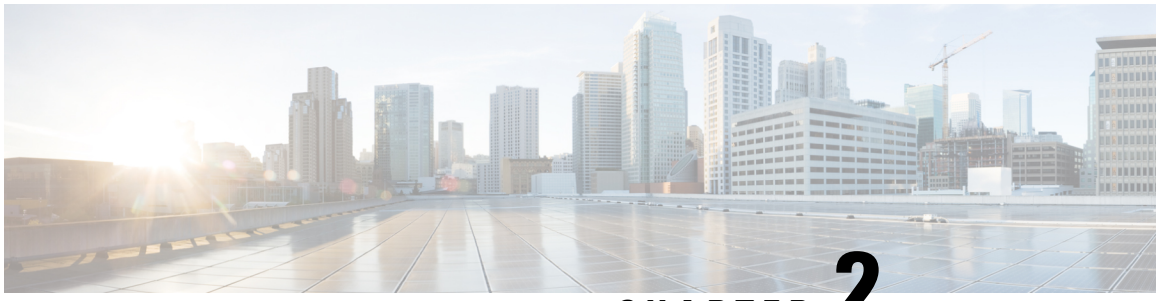
New and Changed Information

The following table provides an overview of the significant changes to the organization and features in this guide up to the specified release. The table does not provide an exhaustive list of all changes made to the guide or of the new features up to that release.

Table 1: New Features and Changed Behavior in Cisco APIC Release 5.3(1)

Feature or Change	Description	Where Documented
N/A	This document has no changes from the previous release.	N/A

New and Changed Information



CHAPTER 2

Alias, Annotations, and Tags

Alias, Annotations, and Tags

To simplify the identifying, addressing, and grouping of objects, ACI provides several methods for the user to add label metadata to objects. These methods are summarized in the list below:

- **Name Alias**: A cosmetic substitute for a GUI entity.
- **Global Alias**: A label, unique within the fabric, that can serve as a substitute for an object's Distinguished Name (DN).
- **Tag Instance / Annotation**: A simple note or description.
- **Policy Tag**: A label for grouping of objects, which need not be of the same class.

Alias

In the ACI object model, every object has a unique Distinguished Name (DN), which is an often lengthy identifier that includes the names of its parent object hierarchy and itself. For example, consider a tenant named **Tenant2468** that contains an application profile named **ap13**, which contains an application endpoint group named **aepg35**. The DN of that application endpoint group, generated by APIC, is **uni/tn-Tenant2468/ap-ap13/epg-aepg35**. After each of these objects is created, ACI typically does not allow their names to be changed, as that would cause a change in the DNs of all descendant objects of the renamed object. To overcome this inconvenience, ACI provides two alias functions — Name Alias for the GUI and Global Alias for the API.

Name Alias

The Name Alias feature (or simply "Alias" where the setting appears in the GUI) changes the displayed name of objects in the APIC GUI. While the underlying object name cannot be changed, the administrator can override the displayed name by entering the desired name in the **Alias** field of the object properties menu. In the GUI, the alias name then appears along with the actual object name in parentheses, as *name_alias (object_name)*. Many object types, such as tenants, application profiles, bridge domains, and EPGs, support the alias property. In the object model, the name alias property is `objectClass.nameAlias`. The property for a tenant object, for example, is `fvTenant.nameAlias`.

Using the preceding example of a tenant, suppose the administrator prefers to see the tenant name "AcmeManufacturing" instead of "Tenant2468." By entering the preferred name in the **Alias** field of the Tenant2468 tenant properties, the GUI would now display **AcmeManufacturing (Tenant2468)**.

The name alias property is purely cosmetic for the APIC GUI. The alias need not be unique in any scope, and the same value can be used as name alias for other objects.

Global Alias

The Global Alias feature simplifies querying a specific object in the API. When querying an object, you must specify a unique object identifier, which is typically the object's DN. As an alternative, this feature allows you to assign to an object a label that is unique within the fabric. Using the preceding example, without a global alias, you would query the application endpoint by its DN using this API request:

```
GET: https://APIC_IP/api/mo/uni/tn-Tenant2468/ap-ap13/epg-aepg35.json
```

By configuring a simpler yet unique name in the **Global Alias** field of the object properties menu, you can use the global alias along with a different API command to query the object:

```
GET: https://APIC_IP/api/alias/global_alias.json
```

Using the preceding example, by entering "AcmeEPG35" in the **Global Alias** field of the application endpoint group's configuration properties, the query URL would now be:

```
GET: https://APIC_IP/api/alias/AcmeEPG35.json
```

In the APIC object model, the global alias is a child object (tagAliasInst) attached to the object that is being aliased. In the preceding example, the global alias object would be a child object of the application endpoint group object.

For additional information, see the "Tags and Aliases" chapter of the *APIC REST API Configuration Guide*.

Creating a Name Alias or Global Alias

This example procedure shows you how to create a name alias and a global alias for an application profile of a tenant. Many other objects support these alias features using the same procedure after navigating to the object.

Procedure

Step 1 On the menu bar, choose **Tenants** and select the applicable Tenant.

Step 2 In the **Navigation** pane, expand ***tenant_name*** > **Application Profiles** > ***application_profile_name***.

Step 3 In the **Work** pane, click the **Policy** tab.

The **Properties** page of the application profile appears.

Step 4 In the **Alias** field, enter a name alias.

The alias need not be unique in any scope.

Step 5 In the **Global Alias** field, enter an alias for the distinguished name (DN) of the application profile.

The global alias must be unique within the fabric.

Step 6 Click **Submit**.

If you configured a name alias, the application profile is now identified in the **Navigation** pane as *alias* (*name*). For example, if the **Name** is **ap1234** and you configured an **Alias** as **SanJose**, the application profile appears as **SanJose** (**ap1234**).

If you configured a global alias, you can now substitute that value for the distinguished name (DN) of the application profile in API commands that support the global alias.

Annotations

You can add arbitrary key:value pairs of metadata to an object as annotations (`tagAnnotation`). Annotations are provided for the user's custom purposes, such as descriptions, markers for personal scripting or API calls, or flags for monitoring tools or orchestration applications such as Cisco Multi-Site Orchestrator (MSO). Because APIC ignores these annotations and merely stores them with other object data, there are no format or content restrictions imposed by APIC.

Evolution of Annotations

APIC support for user-defined annotation information has changed over time in the following steps:

- Prior to Cisco APIC Release 4.2(4), APIC supported tag instances (`tagInst`), which stored a simple string. In APIC GUI menus, these were labeled as "Tags."
- In Cisco APIC Release 4.2(4), because many modern systems use a key and value pair as a label, changes were made to move to key:value annotations (`tagAnnotation`) as the main label option for API. The shortcut API to query objects via tag instances (**/api/tag/***your_tag***.json**) was deprecated. The APIC GUI continued to use the simple string tag instances (`tagInst`), labeled as "Tags.".
- In Cisco APIC Release 5.1(1), tag instances (`tagInst`) were deprecated in the GUI. GUI menus still used the term "Tags," but actually configured annotations (`tagAnnotation`). Also beginning with this release, a list of all annotations can be viewed from **Fabric > Fabric policies > Tags**.
- In Cisco APIC release 5.2(1), GUI menu labels were changed from "Tags" to "Annotations." This change was made to avoid confusion with Policy Tags.

Creating an Annotation

This example procedure shows you how to create an annotation for a tenant. Many other objects support the annotation feature using the same procedure after navigating to the object.

Procedure

Step 1 On the menu bar, choose **Tenants** and select the applicable Tenant.

Step 2 In the **Navigation** pane, click the ***tenant_name***.

Step 3 In the **Work** pane, click the **Policy** tab.

The properties menu of the tenant appears.

Step 4 Next to **Annotations**, click the + symbol to add a new annotation.

Step 5 In the annotation key box, choose an existing key or type a new key.

Step 6 In the annotation value box, type a value.

Allowed characters for key and value are a-z, A-Z, 0-9, period, colon, dash, or underscore.

Step 7 Click the ✓ symbol to save the annotation.

You can add more annotations by repeating these steps.

Policy Tags

Policy tags (`tagTag`), or simply tags, are user-definable key and value pairs for use by ACI features. You can configure multiple tags on a single object, and you can apply the same tag on multiple objects. Because many object classes support policy tags, you can use policy tags to group disparate objects. For example, a policy tag can be used to group endpoints, subnets, and VMs together as one Endpoint Security Group (ESG) using ESG tag selectors in Cisco APIC Release 5.2(1).

ACI features using policy tags include:

- Endpoint Security Group (ESG)

Creating a Policy Tag

This example procedure shows you how to create a policy tag for a static endpoint. Several other objects support policy tags using the same procedure after navigating to the object.

Procedure

Step 1 On the menu bar, choose **Tenants** and select the applicable Tenant.

Step 2 In the **Navigation** pane, expand ***tenant_name*** > **Application Profiles** > ***application_profile_name*** > **Application EPGs** > ***application_epg_name*** > **Static Endpoint**.

Step 3 In the **Work** pane, double-click the static endpoint to be tagged.

The Static Endpoint properties dialog box appears.

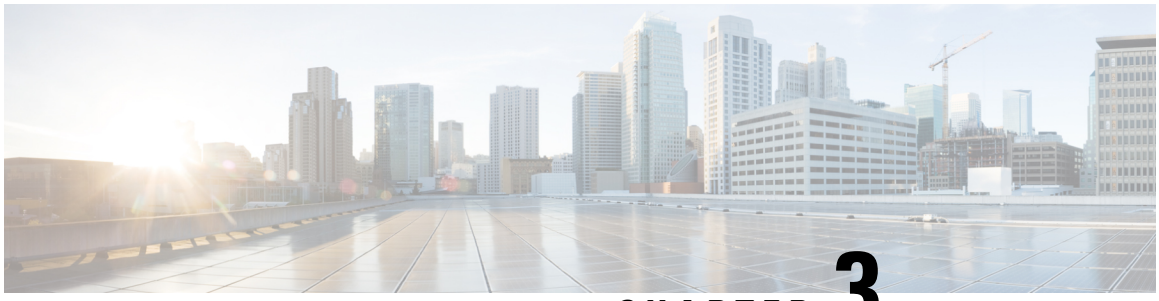
Step 4 Next to **Policy Tags**, click the + symbol to add a new policy tag.

Step 5 In the tag key box, choose an existing key or type a new key.

Step 6 In the tag value box, type a tag value.

Allowed characters for key and value are a-z, A-Z, 0-9, period, colon, dash, or underscore.

Step 7 Click the ✓ symbol to save the tag.



CHAPTER 3

Precision Time Protocol

About PTP

The Precision Time Protocol (PTP) is a time synchronization protocol defined in IEEE 1588 for nodes distributed across a network. With PTP, you can synchronize distributed clocks with an accuracy of less than 1 microsecond using Ethernet networks. PTP's accuracy comes from the hardware support for PTP in the Cisco Application Centric Infrastructure (ACI) fabric spine and leaf switches. The hardware support allows the protocol to compensate accurately for message delays and variation across the network.

Note This document uses the term "client" for what the IEEE1588-2008 standard refers to as the "slave." The exception is instances in which the word "slave" is embedded in the Cisco Application Policy Infrastructure Controller (APIC) CLI commands or GUI.

PTP is a distributed protocol that specifies how real-time PTP clocks in the system synchronize with each other. These clocks are organized into a master-client synchronization hierarchy with the grandmaster clock, which is the clock at the top of the hierarchy, determining the reference time for the entire system. Synchronization is achieved by exchanging PTP timing messages, with the members using the timing information to adjust their clocks to the time of their master in the hierarchy. PTP operates within a logical scope called a PTP domain.

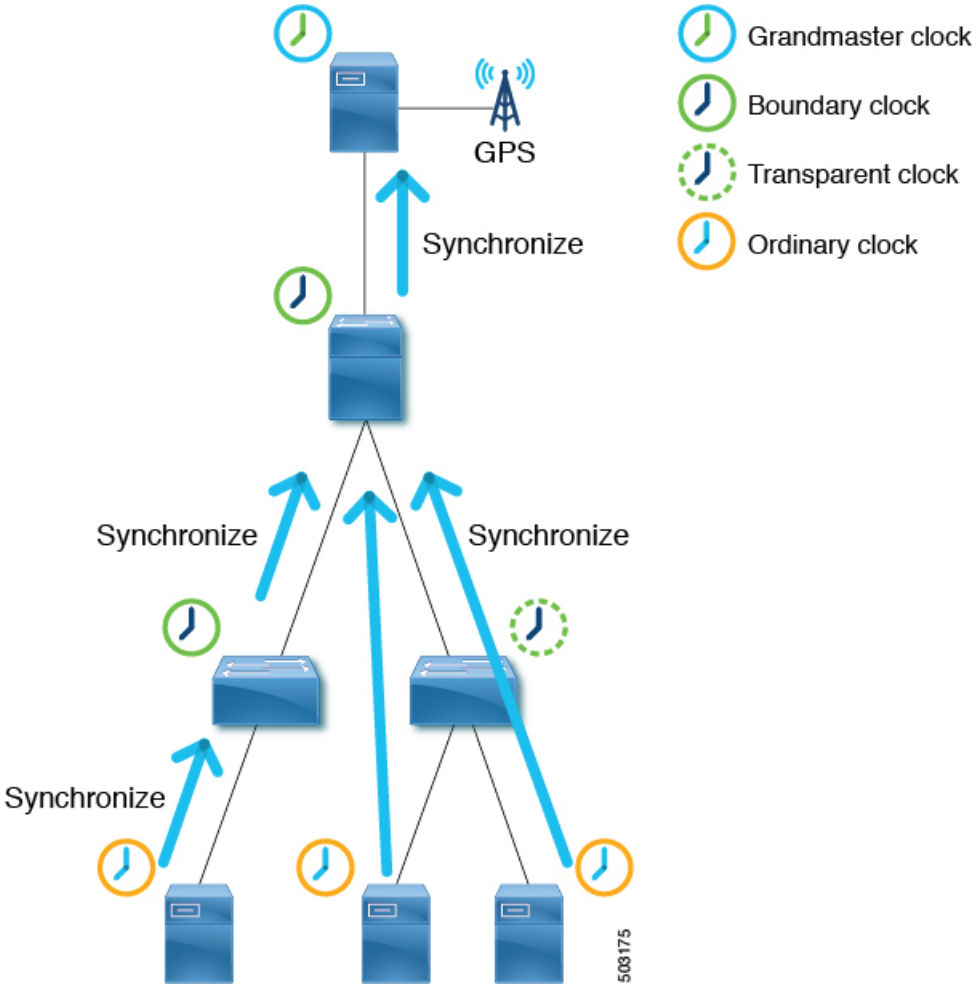
The PTP process consists of two phases: establishing the master-client hierarchy and synchronizing the clocks. Within a PTP domain, each port of an ordinary or boundary clock uses the following process to determine its state:

1. Establish the master-client hierarchy using the Best Master Clock Algorithm (BMCA):
   - Examine the contents of all received `Announce` messages (issued by ports in the master state).
   - Compare the data sets of the foreign master (in the `Announce` message) and the local clock for priority, clock class, accuracy, and so on.
   - Determine its own state as either master or client.
2. Synchronize the clocks:

- Use messages, such as `Sync` and `Delay_Req`, to synchronize the clock between the master and clients.

PTP Clock Types

The following illustration shows the hierarchy of the PTP clock types:



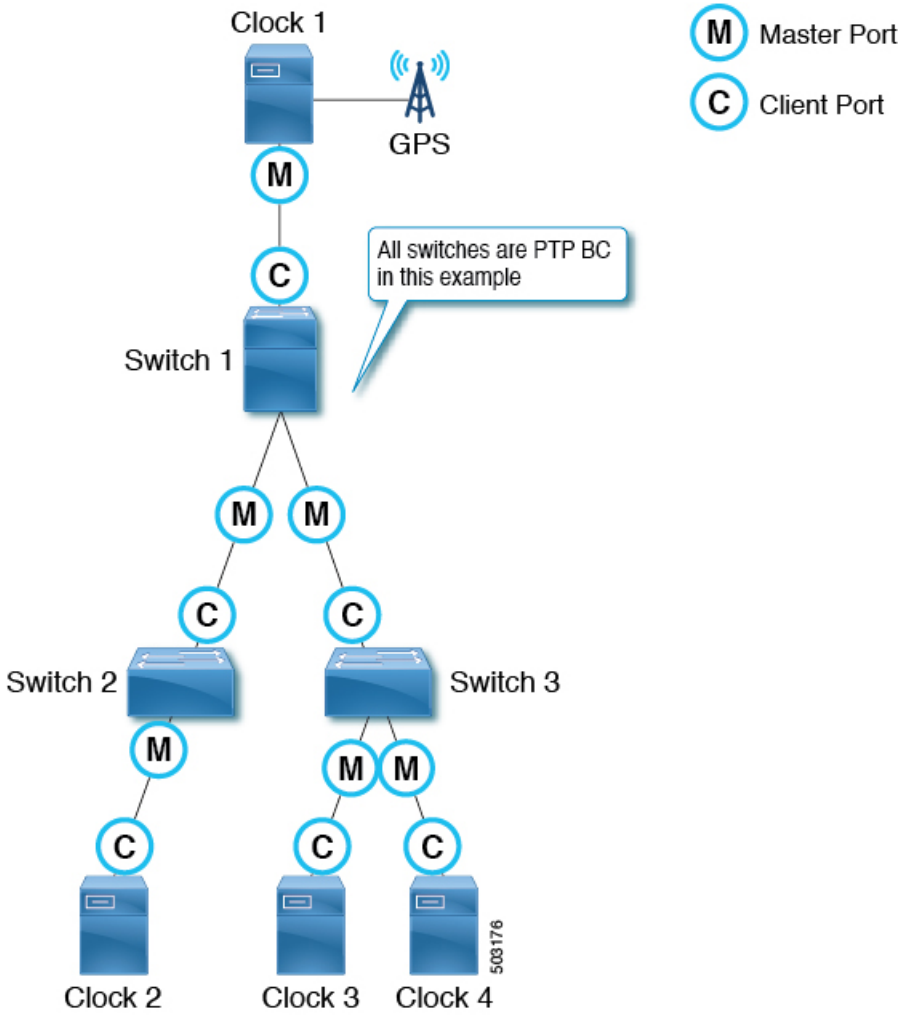
PTP has the following clock types:

Type	Description
Grandmaster Clock (GM, GMC)	The source of time for the entire PTP topology. The grandmaster clock is selected by the Best Master Clock Algorithm (BMCA).

Type	Description
Boundary Clock (BC)	A device with multiple PTP ports. A PTP boundary clock participates in the BMCA and each port has a status, such as master or client. A boundary clock synchronizes with its parent/master so that the client clocks behind itself synchronize to the PTP boundary clock itself. To ensure that, a boundary clock terminates PTP messages and replies by itself instead of forwarding the messages. This eliminates the delay caused by the node forwarding PTP messages from one port to another.
Transparent Clock (TC)	A device with multiple PTP ports. A PTP transparent clock does not participate in the BMCA. This clock type only transparently forwards PTP messages between the master clock and client clocks so that they can synchronize directly with one another. A transparent clock appends the residence time to the PTP messages passing by so that the clients can take the forwarding delay within the transparent clock device into account. In the case of a peer-to-peer delay mechanism, a PTP transparent clock terminates PTP Pdelay_xxx messages instead of forwarding the messages. **Note** Switches in the ACI mode cannot be a transparent clock.
Ordinary Clock (OC)	A device that may serve a source of time as a grandmaster clock or that may synchronize to another clock (such as a master) with the role as a client (a PTP client).

PTP Topology

Master and Client Ports



The master and client ports work as follows:

- Each PTP node directly or indirectly synchronizes its clock to the grandmaster clock that has the best source of time, such as GPS (Clock 1 in the figure).
- One grandmaster is selected for the entire PTP topology (domain) based on the Best Master Clock Algorithm (BMCA). The BMCA is calculated on each PTP node individually, but the algorithm makes sure that all nodes in the same domain select the same clock as the grandmaster.
- In each path between PTP nodes, based on the BMCA, there will be one master port and at least one client port. There will be multiple client ports if the path is point-to-multipoints, but each PTP node can have only one client port. Each PTP node uses its client port to synchronize to the master port on the other end. By repeating this, all PTP nodes eventually synchronize to the grandmaster directly or indirectly.
    - From Switch 1's point of view, Clock 1 is the master and the grandmaster.

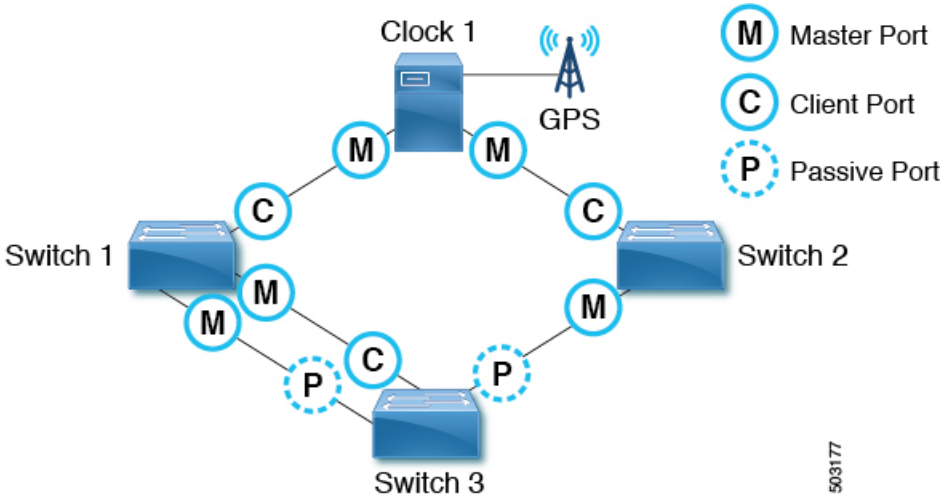
- From Switch 2's point of view, Switch 1 is the master and Clock 1 is the grandmaster.

- Each PTP node should have only one client port, behind which exists the grandmaster. The grandmaster can be multiple hops away.
- The exception is a PTP transparent clock, which does not participate in BMCA. If Switch 3 was a PTP transparent clock, the clock would not have a port status, such as master and client. Clock 3, Clock 4, and Switch 1 would establish a master and client relationship directly.

Passive Ports

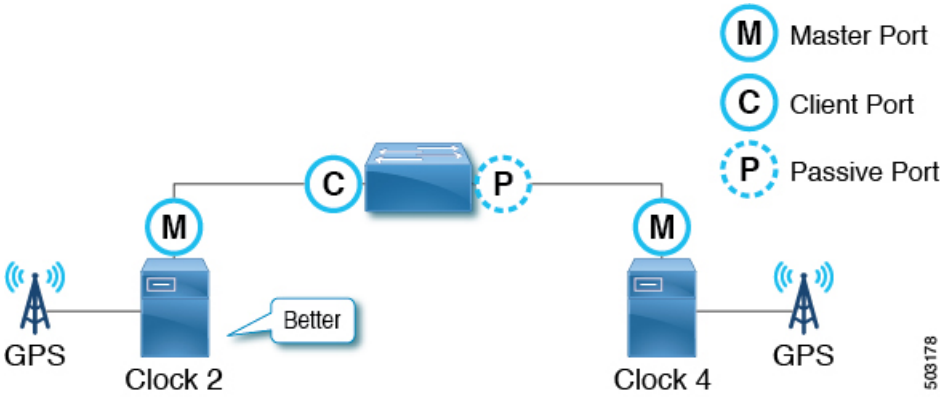
The BMCA can select another PTP port that is in the passive state on top of the master and client. A passive port does not generate any PTP messages, with a few exceptions such as PTP `Management` messages as a response to `Management` messages from other nodes.

Example 1



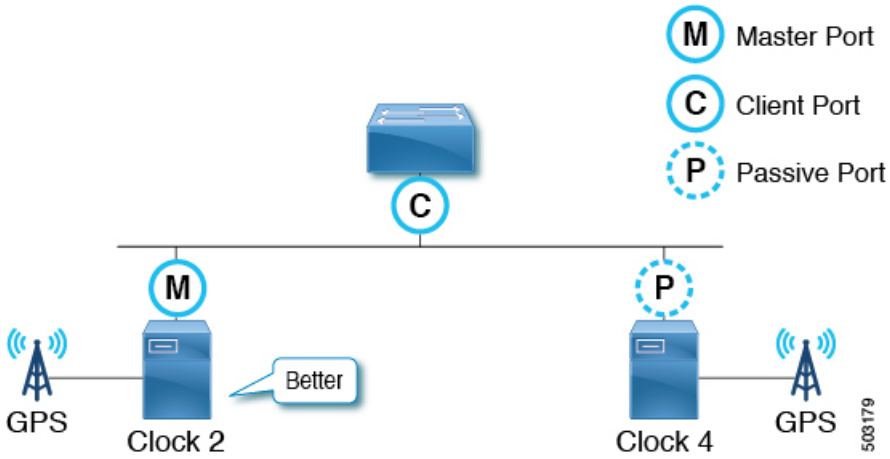
If a PTP node has multiple ports towards the grandmaster, only one of them will be the client port. The other ports toward the grandmaster will be passive ports.

Example 2



If a PTP node detects two master only clocks (grandmaster candidates), the port toward the candidate selected as the grandmaster becomes a client port and the other becomes a passive port. If the other clock can be a client, it forms a master and client relation instead of passive.

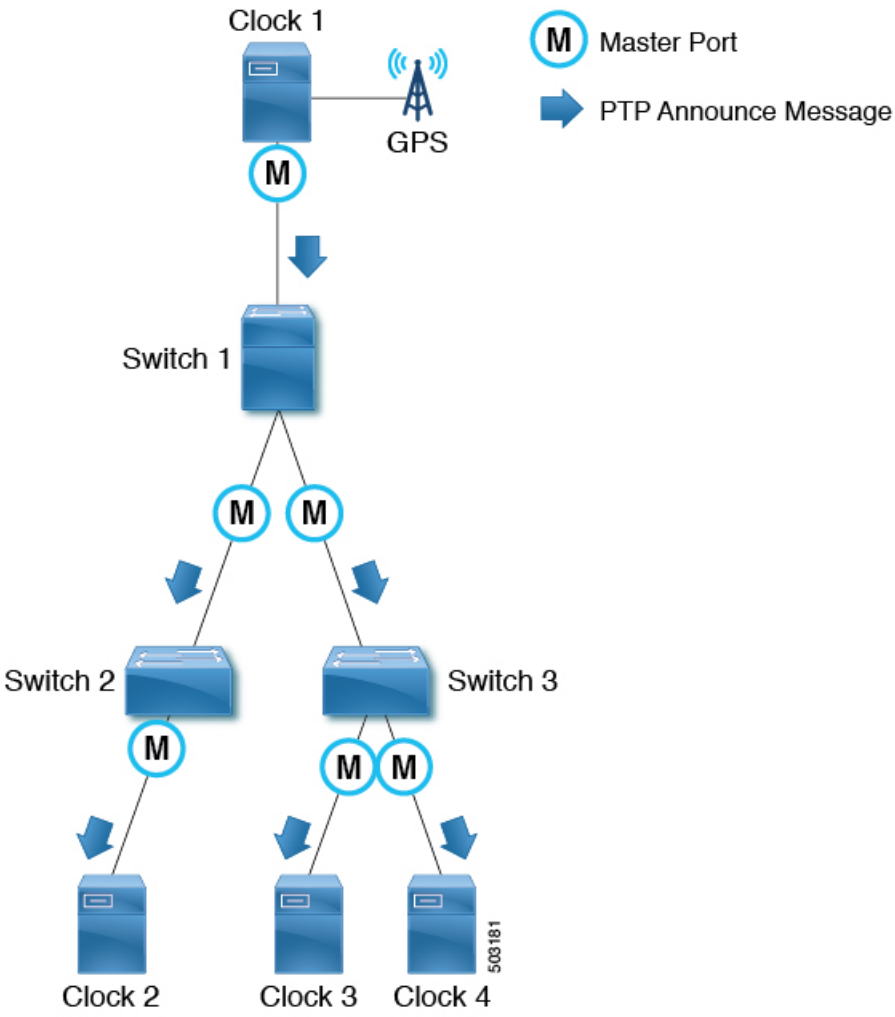
Example 3



If a master-only clock (grandmaster candidate) detects another master-only clock that is better than itself, the clock puts itself in a passive state. This happens when two grandmaster candidates are on the same communication path without a PTP boundary clock in between.

Announce Messages

The `Announce` message is used to calculate the Best Master Clock Algorithm (BMCA) and establish the PTP topology (master-client hierarchy).



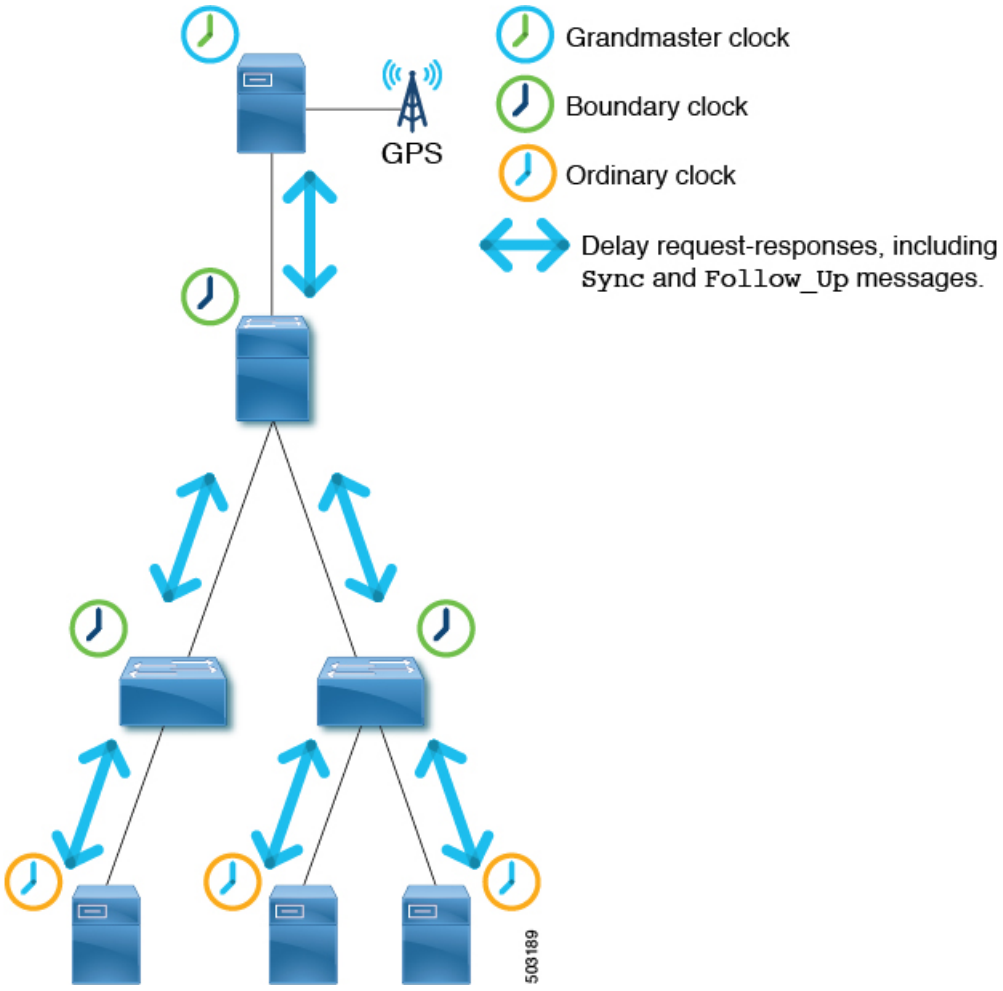
The message works as follows:

- PTP master ports send PTP `Announce` messages to IP address 224.0.1.129 in the case of PTP over IPv4 UDP.
- Each node uses information in the PTP `Announce` messages to automatically establish the synchronization hierarchy (master/client relations or passive) based on the BMCA.
- Some of the information that PTP `Announce` messages contain is as follows:
  - Grandmaster priority 1
  - Grandmaster clock quality (class, accuracy, variance)
  - Grandmaster priority 2
  - Grandmaster identity
  - Step removed
- PTP `Announce` messages are sent with an interval based on $2^{\text{logAnnounceInterval}}$ seconds.

PTP Topology With Various PTP Node Types

PTP Topology With Only End-to-End Boundary Clocks

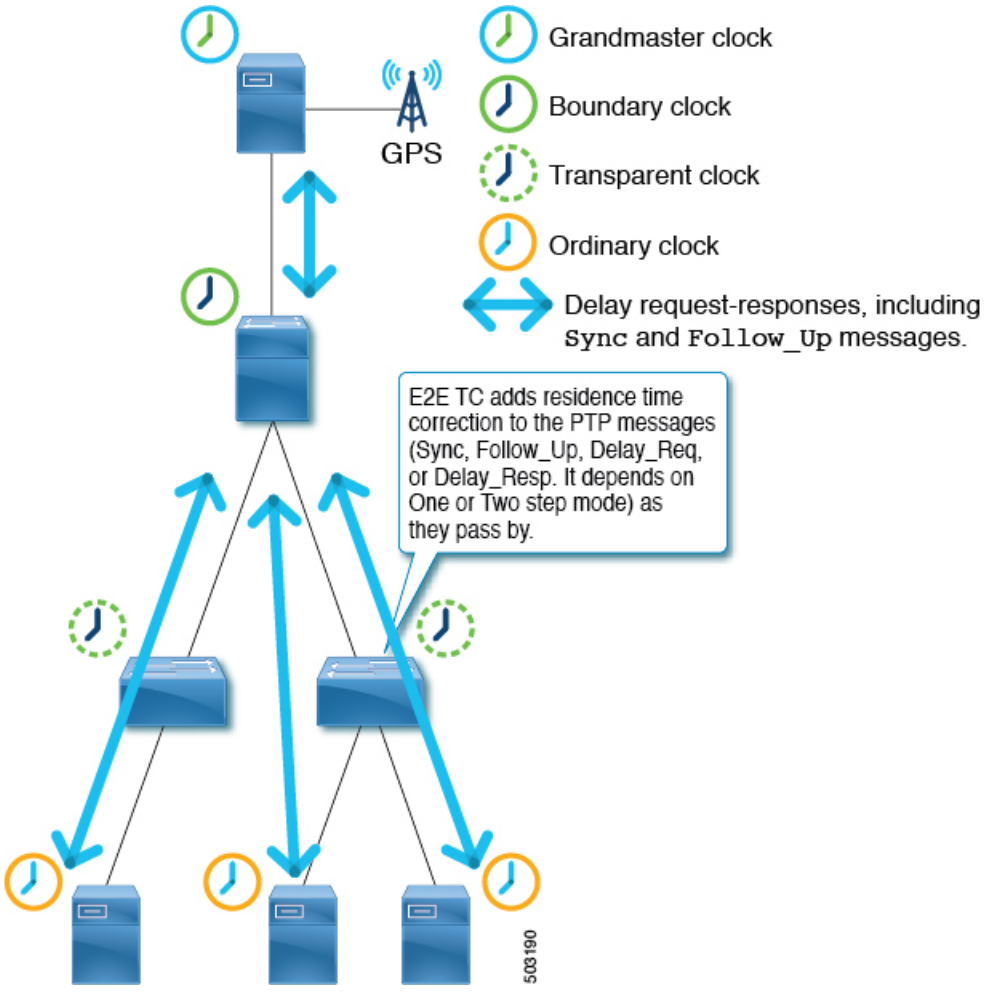
In this topology, the boundary clock nodes terminate all multicast PTP messages, except for `Management` messages.



This ensures that each node processes the `Sync` messages from the closest parent master clock, which helps the nodes to achieve high accuracy.

PTP Topology With a Boundary Clock and End-to-End Transparent Clocks

In this topology, the boundary clock nodes terminate all multicast PTP messages, except for `Management` messages.



End-to-end (E2E) transparent clock nodes do not terminate PTP messages, but simply add a residence time (the time the packet took to go through the node) in the PTP message correction field as the packets pass by so that clients can use them to achieve better accuracy. But, this has lower scalability as the number of PTP messages that need to be handled by one boundary clock node increases.

PTP BMCA

PTP BMCA Parameters

Each clock has the following parameters defined in IEEE 1588-2008 that are used in the Best Master Clock Algorithm (BMCA):

Order	Parameter	Possible Values	Description
1	Priority 1	0 to 255	A user configurable number. The value is normally 128 or lower for grandmaster-candidate clocks (master-capable devices) and 255 for client-only devices.

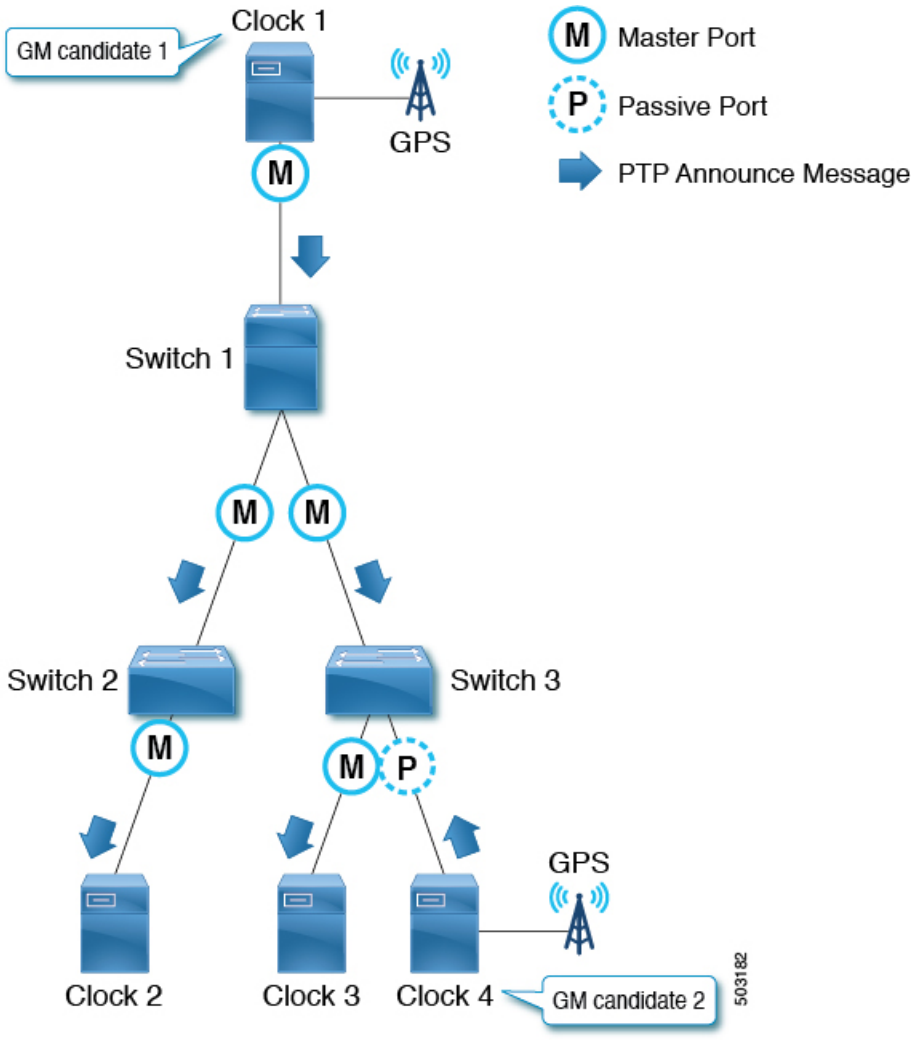
Order	Parameter	Possible Values	Description
2	Clock Quality - Class	0 to 255	Represents the status of the clock devices. For example, 6 is for devices with a primary reference time source, such as GPS. 7 is for devices that used to have a primary reference time source. 127 or lower are for master-only clocks (grandmaster candidates). 255 is for client-only devices.
3	Clock Quality - Accuracy	0 to 255	The accuracy of the clock. For example, 33 (0x21) means < 100 ns, while 35 (0x23) means < 1 us.
4	Clock Quality - Variance	0 to 65535	The precision of the timestamps encapsulated in the PTP messages.
5	Priority 2	0 to 255	Another user-configurable number. This parameter is typically used when the setup has two grandmaster candidates with identical clock quality and one is a standby.
6	Clock Identity	This is an 8-byte value that is typically formed using a MAC address	This parameter serves as the final tie breaker, and is typically a MAC address.
7	Steps Removed	Not configurable	This parameter represents the number of hops from the announced clock and is the last tie breaker when receiving the clock of the same grandmaster from two different ports. If the steps removed is the same for the candidates, the port ID and number is used as a tiebreaker. You cannot configure the value of this parameter.

These parameters of the grandmaster clock are carried by the PTP `Announce` messages. Each PTP node compares these values in the order as listed in the table from all `Announce` messages that the node receives and also the node's own values. For all parameters, the lower number wins. Each PTP node will then create `Announce` messages using the parameters of the best clock among the ones the node is aware of, and the node will send the messages from its own master ports to the next client devices.

Note For more information about each parameter, see clause 7.6 in IEEE 1588-2008.

PTP BMCA Examples

In the following example, Clock 1 and Clock 4 are the grandmaster candidates for this PTP domain:



Clock 1 has the following parameter values:

Parameter	Value
Priority 1	127
Clock Quality - Class	6
Clock Quality - Accuracy	0x21 (< 100ns)
Clock Quality - Variance	15652
Priority 2	128
Clock Identity	0000.1111.1111
Step Removed	*

Clock 4 has the following parameter values:

Parameter	Value
Priority 1	127
Clock Quality - Class	6
Clock Quality - Accuracy	0x21 (< 100ns)
Clock Quality - Variance	15652
Priority 2	129
Clock Identity	0000.1111.2222
Step Removed	*

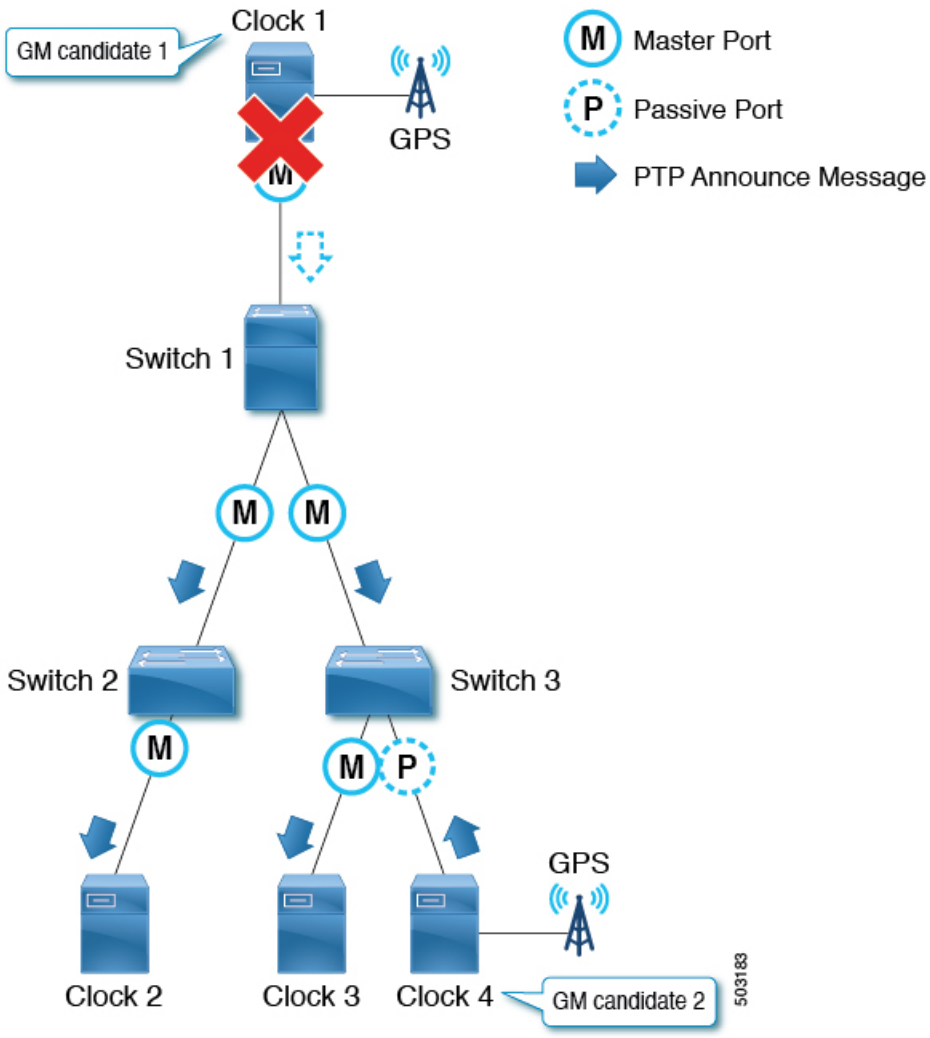
Both clocks send PTP `Announce` messages, then each PTP node compares the values in the messages. In this example, because the first four parameters have the same value, `Priority 2` decides the active grandmaster, which is Clock 1.

After all switches (1, 2, and 3) have recognized that Clock 1 is the best master clock (that is, Clock 1 is the grandmaster), those switches send PTP `Announce` messages with the parameters of Clock 1 from their master ports. On Switch 3, the port connected to Clock 4 (a grandmaster candidate) becomes a passive port because the port is receiving PTP `Announce` messages from a master-only clock (class 6) with parameters that are not better than the current grandmaster that is received by another port.

The `Step Removed` parameter indicates the number of hops (PTP boundary clock nodes) from the grandmaster. When a PTP boundary clock node sends PTP `Announce` messages, it increments the `Step Removed` value by 1 in the message. In this example, Switch 2 receives the PTP `Announce` message from Switch 1 with the parameters of Clock 1 and a `Step Removed` value of `1`. Clock 2 receives the PTP `Announce` message with a `Step Removed` value of `2`. This value is used only when all the other parameters in the PTP `Announce` messages are the same, which happens when the messages are from the same grandmaster candidate clock.

PTP BMCA Failover

If the current active grandmaster (Clock 1) becomes unavailable, each PTP port recalculates the Best Master Clock Algorithm (BMCA).



The availability is checked using the `Announce` messages. Each PTP port declares the timeout of the `Announce` messages after the `Announce` messages were consecutively missing for `Announce Receipt Timeout` times. In other words, for `Announce Receipt Timeout` x $2^{\text{logAnnounceInterval}}$ seconds. This timeout period should be uniform throughout a PTP domain as mentioned in Clause 7.7.3 in IEEE 1588-2008. When the timeout is detected, each switch starts recalculating the BMCA on all PTP ports by sending `Announce` messages with the new best master clock data. The recalculation can result in a switch initially determining that the switch itself is the best master clock, because most of the switches are aware of only the previous grandmaster.

When the client port connected toward the grandmaster goes down, the node (or the ports) does not need to wait for the announce timeout and can immediately start re-calculating the BMCA by sending `Announce` messages with the new best master clock data.

The convergence can take several seconds or more depending on the size of the topology, because each PTP port recalculates the BMCA from the beginning individually to find the new best clock. Prior to the failure of the active grandmaster, only Switch 3 knows about Clock 4, which should take over the active grandmaster role.

Also, when the port status changes to master from non-master, the port changes to the `PRE_MASTER` status first. The port takes `Qualification Timeout` seconds for the port to become the actual master, which is typically equal to:

```
(Step Removed + 1) x the announce interval
```

This means that if the other grandmaster candidate is connected to the same switch as (or close to) the active grandmaster, the port status changes will be minimum and the convergence time will be shorter. See Clause 9.2 in IEEE 1588-2008 for details.

PTP Alternate BMCA (G.8275.1)

The PTP Telecom profile (G.8275.1) uses the alternate Best Master Clock Algorithm (BMCA) defined in G.8275.1, which has a different algorithm than the regular BMCA defined in IEEE 1588-2008. One of the biggest differences is that if there are two grandmaster candidates with the same quality, the alternate BMCA from G.8275.1 allows each PTP node to pick the closest grandmaster instead of forcing all PTP nodes to pick the same clock as the grandmaster by comparing `Steps Removed` before `Clock Identity`. Another difference is the new parameter Local `Priority`, which provides users with manual control over which port to be preferred as the client port. This makes it easier to select the same port as the source for both the PTP Telecom profile and SyncE on each node, which is often preferred for the hybrid mode operation.

PTP Alternate BMCA Parameters

Each clock has the following parameters defined in G.8275.1 that are used in the alternate Best Master Clock Algorithm (BMCA) for the PTP Telecom profile (G.8275.1):

Order	Parameter	Possible Values	Description
1	Clock Quality - Class	0 to 255	Represents the status of the clock devices. For example, 6 is for devices with a primary reference time source, such as GPS. 7 is for devices that used to have a primary reference time source. 127 and lower are for master-only clocks (grandmaster candidates). 255 is for client-only devices.
2	Clock Quality - Accuracy	0 to 255	The accuracy of the clock. For example, 33 (0x21) means < 100 ns, while 35 (0x23) means < 1 us.
3	Clock Quality - Variance	0 to 65535	The precision of the timestamps encapsulated in the PTP messages.
4	Priority 2	0 to 255	A user-configurable number. This parameter is typically used when the setup has two grandmaster candidates with identical clock quality and one is a standby.
5	Local Priority	1 to 255	The clock of the node itself uses the clock local priority configured on the node. A clock received from another node is given the local priority configured for incoming port.

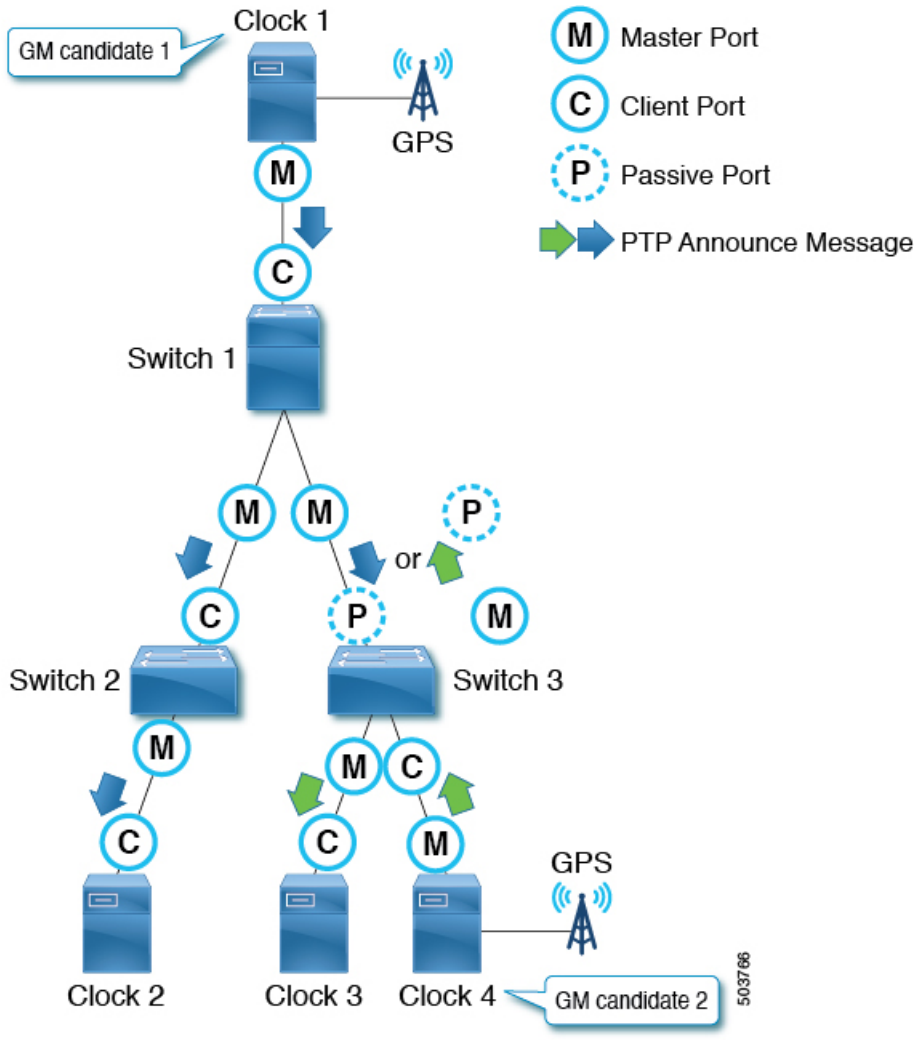
Order	Parameter	Possible Values	Description
6	Steps Removed	Not configurable	This parameter represents the number of hops from the announced clock. The comparison of this allows each Telecom boundary clock to synchronize with a different and closer grandmaster when there are multiple active grandmaster candidates. If the steps removed is the same for the candidates, the port ID and number is used as a tiebreaker. This comparison is performed only when the `Clock Quality - Class` value is 127 or less, which indicates that the clock is a grandmaster candidate.
7	Clock Identity	This is an 8-byte value that is typically formed using a MAC address	This parameter serves as the tie breaker when the `Clock Quality - Class` value is greater than 127, which indicates that the quality of the clock is not designed to be a grandmaster. The value is typically a MAC address.
8	Steps Removed	Not configurable	This parameter represents the number of hops from the announced clock and is the last tie breaker when receiving the clock of the same grandmaster from two different ports. If the steps removed is the same for the candidates, the port ID and number is used as a tiebreaker.

These parameters of the grandmaster clock, except for `Local Priority`, are carried by the PTP `Announce` messages. Each PTP node compares these values in the order as listed in the table from all `Announce` messages that the node receives and also the node's own values. For all parameters, the lower number wins. Each PTP node will then create `Announce` messages using the parameters of the best clock among the ones the node is aware of, and the node will send the messages from its own master ports to the next client devices.

Note For more information about each parameter, see clause 6.3 in G.8275.1.

PTP Alternate BMCA Examples

In the following example, Clock 1 and Clock 4 are the grandmaster candidates for this PTP domain with the same quality and priority:



Clock 1 has the following parameter values:

Parameter	Value
Clock Quality - Class	6
Clock Quality - Accuracy	0x21 (< 100ns)
Clock Quality - Variance	15652
Priority 2	128
Steps Removed	*
Clock Identity	0000.1111.1111

Clock 4 has the following parameter values:

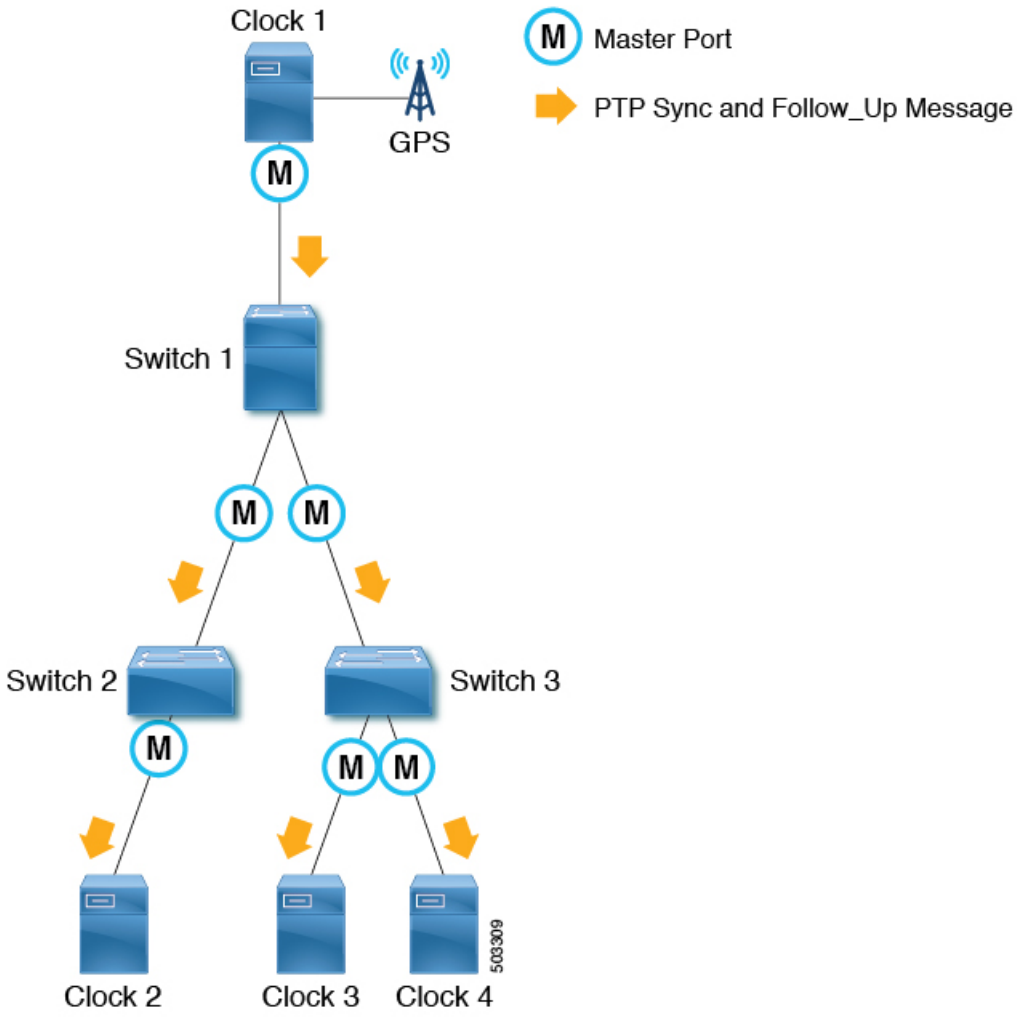
Parameter	Value
Clock Quality - Class	6
Clock Quality - Accuracy	0x21 (< 100ns)
Clock Quality - Variance	15652
Priority 2	128
Steps Removed	*
Clock Identity	0000.1111.2222

Both Clock 1 and Clock 4 send PTP `Announce` messages, then each PTP node compares the values in the messages. Because the values for the `Clock Quality - Class` through `Priority 2` parameters are the same, `Steps Removed` decides the active grandmaster for each PTP node.

For Switch 1 and 2, Clock 1 is the grandmaster. For Switch 3, Clock 4 is the grandmaster.

PTP Clock Synchronization

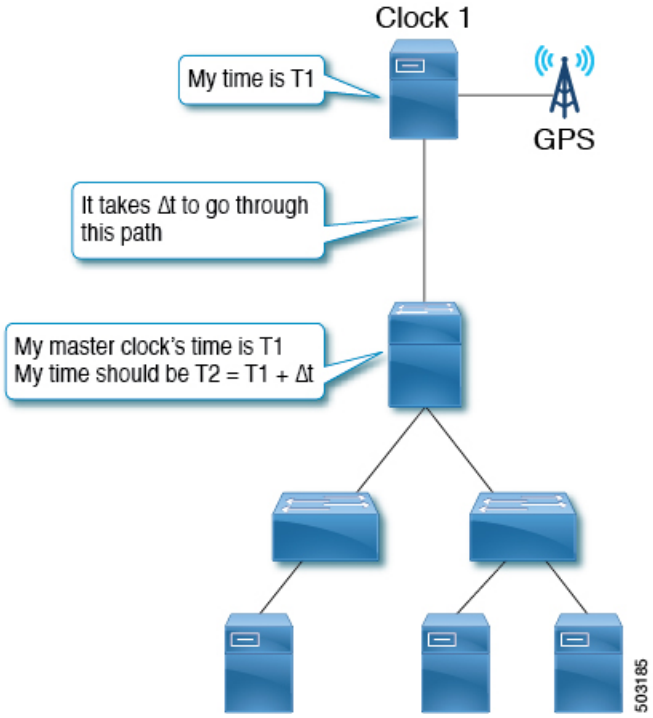
The PTP master ports send PTP `Sync` and `Follow_Up` messages to IP address 224.0.1.129 in the case of PTP over IPv4 UDP.



In One-Step mode, the `Sync` messages carry the timestamp of when the message was sent out. `Follow_Up` messages are not required. In Two-Step mode, `Sync` messages are sent out without a timestamp. `Follow_Up` messages are sent out immediately after each `Sync` message with the timestamp of when the `Sync` message was sent out. Client nodes use the timestamp in the `Sync` or `Follow_Up` messages to synchronize their clock along with an offset calculated by `meanPathDealy`. `Sync` messages are sent with the interval based on $2^{\text{logSyncInterval}}$ seconds.

PTP and meanPathDelay

`meanPathDelay` is the mean time that PTP packets take to reach from one end of the PTP path to the other end. In the case of the E2E delay mechanism, this is the time taken to travel between a PTP master port and a client port. PTP needs to calculate `meanPathDelay` (Δt in the following illustration) to keep the synchronized time on each of the distributed devices accurate.



There are two mechanisms to calculate `meanPathDelay`:

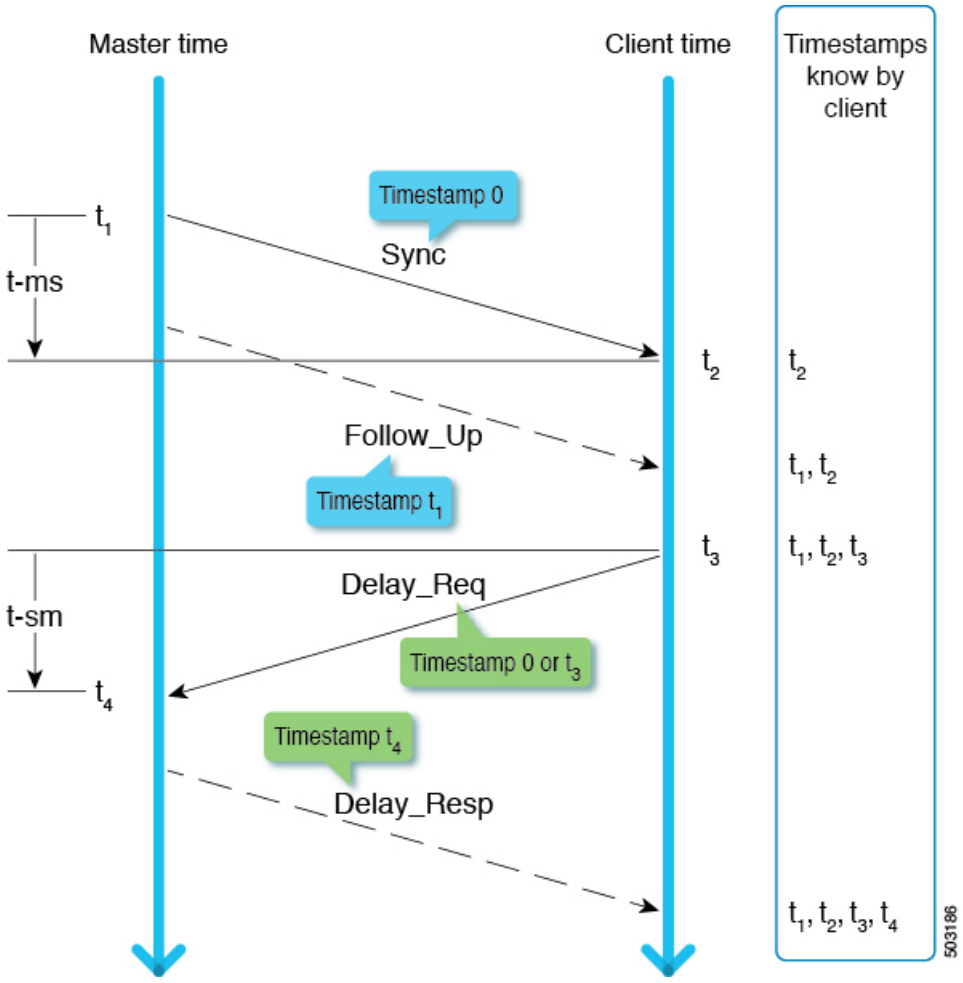
- Delay Request-Response (E2E): End-to-end transparent clock nodes can only support this.
- Peer Delay Request-Response (P2P): Peer-to-peer transparent clock nodes can only support this.

Boundary clock nodes can support both mechanisms by definition. In IEEE 1588-2008, the delay mechanisms are called "Delay" or "Peer Delay." However, the Delay Request-Response mechanism is more commonly referred to as the "E2E delay mechanism," and the Peer Delay mechanism is more commonly referred to as the "P2P delay mechanism."

meanPathDelay Measurement

Delay Request-Response

The delay request-response (E2E) mechanism is initiated by a client port and the `meanPathDelay` is measured on the client node side. The mechanism uses `Sync` and `Follow_Up` messages, which are sent from a master port regardless of the E2E delay mechanism. The `meanPathDelay` value is calculated based on 4 timestamps from 4 messages.



t-ms (t2 – t1) is the delay for master to client direction. t-sm (t4 – t3) is the delay for client to master direction. `meanPathDelay` is calculated as follows:

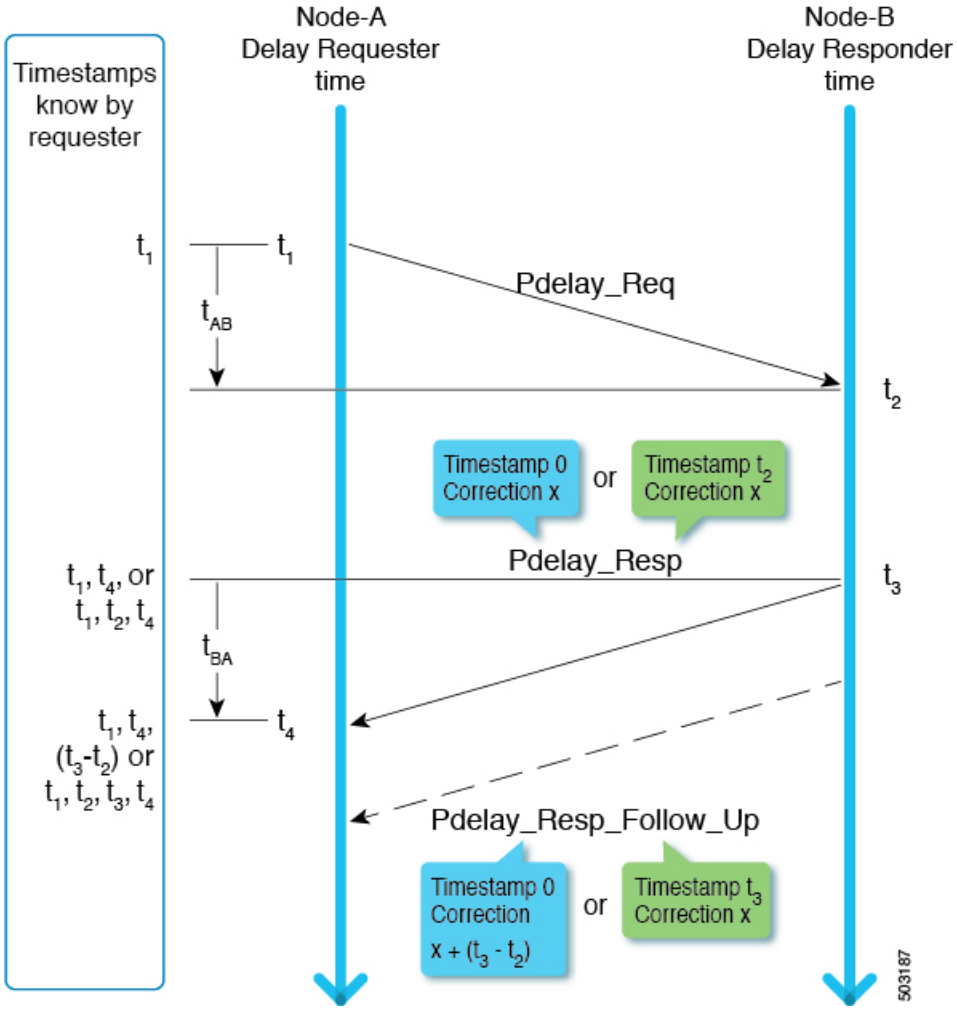
```
(t-ms + t-sm) / 2
```

`Sync` is sent with the interval based on $2^{logSyncInterval}$ sec. `Delay_Req` is sent with the interval based on $2^{logMinDelayReqInterval}$ sec.

Note This example focuses on Two-Step mode. See Clause 9.5 from IEEE 1588-2008 for details on the transmission timing.

Peer Delay Request-Response

The peer delay request-response (P2P) mechanism is initiated by both master and client port and the `meanPathDelay` is measured on the requester node side. `meanPathDelay` is calculated based on 4 timestamps from 3 messages dedicated for this delay mechanism.



In the two-step mode, t2 and t3 are delivered to the requester in one of the following ways:

- As (t3-t2) using `Pdelay_Resp_Follow_Up`
- As t2 using `Pdelay_Resp` and as t3 using `Pdelay_Resp_Follow_Up`

`meanPathDelay` is calculated as follows:

```
(t4-t1) - (t3-t1) / 2
```

`Pdelay_Req` is sent with the interval based on $2^{logMinPDelayReqInterval}$ seconds.

Note Cisco Application Centric Infrastructure (ACI) switches do not support the peer delay request-response (P2P) mechanism.

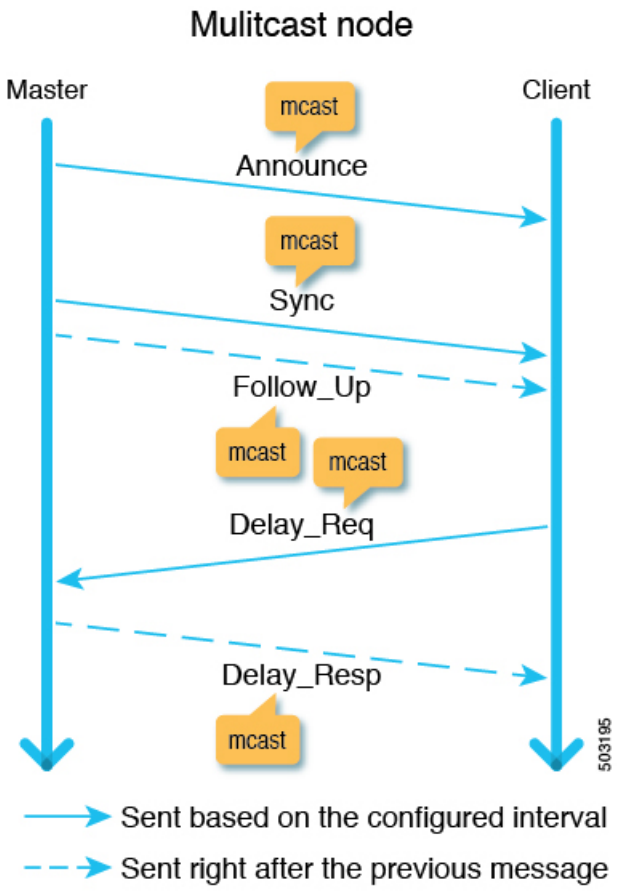
See clause 9.5 from IEEE 1588-2008 for details on the transmission timing.

PTP Multicast, Unicast, and Mixed Mode

The following sections describe the different PTP modes using the delay request-response (E2E delay) mechanism.

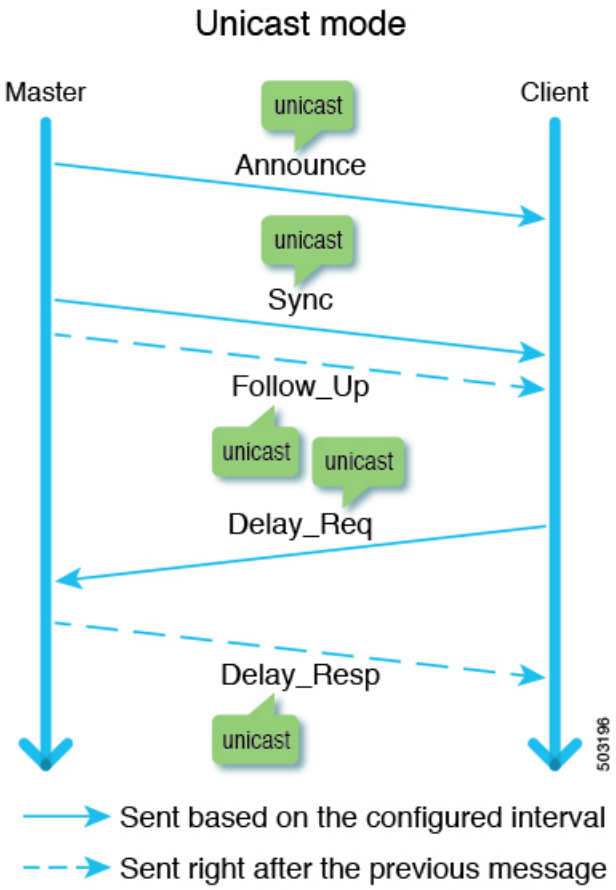
Multicast Mode

All PTP messages are multicast. Transparent clock or PTP unaware nodes between the master and clients result in inefficient flooding of the `Delay` messages. However, the flooding is efficient for `Announce`, `Sync`, and `Follow_Up` messages because these messages should be sent toward all client nodes.



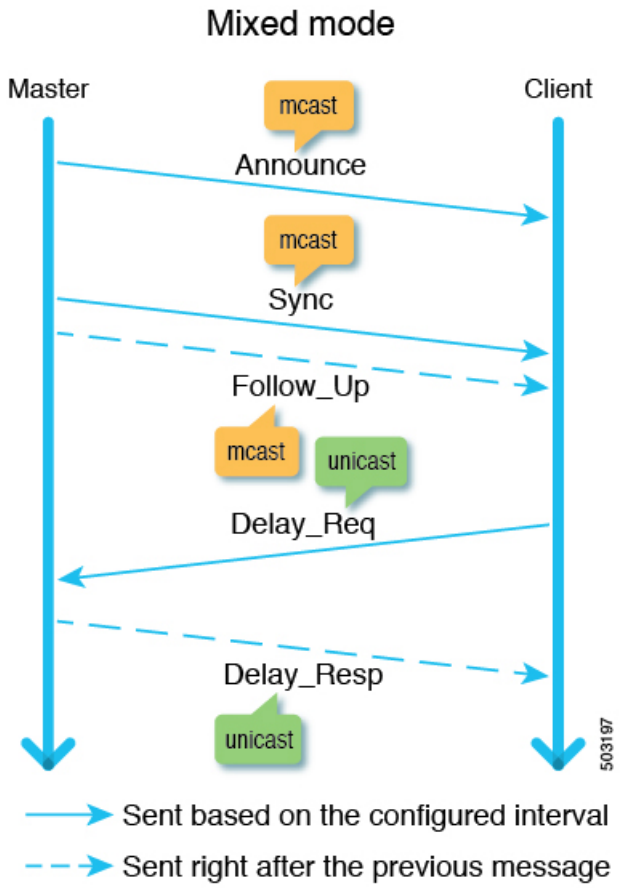
Unicast Mode

All PTP messages are unicast, which increases the number of messages that the master must generate. Hence, the scale, such as the number of client nodes behind one master port, is impacted.



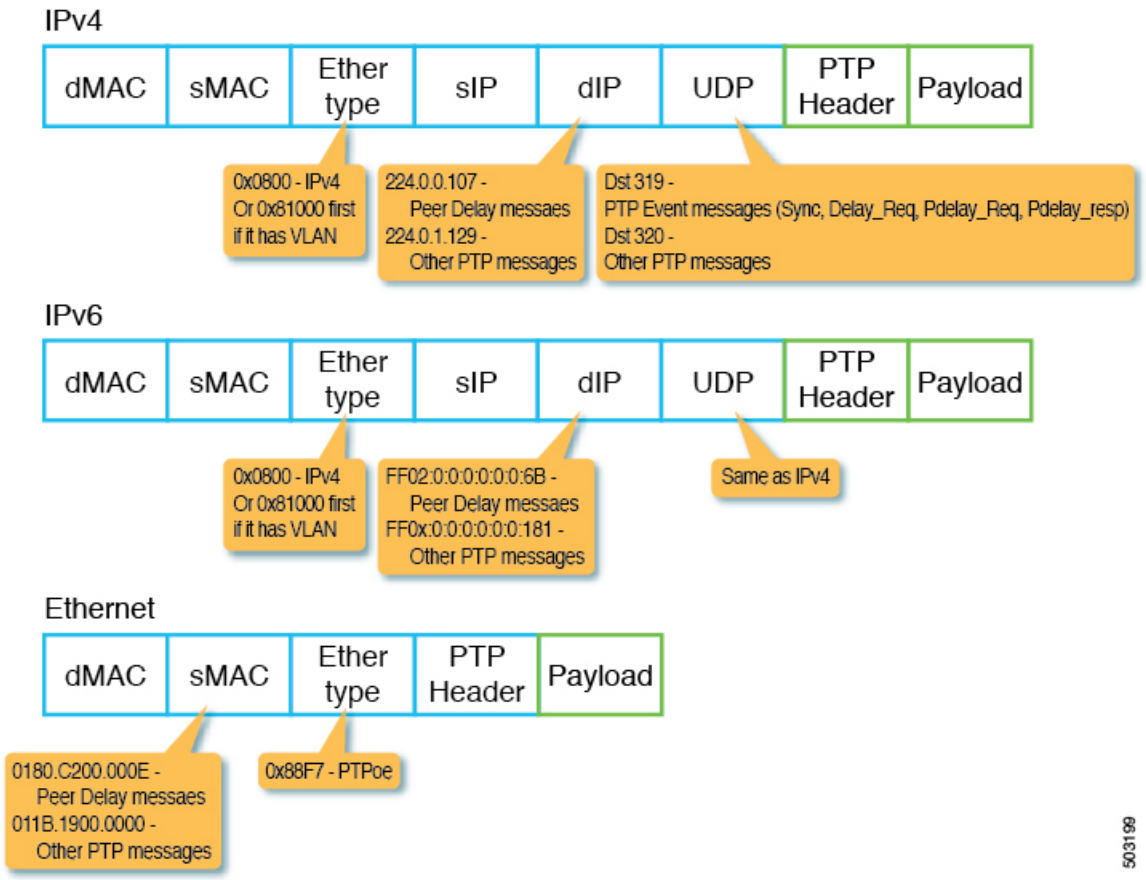
Mixed Mode

Only `Delay` messages are unicast, which resolves the problems that exist in multicast mode and unicast mode.



PTP Transport Protocol

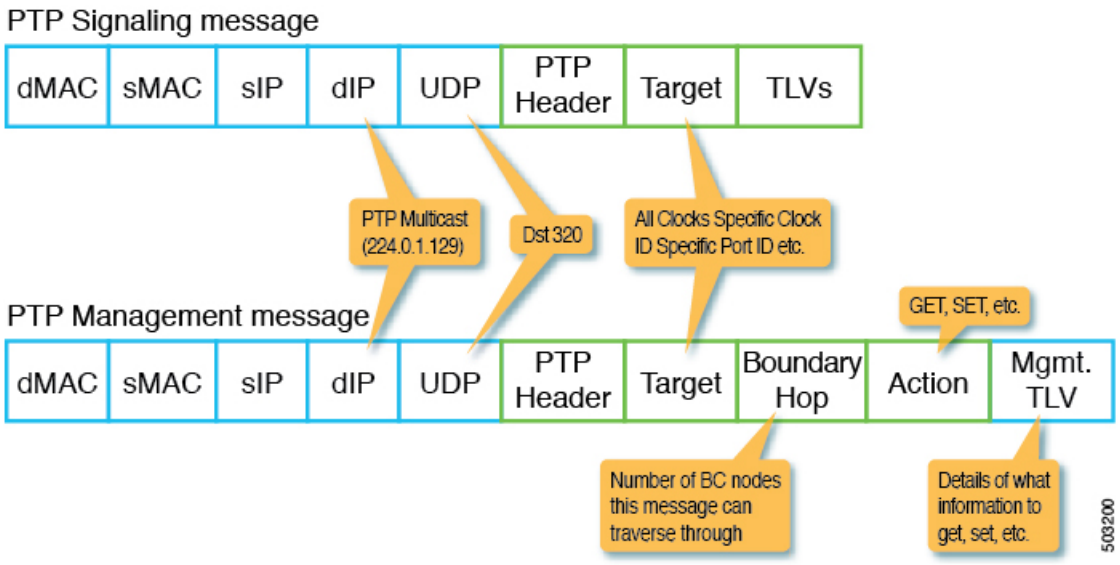
The following illustration provides information about the major transport protocols that PTP supports:



Note Cisco Application Centric Infrastructure (ACI) switches only support IPv4 and Ethernet as a PTP transport protocol.

PTP Signaling and Management Messages

The following illustration shows the `Signaling` and `Management` message parameters in the header packet for PTP over IPv4 UDP:



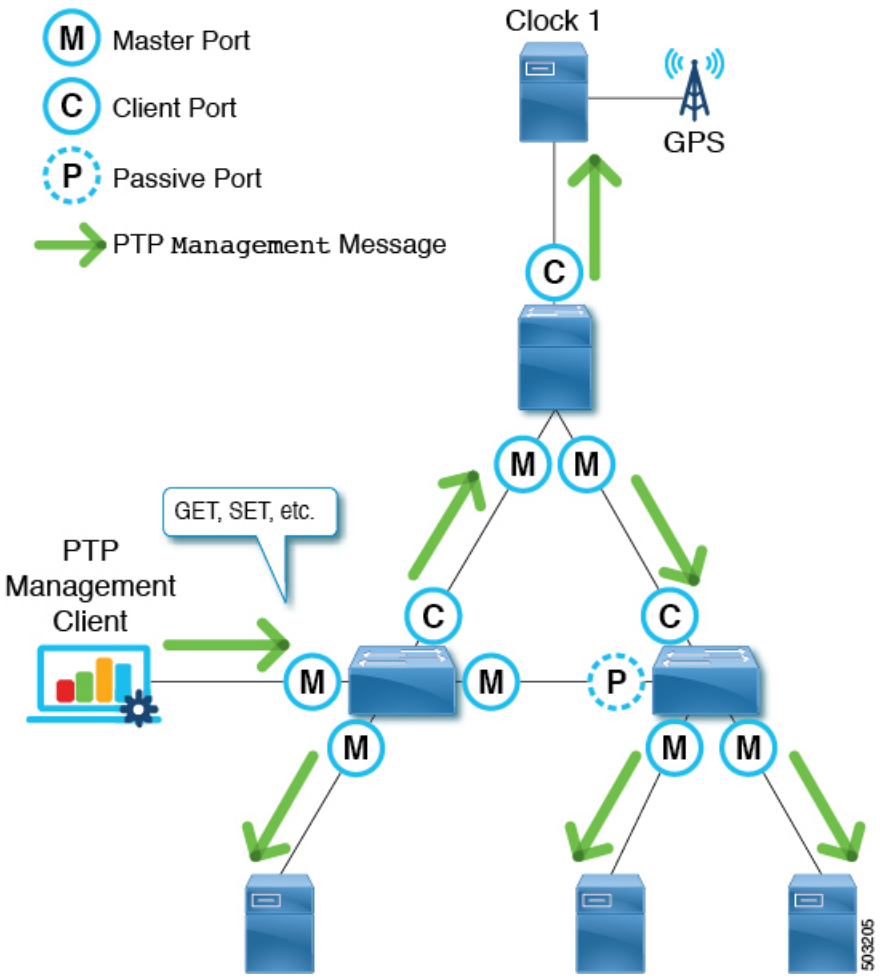
A `Management` message is used to configure or collect PTP parameters, such as the current clock and offset from its master. With the message, a single PTP management node can manage and monitor PTP-related parameters without relying on an out-of-band monitoring system.

A `Signaling` message also provides various types of type, length, and value (TLVs) to do additional operations. There are other TLVs that are used by being appended to other messages. For example, the `PATH_TRACE` TLV as defined in clause 16.2 of IEEE 1588-2008 is appended to `Announce` messages to trace the path of each boundary clock node in the PTP topology.

Note Cisco Application Centric Infrastructure (ACI) switches do not support management, signal, or other optional TLVs.

PTP Management Messages

PTP `Management` messages are used to transport management types, lengths, and values (TLVs) toward multiple PTP nodes at once or to a specific node.



The targets are specified with the targetPortIdentity (clockID and portNumber) parameter. PTP `Management` messages have the actionField that specify actions such as GET, SET, and COMMAND to inform the targets of what to do with the delivered management TLV.

PTP `Management` messages are forwarded by the PTP boundary clock, and only to the Master, Client, Uncalibrated, or Pre_Master ports. A message is forwarded to those ports only when the message is received on a port in the Master, Client, Uncalibrated, or Pre_Master port. BoundaryHops in the message is decremented by 1 when the message is forwarded.

The SMTPE ST2059-2 profile defines that the grandmaster should send PTP `Management` messages using the action COMMAND with the synchronization metadata TLV that is required for the synchronization of audio/video signals.

Note Cisco Application Centric Infrastructure (ACI) switches do not process `Management` messages, but forward them to support the SMTPE ST2059-2 PTP profile.

PTP Profiles

The precision time protocol (PTP) has a concept called the *PTP profile*. A PTP profile is used to define various parameters that are optimized for different use cases of PTP. Some of those parameters include, but not limited to, the appropriate range of PTP message intervals and the PTP transport protocols. A PTP profile is defined by many organizations/standards in different industries. For example:

- IEEE 1588-2008: This standard defines a default PTP profile called the `Default Profile`.
- AES67-2015: This standard defines a PTP profile for audio requirements. This profile is also called the `Media Profile`.
- SMPTE ST2059-2: This standard defines a PTP profile for video requirements.
- ITU-T G.8275.1: Also known as the Telecom profile with Full Timing Support. This standard is recommended for telecommunications with Full Timing Support. Full Timing Support is the term defined by ITU to describe a telecommunication network that can provide devices with the PTP G.8275.1 profile on every hop. G.8275.2, which is not supported by Cisco Application Centric Infrastructure (ACI), is for Partial Timing Support that may have devices in the path that do not support PTP.

  The telecommunication industry requires both frequency and time/phase synchronization. G.8275.1 is used to synchronize time and phase. The frequency can be synchronized either using PTP through the packet network with another PTP G.8265.1 profile, which is not supported by Cisco ACI, or using the physical layer such as the synchronous digital hierarchy (SDH), synchronous optical networking (SONET) through a dedicated circuit, or synchronous Ethernet (SyncE) through Ethernet. Synchronizing the frequency using SyncE and time/phase using PTP is called the *hybrid mode*.

  The key differences of G.8275.1 compared to the other profiles are as follows:

  - G.8275.1 uses the alternate BMCA with the additional parameter Local Priority that does not exist in the other profiles.
  - G.8275.1 uses PTP over Ethernet with all PTP messages using the same destination MAC address (forwardable and non-forwardable), which you can choose.
  - G.8275.1 expects the telecom boundary clock (T-BC) to follow the accuracy (maximum time error; max|TE|) defined by G.8273.2.
    - Class A: 100 ns
    - Class B: 70 ns
    - Class C: 30 ns

The following table shows some of the parameters defined in each standard for each PTP profile:

Profiles	logAnnounce Interval	logSync Interval	logMinDelayReq Interval	announceReceipt Timeout	Domain Number	Mode	Transport Protocol
Default Profile	0 to 4 (1) [= 1 to 16 sec]	-1 to +1 (0) [= 0.5 to 2 sec]	0 to 5 (0) [= 1 to 32 sec]	2 to 10 announce intervals (3)	0 to 255 (0)	Multicast / Unicast	Any/IPv4

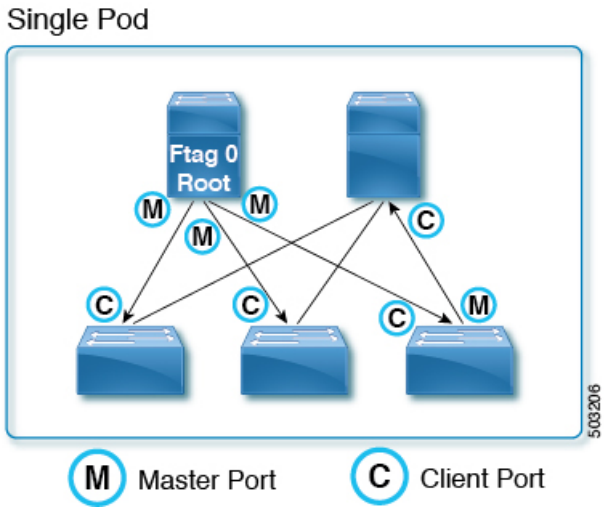
Profiles	logAnnounce Interval	logSync Interval	logMinDelayReq Interval	announceReceipt Timeout	Domain Number	Mode	Transport Protocol
AES67-2015 (Media Profile)	0 to 4 (1) [= 1 to 16 sec]	-4 to +1 (-3) [= 1/16 to 2 sec]	-3 to +5 (0) [= 1/8 to 32 sec] Or *logSyncInterval* to *logSyncInterval* + 5 seconds	2 to 10 announce intervals (3)	0 to 255 (0)	Multicast / Unicast	UDP/IPv4
SMTPE ST2059-2-2015	-3 to +1 (-2) [= 1/8 to 2 sec]	-7 to -1 (-3) [= 1/128 to 0.5 sec]	*logSyncInterval* to *logSyncInterval* + 5 seconds	2 to 10 announce intervals (3)	0 to 127 (127)	Multicast / Unicast	UDP/IPv4
ITU-T G.8275.1	-3	-4	-4	2 to 4	24 to 43 (24)	Multicast Only	Ethernet

Cisco ACI and PTP

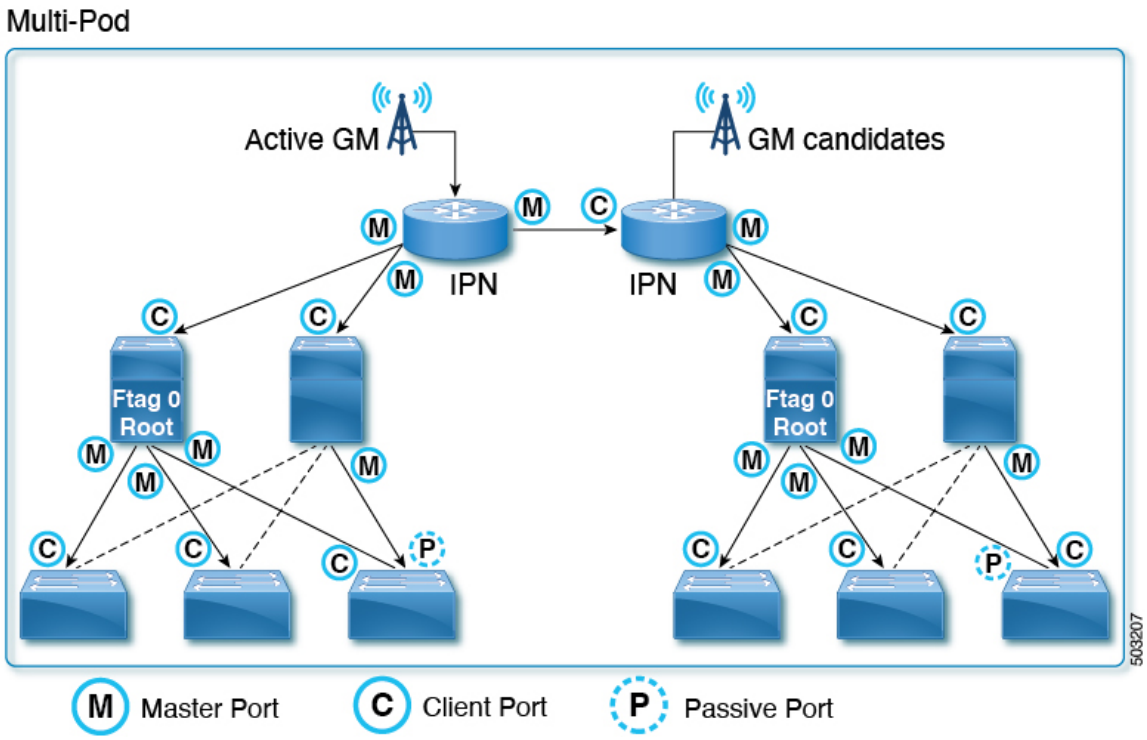
In the Cisco Application Centric Infrastructure (ACI) fabric, when the PTP feature is globally enabled in the Cisco Application Policy Infrastructure Controller (APIC), the software automatically enables PTP on specific interfaces of all the supported spine and leaf switches to establish the PTP master-client topology within the fabric. Starting in Cisco APIC release 4.2(5), you can enable PTP on leaf switch front panel ports and extend PTP topology to outside of the fabric. In the absence of an external grandmaster clock, one of the spine switch is chosen as the grandmaster. The master spine switch is given a different PTP priority that is lower by 1 than the other spines and leaf switches.

Implementation in Cisco APIC Release 3.0(1)

Starting in Cisco Application Policy Infrastructure Controller (APIC) release 3.0(1), PTP was partially introduced to synchronize time only within Cisco Application Centric Infrastructure (ACI) fabric switches. PTP was required to provide the latency measurement feature that was also introduced in Cisco APIC release 3.0(1). For this purpose, a single option was introduced to enable or disable PTP globally. When PTP is enabled globally, all leaf and spine switches are configured as PTP boundary clocks. PTP is automatically enabled on all fabric ports that are used by the ftag tree with ID 0 (ftag0 tree), which is one of the internal tree topologies that is automatically built based on Cisco ACI infra ISIS for loop-free multicast connectivity between all leaf and spine switches in each pod. The root spine switch of the ftag 0 tree is automatically configured with PTP priority1 254 to be the grandmaster when there are no external grandmasters in the inter-pod network (IPN). Other spine and leaf switches are configured with PTP priority1 255.



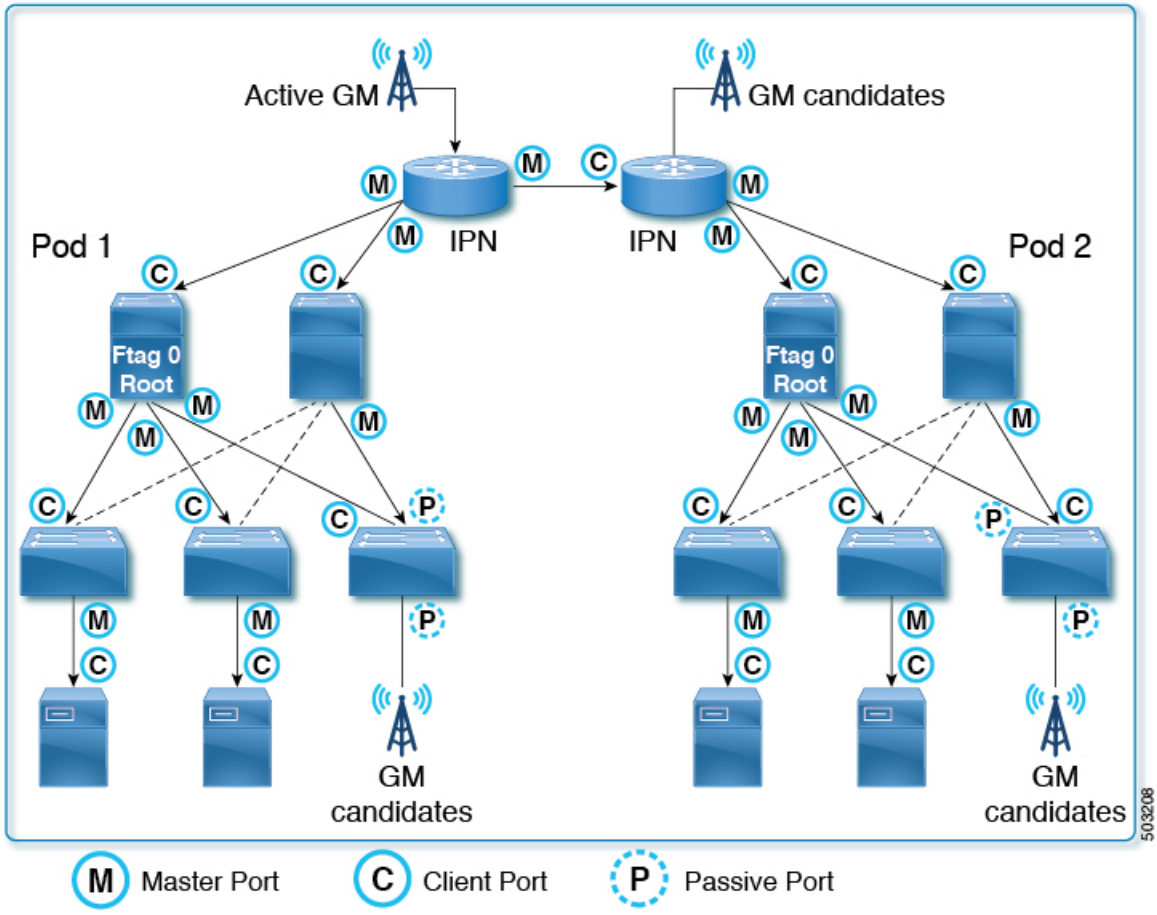
In a Cisco ACI Multi-Pod setup, when PTP is enabled globally, PTP is automatically enabled on the spine sub-interfaces configured for IPN connectivity in the tn-infra Multi-Pod L3Out. Until Cisco APIC release 4.2(5) or 5.1(1), this was the only way to enable PTP on external-facing interfaces. With this, you can have the Cisco ACI fabric look to an external grandmaster through IPN. When a high accuracy is required, we recommend that you have an external grandmaster with a primary reference time source, such as GPS/GNSS. When enabling PTP in a Cisco ACI Multi-Pod setup without an external grandmaster, one of the spine switches can become a grandmaster for all pods assuming PTP is enabled on IPN and IPN's PTP BMCA parameters, such as PTP priorities, are not better than the spine switch's parameters. When using a spine switch as the grandmaster, adding a new pod may unintentionally result in the new grandmaster being selected from the new pod, which can temporarily cause churn in PTP synchronization throughout the fabric. Regardless of external grandmasters, for a better PTP topology that is fewer hops from the grandmaster, we recommend that you connect all spine switches in the fabric to IPN because users do not have control over how ftag0 tree is formed, which decides the PTP topology inside each pod.



In Cisco APIC release 3.0(1), PTP cannot be enabled on any other interfaces on demand, such as the down link (front panel ports) on leaf switches.

Implementation in Cisco APIC Releases 4.2(5) and 5.1(1)

Starting in Cisco APIC releases 4.2(5) and 5.1(1), you can enable PTP on a leaf switch's front panel ports to connect the PTP nodes, clients, or grandmaster. The PTP implementation on fabric ports are still the same as the previous releases, except that the PTP parameters for fabric ports can now be adjusted. With this change, you can use the Cisco ACI fabric to propagate time synchronization using PTP with Cisco ACI switches as PTP boundary clock nodes. Prior to that, the only approach Cisco ACI had was to forward PTP multicast or unicast messages transparently as a PTP unaware switch from one leaf switch to another as a tunnel.



Note The 5.0(x) releases do not support the PTP functionality that was introduced in the 4.2(5) and 5.1(1) releases.

Cisco ACI Software and Hardware Requirements

Supported Software for PTP

The following feature is supported from Cisco Application Policy Infrastructure Controller (APIC) release 3.0(1):

- PTP only within the fabric for the latency measurement feature

The following features are supported from Cisco APIC release 4.2(5):

- PTP with external devices by means of the leaf switches
- PTP on leaf switch front panel ports
- Configurable PTP message intervals
- Configurable PTP domain number

- Configurable PTP priorities
- PTP multicast port
- PTP unicast master port on leaf switch front panel ports
- PTP over IPv4/UDP
- PTP profile (Default, AES67, and SMTPE ST2059-2)

The following features are supported from Cisco APIC release 5.2(1):

- PTP multicast master-only ports
- PTP over Ethernet
- PTP Telecom profile with Full Timing Support (ITU-T G.8275.1)

Supported Hardware for PTP

Leaf switches, spine switches, and line cards with -EX or later in the product ID are supported, such as N9K-X9732C-EX or N9K-C93180YC-FX.

The PTP Telecom profile (G.8275.1) is supported only on the Cisco N9K-C93180YC-FX3 switch. This switch supports Class B (G.8273.2) accuracy when used along with SyncE.

The following leaf switches are not supported:

- N9K-C9332PQ
- N9K-C9372PX
- N9K-C9372PX-E
- N9K-C9372TX
- N9K-C9372TX-E
- N9K-C9396PX
- N9K-C9396TX
- N9K-C93120TX
- N9K-C93128TX

The following spine box switch is not supported:

- N9K-C9336PQ

The following spine switch line card is not supported:

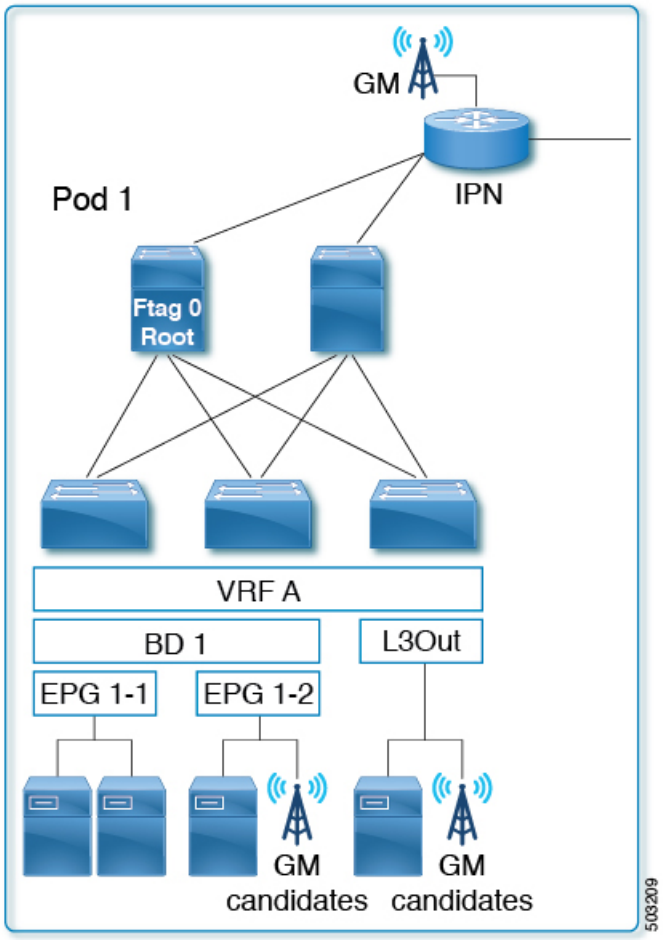
- N9K-X9736PQ

PTP Connectivity

Supported PTP Node Connectivity

External PTP nodes can be connected to the Cisco Application Centric Infrastructure (ACI) fabric using the following methods:

- Inter-pod network
- EPG (on a leaf switch)
- L3Out (on a leaf switch)



PTP is VRF-agnostic, the same as with a standalone NX-OS switch. All PTP messages are terminated, processed, and generated at the interface level on each Cisco ACI switch node as a PTP boundary clock. Regardless of the VRF, bridge domain, EPG, or VLAN, the Best Master Clock Algorithm (BMCA) is calculated across all of the interfaces on each Cisco ACI switch. There is only one PTP domain for the entire fabric.

Any PTP nodes with the E2E delay mechanism (delay req-resp) can be connected to the Cisco ACI switches that are running as a PTP boundary clock.

Note Cisco ACI switches do not support the Peer Delay (P2P) mechanism. Therefore, a P2P transparent clock node cannot be connected to Cisco ACI switches.

Supported PTP Interface Connectivity

Connection Type	Interface Type	Leaf Switch Type (leaf, remote leaf, tier-2 leaf)	Supported / Not Supported (non-Telecom profiles)	Supported / Not Supported (G.8275.1)
Fabric Link (between a leaf and spine switch)	Sub-interface (non-PC)	-	Supported	Not supported
Fabric Link (between a tier-1 and tier-2 leaf switch)	Sub-interface (non-PC)	-	Supported	Not supported
Spine (toward an IPN)	Sub-interface (non-PC)	-	Supported	Not supported
Remote leaf (toward an IPN)	Sub-interface (non-PC)	-	Supported	Not supported
Normal EPG (trunk, access, 802.1P)	Physical, port channel	Any	Supported	Supported
	vPC	Any	Supported	Not supported
L3Out (routed, routed-sub)	Physical, port channel	Any	Supported	Supported
L3Out (SVI – trunk, access, 802.1P)	Physical, port channel, vPC	Any	Not supported	Not supported
L2Out (trunk)	Physical, port channel, vPC	Any	Not supported	Not supported
EPG/L3Out in tn-mgmt	Physical, port channel, vPC	Any	Not supported	Not supported
Service EPG (trunk)[1]	Physical, port channel, vPC	Any	Not supported	Not supported
Any type of FEX interface	Any	Any	Not supported	Not supported
Breakout ports[2]	Any	Any	Not supported	Not supported
Out-of-band management interface	Physical	-	Not supported	Not supported

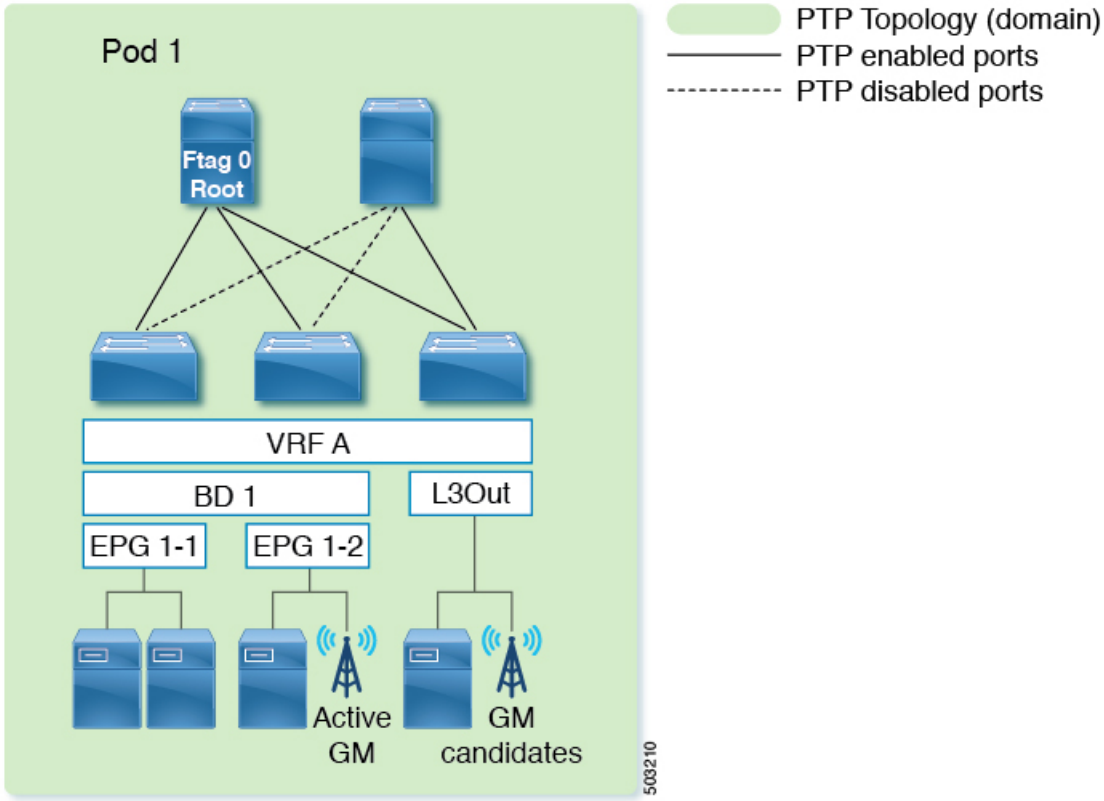
[1] The service EPG is an internal EPG created for a Layer 4 to Layer 7 service graph.
[2] Both fabric links and downlinks.

Grandmaster Deployments

You can deploy the grandmaster candidates using one of the following methods:

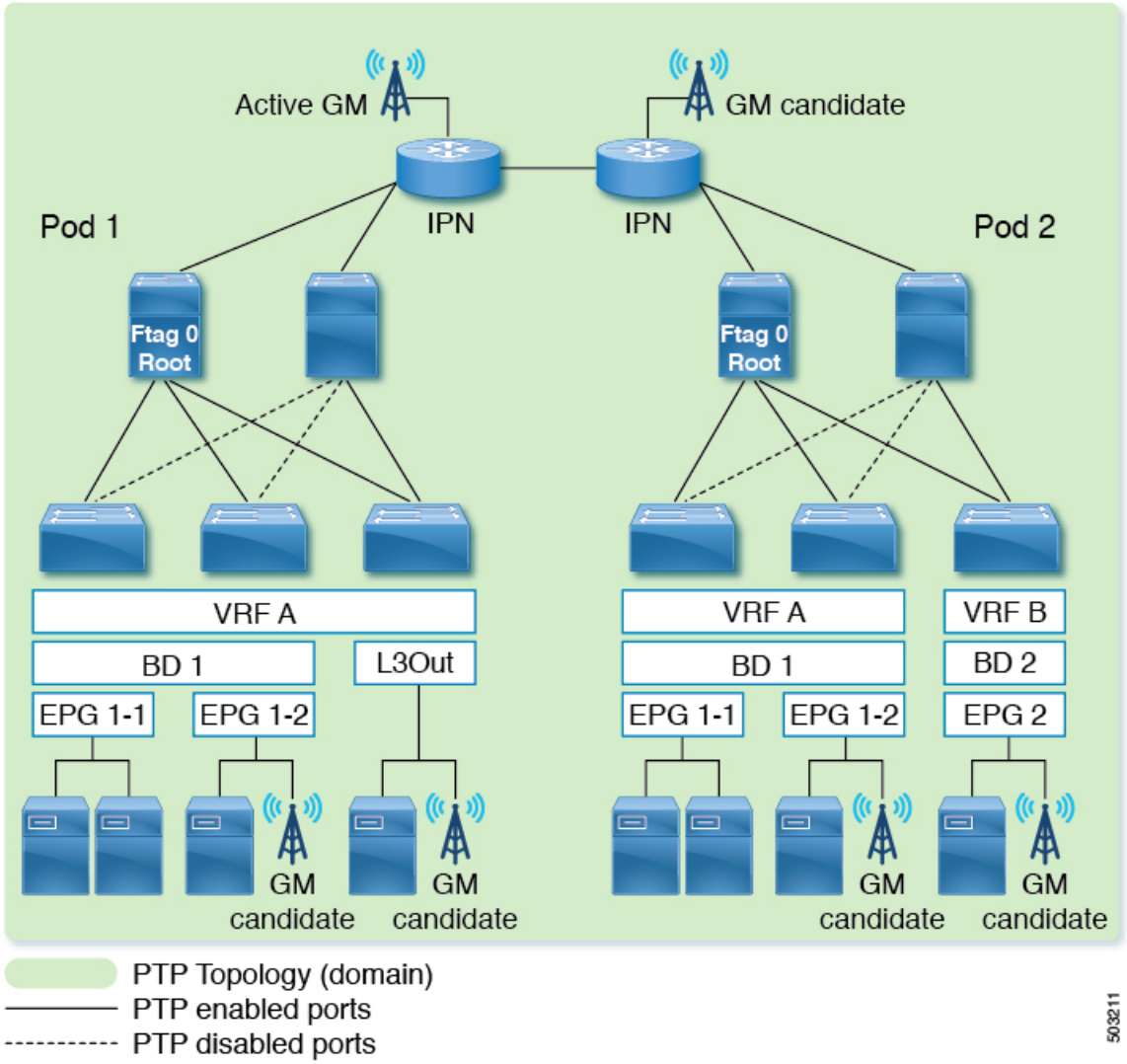
Single Pod

In a single pod deployment, grandmaster candidates can be deployed anywhere in the fabric (L3Out, EPG, or both). The Best Master Clock Algorithm (BMCA) picks one active grandmaster from among all of them.



Multipod With BMCA Across Pods

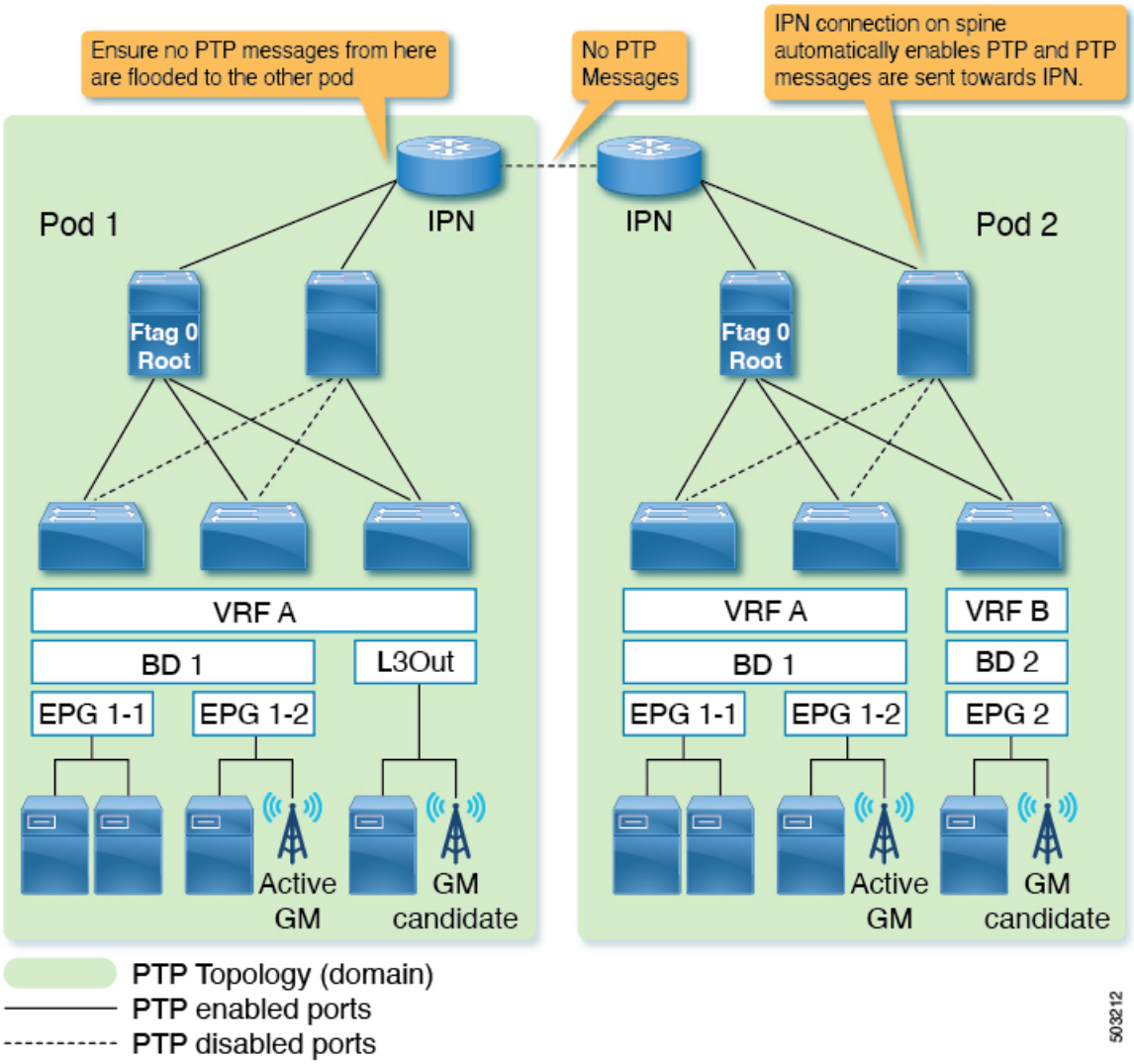
Grandmaster candidates can be deployed anywhere in the fabric (an inter-pod network, L3Out, EPG, or all of them). The BMCA picks one active grandmaster from among all of them across pods. We recommend that you place your grandmasters on inter-pod networks (IPNs) so that the PTP clients in any pod have a similar number of hops to the active grandmaster. In addition, the master/client tree topology will not change drastically when the active grandmaster becomes unavailable.



Multipod With BMCA in Each Pod

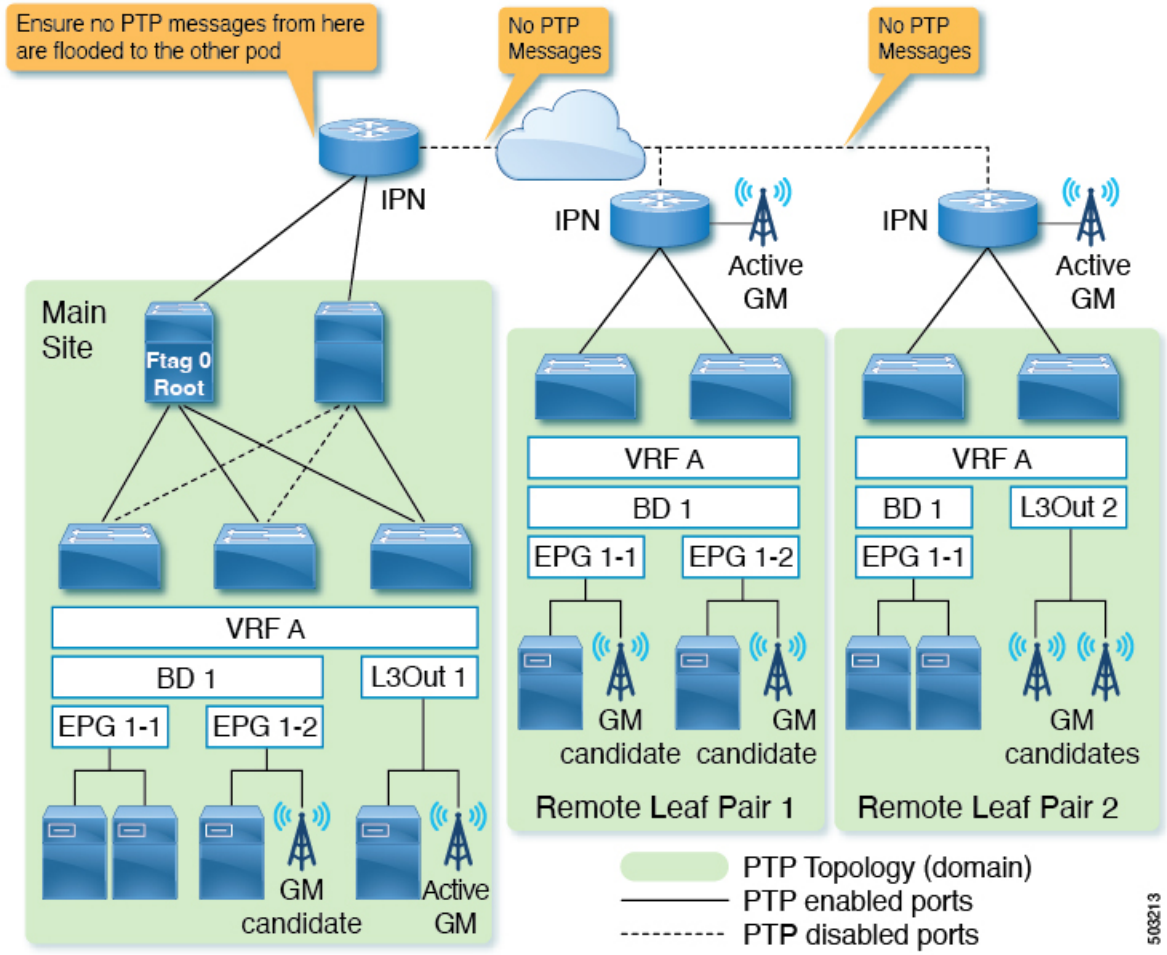
If you must have active grandmasters in each pod because PTP accuracy suffers too much degradation through an IPN domain, PTP messages must not traverse through an IPN across pods. You can accomplish this configuration in one of the following ways:

- Option 1: Ensure sub-interfaces are used between the IPN and spine switches, and disable PTP on the IPN.
- Option 2: If the PTP grandmaster is connected to the IPN in each pod, but the PTP topologies still must be separated, disable PTP on the IPN interfaces that are between the pods.



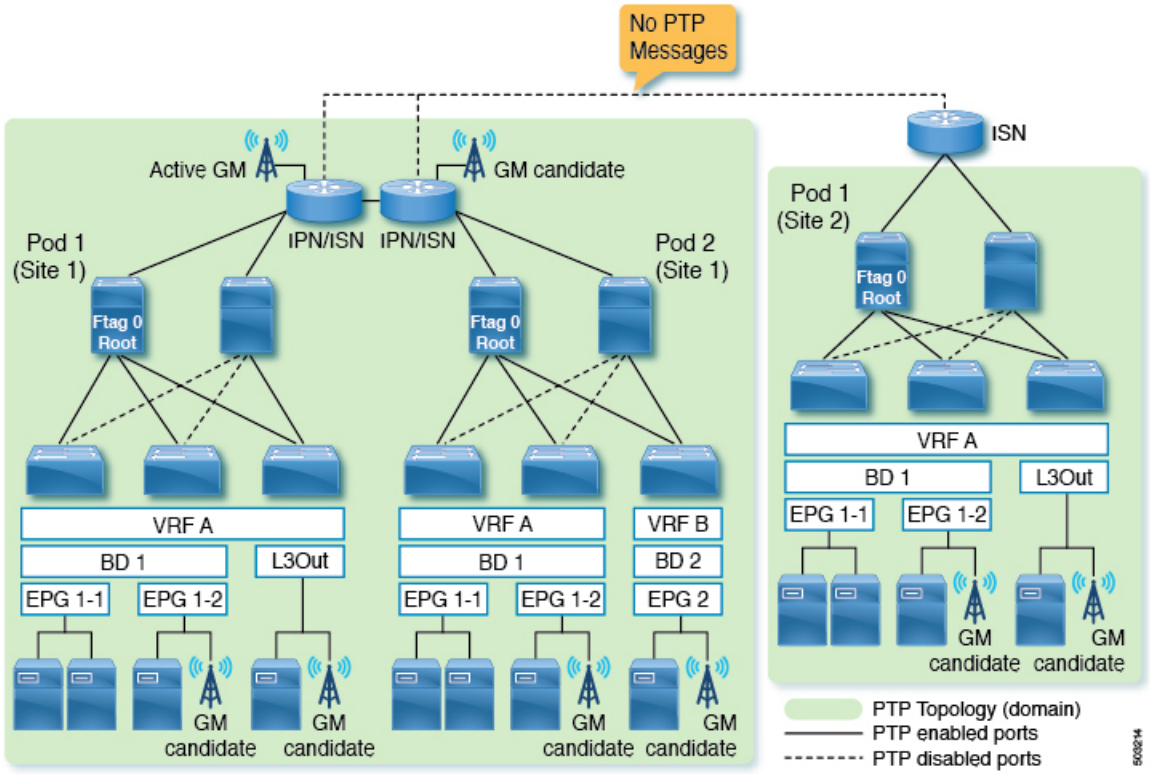
Remote Leaf Switch

Remote leaf switch sites are typically not close to the main data center or to each other, and it is difficult to propagate PTP messages across each location with accurate measurements of delay and correction. Hence, we recommend that you prevent PTP messages from traversing each site (location) so that the PTP topology is established within each site (location). Some remote locations may be close to each other. In such a case, you can enable PTP between those IPNs to form one PTP topology across those locations. You can use the same options mentioned in *Multipod With BMCA in Each Pod* to prevent PTP message propagation.



Cisco ACI Multi-Site

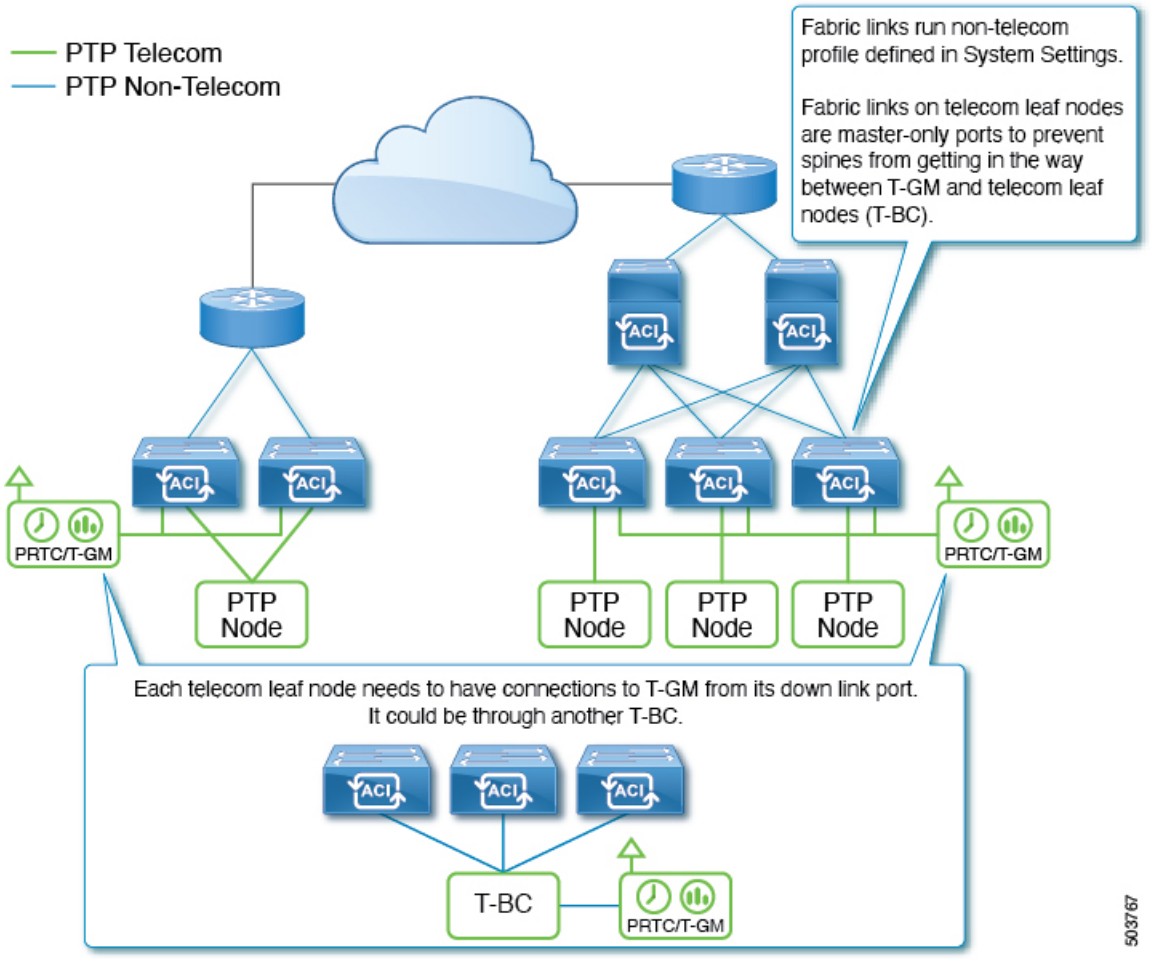
Each site is typically not close to each other, and it is difficult to propagate PTP messages across each site with accurate measurements of delay and correction. Hence, we recommend that you prevent PTP messages from traversing each site so that the PTP topology is established within each site. You can use the same options mentioned in *Multipod With BMCA in Each Pod* to prevent PTP message propagation. Also, Cisco ACI Multi-Site has no visibility or capability of configuring PTP.



Telecom Profile (G.8275.1)

The PTP Telecom profile (G.8275.1) in Cisco Application Centric Infrastructure (ACI) requires SyncE to achieve class B (G.8273.2) accuracy. Also, both the PTP Telecom profile (G.8275.1) and SyncE are supported only on Cisco N9K-C93180YC-FX3 leaf nodes. As a result, spine nodes cannot be used to distribute time, phase, and frequency synchronization for the Telecom profile (G.8275.1).

Because of this, fabric links on telecom leaf node (leaf nodes configured for G.8275.1) run in PTP multicast master only mode. This ensures that the telecom leaf nodes not to lock their clock through the spine nodes. This means that the grandmaster deployment for PTP Telecom profile (G.8275.1) in Cisco ACI requires each telecom leaf node to receive timing from the node's respective down link ports.



PTP Limitations

For general support and implementation information, see Supported Software for PTP, on page 38, Supported Hardware for PTP, on page 39, and PTP Connectivity, on page 40.

The following limitations apply to PTP:

- Cisco Application Centric Infrastructure (ACI) leaf and spine switches can work as PTP boundary clocks. The switches cannot work as PTP transparent clocks.
- Only the E2E delay mechanism (delay request/response mechanism) is supported. The P2P delay mechanism is not supported.
- PTP over IPv4/UDP for the default/Media/SMPTE PTP profiles and PTP over Ethernet for the Telecom (G.8275.1) PTP profile are supported. PTP over IPv6 is not supported.
- Only PTPv2 is supported.
  - Although PTPv1 packets are still redirected to the CPU when PTP is enabled on any of the front panel ports on the leaf switch, the packets will be discarded on the CPU.

- PTP Management TLVs are not recognized by Cisco ACI switches, but they are still forwarded as defined in IEEE1588-2008 to support the SMTPE PTP profile.
- PTP cannot be used as the system clock of the Cisco ACI switches.
- PTP is not supported on Cisco Application Policy Infrastructure Controller (APIC).
- NTP is required for all switches in the fabric.
- PTP offload is not supported. This functionality is the offloading of the PTP packet processing to each line card CPU on a modular spine switch for higher scalability.
- Due to a hardware limitation, interfaces with 1G/100M speed have lower accuracy than 10G interfaces when there are traffic loads. In the 5.2(3) and later releases, this limitation does not apply to the Cisco N9K-C93108TC-FX3P switch for the 1G speed.
- PTP is not fully supported on 100M interfaces due to higher PTP offset corrections.
- The PTP Telecom profile (G.8275.1) is not supported on ports with 1G/10G speed.
- `Sync` and `Delay_Request` messages can support up to a -4 interval (1/16 seconds). Interval values of -5 to -7 are not supported.
- For leaf switch front panel ports, PTP can be enabled per interface and VLAN, but PTP is automatically enabled on all appropriate fabric links (interfaces between leaf and spine switches, tier-1 and tier-2 leaf switches, and interfaces toward the IPN/ISN) after PTP is enabled globally. The appropriate fabric links are the interfaces that belong to ftag0 tree.
- PTP on Cisco ACI interfaces toward the IPN/ISN is enabled with the native VLAN 1 and sent out without a VLAN tag. The interfaces on the ISN/IPN node can send PTP packets toward Cisco ACI spine switches without a VLAN tag or with VLAN ID 4, which is enabled automatically for IPN/ISN connectivity regardless of PTP.
- PTP must be enabled globally for leaf switch front panel interfaces to use PTP. This means that you cannot enable PTP on leaf switch front panel ports without enabling PTP on the fabric links.
- PTP configuration using `tn-mgmt` and `tn-infra` is not supported.
- PTP can be enabled only on one VLAN per interface.
- PTP cannot be enabled on the interface and the VLAN for an L3Out SVI. PTP can be enabled on another VLAN on the same interface using an EPG.
- Only the leaf switch front panel interfaces can be configured as unicast master ports. The interfaces cannot be configured as unicast client ports. Unicast ports are not supported on a spine switch.
- Unicast negotiation is not supported.
- Unicast mode does not work with a PC or vPC when the PC or vPC is connected to a device such as NX-OS, which configures PTP on individual member ports.
- PTP and MACsec should not be configured on the same interface.
- When PTP is globally enabled, to measure the latency of traffic traversing through the fabric, Cisco ACI adds Cisco timestamp tagging (TTag) to traffic going from one ACI switch node to another ACI switch node. This results in an additional 8 bytes for such traffic. Typically, users do not need to take any actions regarding this implementation because the TTag is removed when the packets are sent out to the outside of the ACI fabric. However, when the setup consists of Cisco ACI Multi-Pod, user traffic traversing

across pods will keep the TTag in its inner header of the VXLAN. In such a case, increase the MTU size by 8 bytes on the ACI spine switch interfaces facing toward the Inter-Pod Network (IPN) along with all non-ACI devices in the IPN. IPN devices do not need to support nor be aware of the TTag, as the TTag is embedded inside of the VXLAN payload.

- When PTP is globally enabled, ERSPAN traffic traversing through spine nodes to reach to the ERSPAN destination will have Cisco timestamp tagging (TTag) with ethertype 0x8988. There is no impact to the original user traffic.
- In the presence of leaf switches that do not support PTP, you must connect an external grandmaster to all of the spine switches using IPN or using leaf switches that support PTP. If a grandmaster is connected to one or a subset of spine switches, PTP messages from the spine may be blocked by the unsupported leaf switch before they reach other switches depending on the ftag0 tree status. PTP within leaf and spine switches are enabled based on ftag0 tree, which is automatically built based on Cisco ACI infra ISIS for loop free multicast connectivity between all leaf and spine switches in each pod.
- When the PTP Telecom profile is deployed, the Telecom grandmaster clock (T-GM) and Telecom boundary clock (T-BC) timestamps should be within 2 seconds for the T-BC to lock with the T-GM.
- You cannot enable PTP on a VLAN that is deployed on a leaf node interface using VMM domain integration.

Configuring PTP

PTP Configuration Basic Flow

The following steps provide an overview of the PTP configuration process:

Procedure

Step 1 Enable PTP globally and set PTP parameters for all fabric interfaces.

Step 2 For the PTP Telecom profile (G.8275.1) only, create a PTP node policy and apply it to a switch profile through a switch policy group.

Step 3 Create PTP user profile for leaf front panel interfaces under **Fabric** > **Access Policies** > **Policies** > **Global**.

Step 4 Enable PTP under **EPG** > **Static Ports** with the PTP user profile.

Step 5 Enable PTP under **L3Out** > **Logical Interface Profile** > **Routed or Sub-Interface** with the PTP user profile.

Configuring the PTP Policy Globally and For the Fabric Interfaces Using the GUI

This procedure enables the precision time protocol (PTP) globally and for the fabric interfaces using the Cisco Application Policy Infrastructure Controller (APIC) GUI. When PTP is enabled globally, ongoing TEP to TEP latency measurements get enabled automatically.

Procedure

Step 1 On the menu bar, choose **System** > **System Settings**.

Step 2 In the Navigation pane, choose **PTP and Latency Measurement**.

Step 3 In the Work pane, set the interface properties as appropriate for your desired configuration. At the least, you must set **Precision Time Protocol** to **Enabled**.

See the online help page for information about the fields. If any interval value that you specify is outside of the chosen PTP profile standard range, the configuration is rejected.

The PTP profile, intervals, and timeout fields apply to fabric links. The other fields apply to all of the leaf and spine switches.

Step 4 Click **Submit**.

Configuring a PTP Node Policy and Applying the Policy to a Switch Profile Using a Switch Policy Group Using the GUI

A PTP node policy is required for the leaf nodes to run PTP Telecom profile (G.8275.1) because it uses the Alternate BMCA with additional parameters. Also, the allowed range of the domain number, priority 1, and priority 2 are different from other PTP profiles. You can apply the PTP node policy to a leaf switch using a leaf switch profile and a policy group.

Note For media profile deployment, you do not need to create a node policy.

Procedure

Step 1 On the menu bar, choose **Fabric** > **Access Policies**.

Step 2 In the Navigation pane, choose **Switches** > **Leaf Switches** > **Profiles**.

Step 3 Right-click **Profiles** and choose **Create Leaf Profile**.

Step 4 In the **Create Leaf Profile** dialog, in the **Name** field, enter a name for the profile.

Step 5 In the **Leaf Selectors** section, click **+**.

Step 6 Enter a name, choose the switches, and choose to create a policy group.

Step 7 In the **Create Access Switch Policy Group** dialog, enter a name for the policy group.

Step 8 In the **PTP Node Policy** drop-down list, choose **Create PTP Node Profile**.

Step 9 In the **Create PTP Node Profile** dialog, set the values as desired for your configuration.

- **Node Domain**: The value must be between 24 and 43, inclusive. The Telecom leaf nodes that need to be in the same PTP topology should use the same domain number.
- **Priority 1**: The value must be 128.
- **Priority 2**: The value must be between 0 and 255, inclusive.

See the online help page for information about the fields.

Step 10 Click **Submit**.

The **Create PTP Node Profile** dialog closes.

Step 11 In the **Create Access Switch Policy Group** dialog, set any other policies as desired for your configuration.

Step 12 Click **Submit**.

The **Create Access Switch Policy Group** dialog closes.

Step 13 In the **Leaf Selectors** section, click **Update**.

Step 14 Click **Next**.

Step 15 In the **STEP 2 > Associations** screen, associate the interface profiles as desired.

Step 16 Click **Finish**.

Creating the PTP User Profile for Leaf Switch Front Panel Ports Using the GUI

This procedure creates the PTP user profile for leaf switch front panel ports using the Cisco Application Policy Infrastructure Controller (APIC) GUI. A PTP user profile is applied to the leaf switch front panel interfaces using an EPG or L3Out.

Before you begin

You must enable PTP globally to use PTP on leaf switch front panel ports that face external devices.

Procedure

Step 1 On the menu bar, choose **Fabric** > **Access Policies**.

Step 2 In the Navigation pane, choose **Policies** > **Global** > **PTP User Profile**.

Step 3 Right-click **PTP User Profile** and choose **Create PTP User Profile**.

Step 4 In the **Create PTP User Profile** dialog, set the values as desired for your configuration.

See the online help page for information about the fields. If any interval value that you specify is outside of the chosen PTP profile standard range, the configuration is rejected.

Step 5 Click **Submit**.

Enabling PTP on EPG Static Ports Using the GUI

This procedure enables PTP on EPG static ports using the Cisco Application Policy Infrastructure Controller (APIC) GUI. You can enable PTP with multicast dynamic, multicast master, or unicast master mode.

Before you begin

You must first create a PTP user profile for the leaf switch front panel ports and enable PTP globally.

Procedure

Step 1 On the menu bar, choose **Tenants** > **All Tenants**.

Step 2 In the Work pane, double-click the tenant's name.

Step 3 In the Navigation pane, choose **Tenant** ***tenant_name*** > **Application Profiles** > ***app_profile_name*** > **Application EPGs** > ***app_epg_name*** > **Static Ports** > ***static_port_name***.

Step 4 In the Work pane, for the **PTP State** toggle, choose **Enable**. You might need to scroll down to see **PTP State**.

PTP-related fields appear.

Step 5 Configure the PTP fields as required for your configuration.

- **PTP Mode**: Choose **multicast dynamic**, **multicast master**, or **unicast master**, as appropriate.
- **PTP Source Address**: PTP packets from this interface and VLAN are sent with the specified IP address as the source. The leaf switch TEP address is used by default or when you enter "0.0.0.0" as the value. This value is optional for multicast mode. Use the bridge domain SVI or EPG SVI for unicast mode. The source IP address must be reachable by the connected PTP node for unicast mode.
- **PTP User Profile**: Choose the PTP user profile that you created for the leaf switch front panel ports to specify the message intervals.

See the online help page for additional information about the fields.

A node-level configuration takes precedence over the fabric-level configuration on a node where the PTP Telecom profile (G.8275.1) is deployed.

Step 6 Click **Submit**.

Enabling PTP on L3Out Interfaces Using the GUI

This procedure enables PTP on L3Out interfaces using the Cisco Application Policy Infrastructure Controller (APIC) GUI. You can enable PTP with multicast dynamic, multicast master, or unicast master mode.

Before you begin

You must first create a PTP user profile for the leaf switch front panel ports and enable PTP globally.

Procedure

Step 1 On the menu bar, choose **Tenants** > **All Tenants**.

Step 2 In the Work pane, double-click the tenant's name.

Step 3 In the Navigation pane, choose **Tenant** ***tenant_name*** > **Networking** > **L3Outs** > ***l3out_name*** > **Logical Node Profiles** > ***node_profile_name*** > **Logical Interface Profiles** > ***interface_profile_name***.

Step 4 In the Work pane, choose **Policy** > **Routed Sub-Interfaces** or **Policy** > **Routed Interfaces**, as appropriate.

Step 5 If you want to enable PTP on an existing L3Out, perform the following sub-steps:

a) Double-click the desired interface to view its properties.

b) Scroll down if necessary to find the PTP properties, set the **PTP State** to **Enable**, and enter the same values that you used for the EPG static ports.

See the online help page for information about the fields.

c) Click **Submit**.

Step 6 If you want to enable PTP on a new L3Out, perform the following sub-steps:

a) Click **+** at the upper right of the table.

b) In **Step 1 > Identity**, enter the appropriate values.

c) In **Step 2 > Configure PTP**, set the **PTP State** to **Enable**, and enter the same values that you used for the EPG static ports.

See the online help page for information about the fields.

d) Click **Finish**.

Configuring the PTP Policy Globally and For the Fabric Interfaces Using the REST API

This procedure enables PTP globally and for the fabric interfaces using the REST API. When PTP is enabled globally, ongoing TEP to TEP latency measurements get enabled automatically.

To configure the PTP policy globally and for the fabric interfaces, send a REST API POST similar to the following example:

POST: `/api/mo/uni/fabric/ptpmode.xml`

```
<latencyPtpMode
    state="enabled"                          # PTP admin state
    systemResolution="11"                    # Latency Resolution (can be skipped for PTP)
    prio1="255"                              # Global Priority1
    prio2="255"                              # Global Priority2
    globalDomain="0"                         # Global Domain
    fabProfileTemplate="aes67"               # PTP Profile
    fabAnnounceIntvl="1"                     # Announce Interval (2^x sec)
    fabSyncIntvl="-3"                        # Sync Interval (2^x sec)
    fabDelayIntvl="-2"                       # Delay Request Interval (2^x sec)
    fabAnnounceTimeout="3"                   # Announce Timeout
/>
```

Configuring a PTP Node Policy and Applying the Policy to a Switch Profile Using a Switch Policy Group Using the REST API

A PTP node policy is required for the leaf nodes to run PTP Telecom profile (G.8275.1) because it uses the Alternate BMCA with additional parameters. Also, the allowed range of the domain number, priority 1, and priority 2 are different from other PTP profiles. You can apply the PTP node policy to a leaf switch using a leaf switch profile and a policy group.

POST: `/api/mo/uni.xml`

```
<infraInfra>
    <!-- Switch Profile -->
    <infraNodeP name="L101_SWP" dn="uni/infra/nprof-L101_SWP">
        <infraRsAccPortP tDn="uni/infra/accportprof-L101_IFP"/>
        <infraLeafS name="L101" type="range">
            <infraNodeBlk name="L101" to_="101" from_="101"/>
            <!-- Associate Switch Policy Group for node-101 -->
```

```
            <infraRsAccNodePGrp tDn="uni/infra/funcprof/accnodepgrp-Telecom_PG_1"/>
        </infraLeafS>
    </infraNodeP>

    <infraFuncP>
        <!-- Switch Policy Group with PTP Node and SyncE Policy -->
        <infraAccNodePGrp name="Telecom_PG_1"
          dn="uni/infra/funcprof/accnodepgrp-Telecom_PG_1">
            <infraRsSynceInstPol tnSynceInstPolName="SyncE_QL1"/>
            <infraRsPtpInstPol tnPtpInstPolName="Telecom_domain24"/>
        </infraAccNodePGrp>
    </infraFuncP>

    <!-- PTP Node policy -->
    <ptpInstPol
      dn="uni/infra/ptpInstP-Telecom_domain24"
      name="Telecom_domain24"
      operatingMode="hybrid"
      nodeProfile="telecom_full_path"
      nodePrio1="128"
      nodePrio2="128"
      nodeDomain="24"/>

    <!-- SyncE Node policy -->
    <synceInstPol
      dn="uni/infra/synceInstP-SyncE_QL1"
      name="SyncE_QL1"
      qloption="op1"
      adminSt="disabled"/>
</infraInfra>
```

Creating the PTP User Profile for Leaf Switch Front Panel Ports Using the REST API

A PTP user profile is applied to the leaf switch front panel interfaces using an EPG or L3Out. You also must enable PTP globally to use PTP on leaf switch front panel ports that face external devices.

To create the PTP user profile, send a REST API POST similar to the following example:

POST: `/api/mo/uni/infra/ptpprofile-Ptelecomprofile.xml`

```
<ptpProfile
    name="Ptelecomprofile"                    # PTP user profile name
    profileTemplate="telecom_full_path"       # PTP profile
    announceIntvl="-3"                        # Announce interval (2^x sec)
    syncIntvl="-4"                            # Sync interval (2^x sec)
    delayIntvl="-4"                           # Delay request interval (2^x sec)
    announceTimeout="3"                       # Announce timeout
    annotation=""                             # Annotation key

                                              (Only for Telecom ports)
    ptpoeDstMacType="forwardable"             # Destination MAC for PTP messages
    ptpoeDstMacRxNoMatch="replyWithCfgMac"    # Packet handling
    localPriority="128"                       # Port local priority

                                              (Only for non-Telecom ports on a telecom leaf)
    nodeProfileOverride="no"                  # Node profile override
/>
```

Enabling PTP on EPG Static Ports Using the REST API

Before you can enable PTP on EPG static ports, you must first create a PTP user profile for the leaf switch front panel ports and enable PTP globally.

To enable PTP on EPG static ports, send a REST API POST similar to the following example:

POST: `/api/mo/uni/tn-TK/ap-AP1/epg-EPG1-1.xml`

Multicast Mode

```
<fvRsPathAtt
  tDn="topology/pod-1/paths-101/pathep-[eth1/1]"
  encap="vlan-2011">
    <ptpEpgCfg
      ptpMode="multicast">                                  # PTP mode
        <ptpRsProfile                                       # PTP user profile
          tDn="uni/infra/ptpprofile-PTP_AES"/>
    </ptpEpgCfg>
</fvRsPathAtt>
```

The possible values for the `ptpMode` parameter are as follows:

- `multicast`: Multicast dynamic.
- `multicast-master`: Multicast master.

Unicast Mode

```
<fvRsPathAtt
  tDn="topology/pod-1/paths-101/pathep-[eth1/1]"
  encap="vlan-2011">
    <ptpEpgCfg
      srcIp="192.168.1.254"                                  # PTP source IP address
      ptpMode="unicast-master">                              # PTP mode
        <ptpRsProfile                                        # PTP user profile
          tDn="uni/infra/ptpprofile-PTP_AES"/>               # PTP unicast destination
        <ptpUcastIp dstIp="192.168.1.11"/>                     IP address
    </ptpEpgCfg>
</fvRsPathAtt>
```

If `ptpEpgCfg` exists, that means that PTP is enabled. If PTP must be disabled on that interface, delete `ptpEpgCfg`.

Enabling PTP on L3Out Interfaces Using the REST API

This procedure enables PTP on L3Out interfaces using the REST API. Before you can enable PTP on the L3Out interfaces, you must first create a PTP user profile for the leaf switch front panel ports and enable PTP globally.

To enable PTP on L3Out interfaces, send a REST API POST similar to the following example:

POST: `/api/node/mo/uni/tn-TK/out-BGP/lnodep-BGP_nodeProfile/lifp-BGP_IfProfile.xml`

Multicast Mode

```
<l3extRsPathL3OutAtt
  tDn="topology/pod-1/paths-103/pathep-[eth1/11]"
  addr="11.0.0.1/30" ifInstT="l3-port">
    <ptpRtdEpgCfg
      ptpMode="multicast">                                # PTP mode
        <ptpRsProfile                                     # PTP user profile
          tDn="uni/infra/ptpprofile-PTP_AES"/>
    </ptpRtdEpgCfg>
</l3extRsPathL3OutAtt>
```

The possible values for the `ptpMode` parameter are as follows:

- `multicast`: Multicast dynamic.
- `multicast-master`: Multicast master.

Unicast Mode

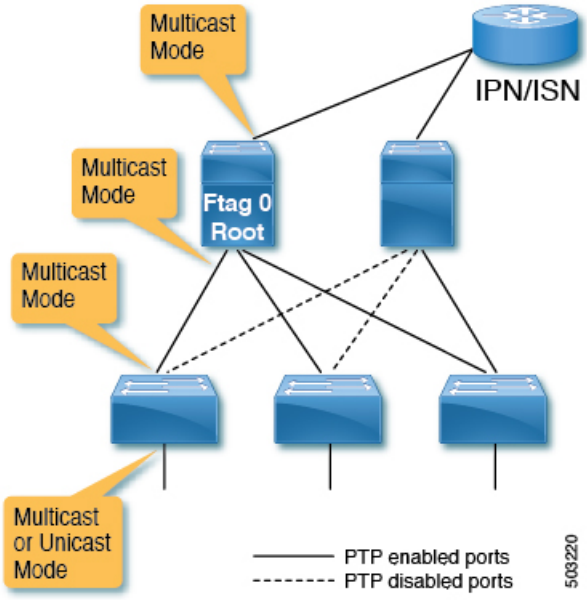
```
<l3extRsPathL3OutAtt
  tDn="topology/pod-1/paths-103/pathep-[eth1/11]"
  addr="11.0.0.1/30" ifInstT="l3-port">
    <ptpRtdEpgCfg
      srcIp="11.0.0.1"                                  # PTP source IP address
      ptpMode="unicast-master">                         # PTP mode
        <ptpRsProfile                                   # PTP user profile
          tDn="uni/infra/ptpprofile-PTP_AES"/>          # PTP unicast destination
        <ptpUcastIp dstIp="11.0.0.4"/>                    IP address
    </ptpRtdEpgCfg>
</l3extRsPathL3OutAtt>
```

If `ptpRtdEpgCfg` exists, that means that PTP is enabled. If PTP needs to be disabled on that interface, delete `ptpRtdEpgCfg`.

PTP Unicast, Multicast, and Mixed Mode on Cisco ACI

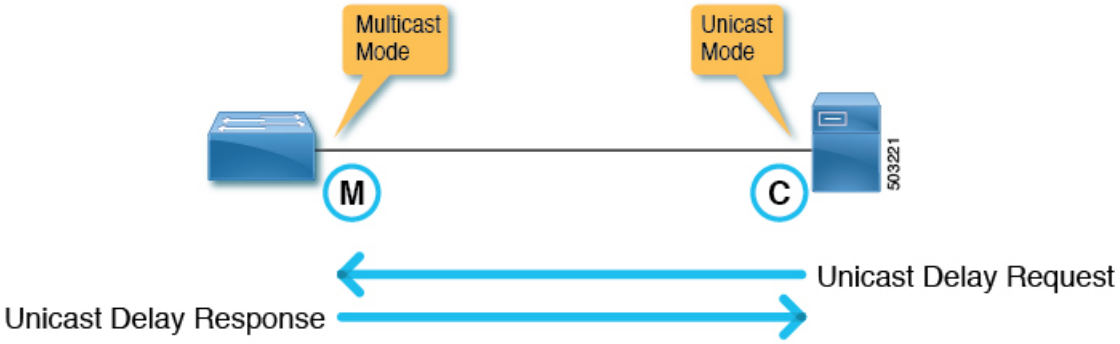
By default, all PTP interfaces run in multicast mode. Only the leaf switch front panel interfaces can be configured in unicast mode. Only unicast master ports are supported; unicast client ports are not supported.

Figure 1: Multicast or Unicast Mode



Mixed mode (a PTP multicast port replying with a unicast delay response) will be automatically activated on a PTP master port in multicast mode when the port receives a unicast delay request. Mixed mode is essentially a multicast master and unicast client.

Figure 2: Mixed Mode



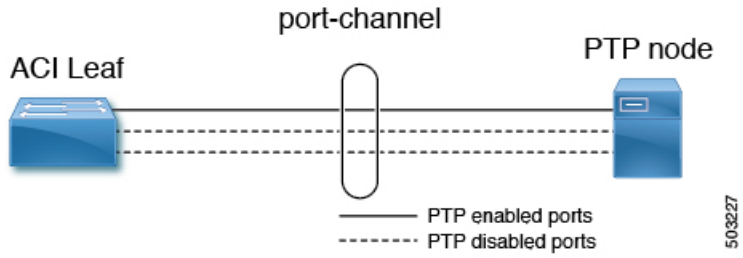
One leaf switch can have multiple PTP unicast master ports. The supported number of client switch IP addresses on each unicast master port is 2. More IP addresses can be configured, but not qualified. The PTP unicast master ports and PTP multicast ports can be configured on the same switch.

PTP Unicast Mode Limitations on Cisco ACI

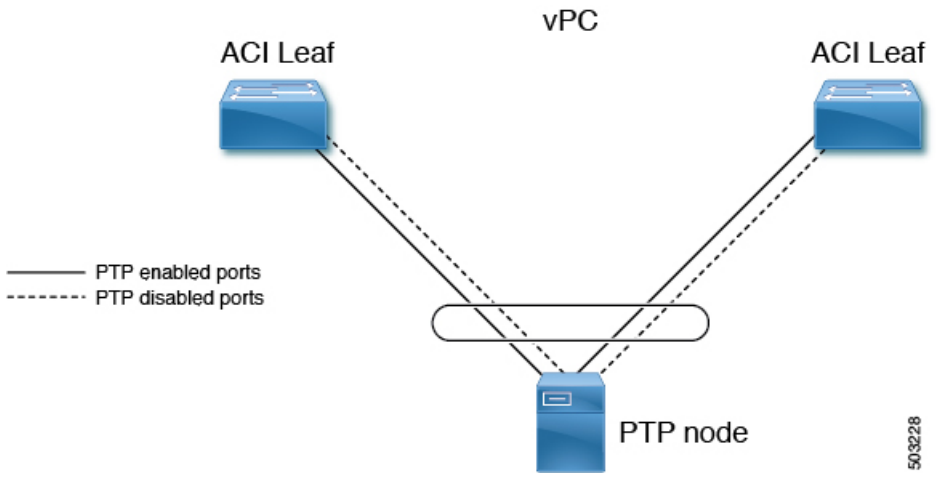
PTP unicast negotiation is not supported. Because Cisco Application Centric Infrastructure (ACI) does not have unicast negotiation to request the messages that Cisco ACI wants or to grant those requests from other nodes, Cisco ACI PTP unicast master ports will send `Announce`, `Sync`, and `Follow_Up` messages with intervals configured using the Cisco Application Policy Infrastructure Controller (APIC) without receiving any requests from its client nodes. Unicast `Delay_Response` messages are sent out as a response to `Delay_Request` messages from the unicast client nodes. Because a unicast master port sends PTP messages such as `Sync` without listening to unicast requests, the Best Master Clock Algorithm (BMCA) is not calculated on the Cisco ACI PTP unicast ports.

PTP PC and vPC Implementation on Cisco ACI

For port channels (PCs) and virtual port channels (vPCs), PTP is enabled per PC or vPC instead of per member port. Cisco Application Centric Infrastructure (ACI) does not allow PTP to be enabled on each member port of the parent PC or vPC individually.



When PTP is enabled on a Cisco ACI PC or vPC, the leaf switch automatically picks a member port from the PC on which PTP is enabled. When the PTP-enabled member port fails, the leaf switch picks another member port that is still up. The PTP port status is inherited from the previous PTP-enabled member port.



When PTP is enabled on a Cisco ACI vPC port, even though vPC is a logical bundle of two port channels on two leaf switches, the behavior is the same as PTP being enabled on a normal port channel. There is no specific implementation for the vPC, such as the synchronization of PTP information between vPC peer leaf switches.

Note Unicast mode does not work with a PC or vPC when the PC or vPC is connected to a device such as NX-OS, which configures PTP on individual member ports.

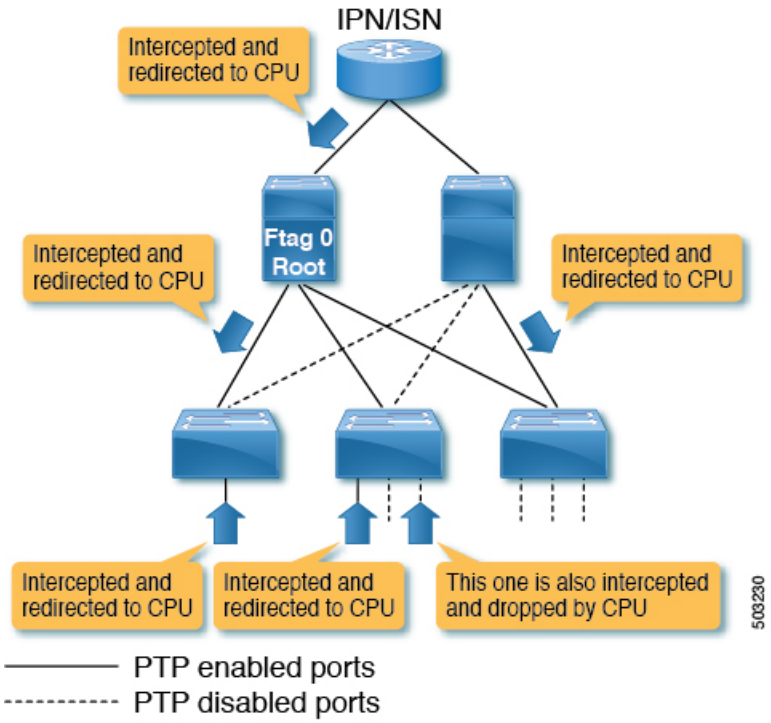
PTP Packet Filtering and Tunneling

PTP Packet Filtering

When PTP handles packets on the fabric ports and PTP is enabled globally, all spine and leaf switches have internal filters to redirect all incoming PTP packets from any fabric ports to the CPU.

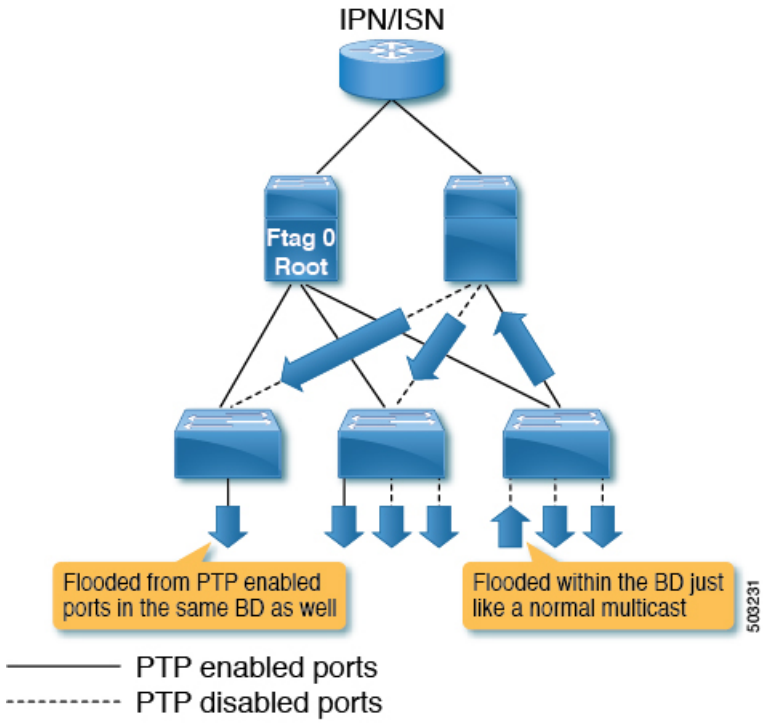
When PTP handles packets on the front panel ports and PTP is enabled on at least one leaf switch front panel port on a given leaf switch, the leaf switch has internal filters to redirect all incoming PTP packets from any front panel ports. Even if a PTP packet is received from a front panel port on which PTP is not enabled, the packet is still intercepted and redirected to the CPU, then discarded.

Figure 3: Packet Filtering On the Front Panel on a Leaf switch With PTP-Enabled Front Panel Ports



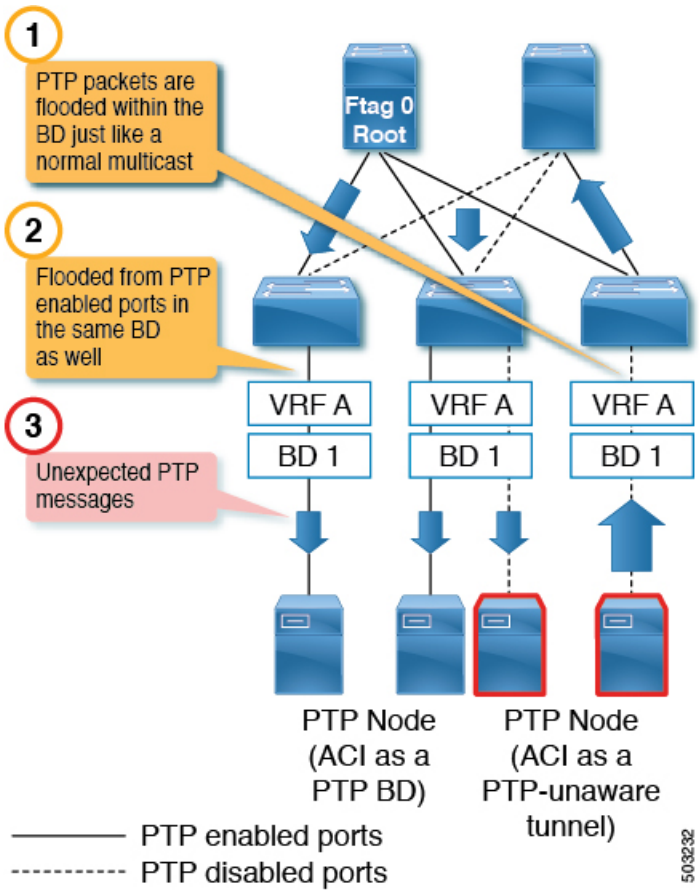
When PTP handles packets on the front panel ports and PTP is not enabled on any leaf switch front panel ports on a given leaf switch, the leaf switch does not have internal filters to redirect PTP packets from any front panel ports. If a PTP packet is received on a front panel port on such a leaf switch, the packet is handled as a normal multicast packet and is forwarded or flooded to other switches using VxLAN. The other switches will also handle this as a normal multicast packet because the PTP packets that are supposed to be intercepted by the Cisco Application Centric Infrastructure (ACI) switches are not encapsulated in VxLAN even between leaf and spine switches. This may cause unexpected PTP behavior on other leaf switches with PTP enabled on the front panel ports. For more information, see Cisco ACI As a PTP Boundary Clock or PTP-Unaware Tunnel, on page 60.

Figure 4: Packet Filtering On the Front Panel on a Leaf switch Without PTP-Enabled Front Panel Ports

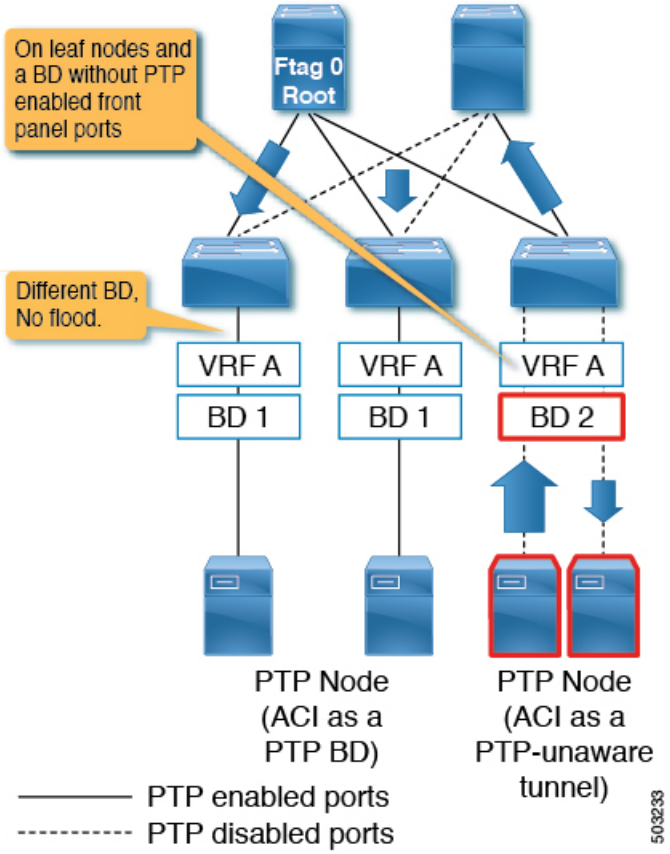


Cisco ACI As a PTP Boundary Clock or PTP-Unaware Tunnel

PTP packets from a leaf switch with no PTP front panel ports are flooded in the bridge domain. The packets are flooded even toward PTP nodes in the same bridge domain that expect Cisco Application Centric Infrastructure (ACI) to regenerate PTP messages as a PTP boundary clock, as shown in the following illustration:

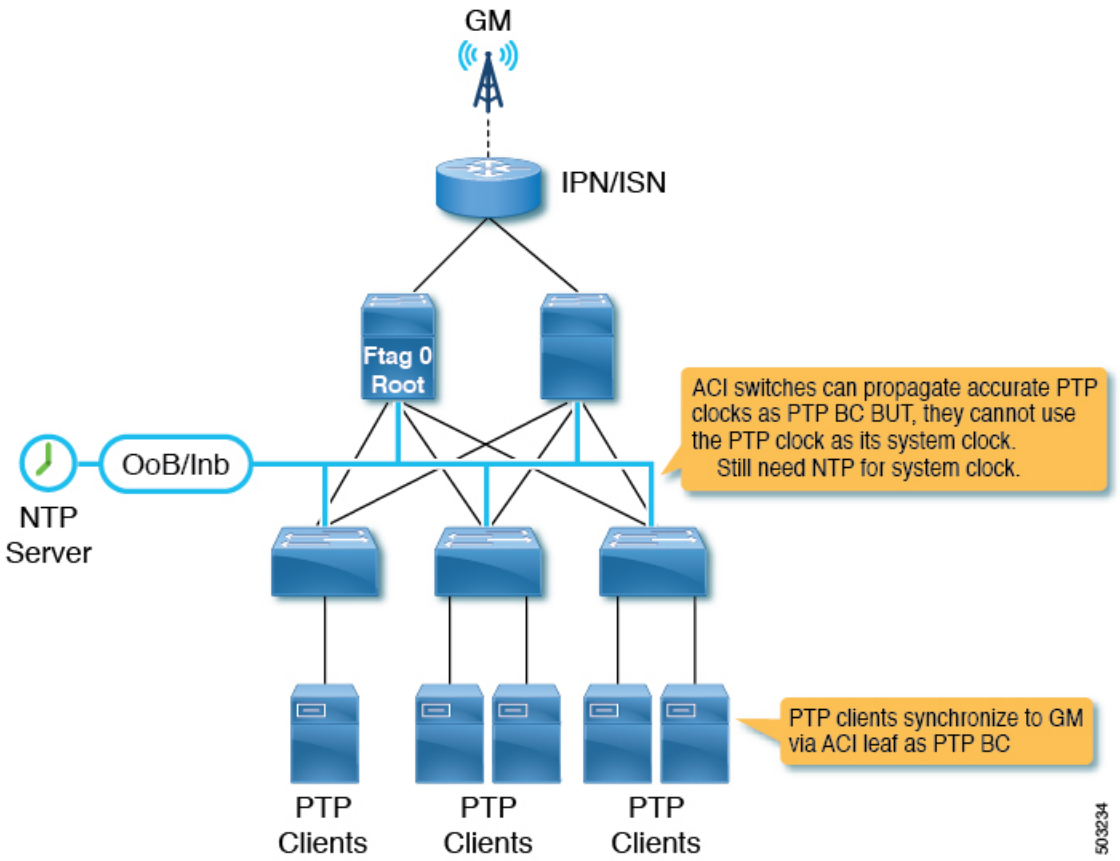


This will result in confusing PTP nodes and its time calculation due to unexpected PTP packets. On the other hand, PTP packets from a leaf switch with PTP front panel ports are always intercepted and never tunneled even if the packets are received on a port on which PTP is not enabled. Therefore, do not mix PTP nodes that need Cisco ACI to be a PTP boundary clock and that need Cisco ACI to be a PTP-unaware tunnel in the same bridge domain and on the same leaf switch. The configuration shown in the following illustration (different bridge domain, different leaf switch) is supported:



PTP and NTP

Cisco Application Centric Infrastructure (ACI) switches run as PTP boundary clocks to provide an accurate clock from the grandmaster to the PTP clients. However, the Cisco ACI switches and Cisco Application Policy Infrastructure Controllers (APICs) cannot use those PTP clocks as their own system clock. The Cisco ACI switches and Cisco APICs still need an NTP server to update their own system clock.



Note For PTP to work accurately and constantly on Cisco ACI, NTP must be configured for all of the switches to keep their system clock as accurate as the PTP grandmaster in the 100 ms order. In other words, the system clock must have less than a 100 ms difference compared to the PTP grandmaster.

PTP Verification

Summary of PTP Verification CLI Commands

You can log into one of the leaf switches and the following commands to verify the PTP configuration:

Command	Purpose
show ptp port interface *slot/port*	Displays the PTP parameters of a specific interface.
show ptp brief	Displays the PTP status.
show ptp clock	Displays the properties of the local clock, including clock identity.
show ptp parent	Displays the properties of the PTP parent.

Command	Purpose
show ptp clock foreign-masters record	Displays the state of foreign masters known to the PTP process. For each foreign master, the output displays the clock identity, basic clock properties, and whether the clock is being used as a grandmaster.
show ptp counters [all |interface Ethernet *slot/port***]**	Displays the PTP packet counters for all interfaces or for a specified interface.
show ptp corrections	Displays the last few PTP corrections.

Showing the PTP Port Information

The following example shows the port interface information:

```
f2-leaf1# vsh -c 'show ptp port int e1/1'
PTP Port Dataset: Eth1/1
Port identity: clock identity: 00:3a:9c:ff:fe:6f:a4:df
Port identity: port number: 0
PTP version: 2
Port state: Master
VLAN info: 20                              <--- PTP messages are sent on this PI-VLAN
Delay request interval(log mean): -2
Announce receipt time out: 3
Peer mean path delay: 0
Announce interval(log mean): 1
Sync interval(log mean): -3
Delay Mechanism: End to End
Cost: 255
Domain: 0
```

The following example shows the information for the specified VLAN:

```
f2-leaf1# show vlan id 20 extended

 VLAN Name                             Encap            Ports
 ---- -------------------------------- ---------------- ------------------------
 20   TK:AP1:EPG1-1                    vlan-2011        Eth1/1, Eth1/2, Eth1/3
```

Showing the PTP Port Status

The following example shows the brief version of the port status:

```
f2-leaf1# show ptp brief

PTP port status
-----------------------------------
Port                  State
--------------------- ------------
Eth1/1                Master
Eth1/51               Passive
Eth1/52               Slave
```

Showing the PTP Switch Information

The following example shows the brief version of the switch status:

```
f2-leaf1# show ptp clock
PTP Device Type : boundary-clock
PTP Device Encapsulation : layer-3
```

```
PTP Source IP Address : 20.0.32.64            <--- Switch TEP. Like a router-id.
                                                   This is not PTP Source Address you
                                                   configure per port.
Clock Identity : 00:3a:9c:ff:fe:6f:a4:df      <--- PTP clock ID. If this node is
                                                   the grandmaster, this ID is the
                                                   grandmaster's ID.
Clock Domain: 0
Slave Clock Operation : Two-step
Master Clock Operation : Two-step
Slave-Only Clock Mode : Disabled
Number of PTP ports: 3
Configured Priority1 : 255
Priority1 : 255
Priority2 : 255
Clock Quality:
        Class : 248
        Accuracy : 254
        Offset (log variance) : 65535
Offset From Master : -8                       <--- -8 ns. the clock difference from the
                                                   closest parent (master)
Mean Path Delay : 344                         <--- 344 ns. Mean path delay measured by
                                                   E2E mechanism.
Steps removed : 2                             <--- 2 steps. 2 PTP BC nodes between the
                                                   grandmaster.
Correction range : 100000
MPD range : 1000000000
Local clock time : Thu Jul 30 01:26:14 2020
Hardware frequency correction : NA
```

Showing the Grandmaster and Parent (Master) Information

The following example shows the PTP grandmaster and parent (master) information:

```
f2-leaf1# show ptp parent

PTP PARENT PROPERTIES

Parent Clock:
Parent Clock Identity: 2c:4f:52:ff:fe:e1:7c:1a          <--- closest parent (master)
Parent Port Number: 30
Observed Parent Offset (log variance): N/A
Observed Parent Clock Phase Change Rate: N/A

Parent IP: 20.0.32.65                                   <--- closest parent's PTP
                                                        source IP address
Grandmaster Clock:
Grandmaster Clock Identity: 00:78:88:ff:fe:f9:2b:13     <--- GM
Grandmaster Clock Quality:                              <--- GM's quality
        Class: 248
        Accuracy: 254
        Offset (log variance): 65535
        Priority1: 128
        Priority2: 255
```

The following example shows the PTP foreign master clock records:

```
f2-leaf1# show ptp clock foreign-masters record

P1=Priority1, P2=Priority2, C=Class, A=Accuracy,
OSLV=Offset-Scaled-Log-Variance, SR=Steps-Removed
GM=Is grandmaster

---------   ----------------------  ---  ----  ----  ---  -----  --------
Interface         Clock-ID          P1   P2    C     A    OSLV   SR
```

```
---------   ----------------------  ---  ----  ----  ---  -----  --------

Eth1/51   c4:f7:d5:ff:fe:2b:eb:8b    128  255   248   254  65535  1
Eth1/52   2c:4f:52:ff:fe:e1:7c:1a    128  255   248   254  65535  1
```

The output shows the master clocks that send grandmaster information to the switch and the switch's connected interface. The clock ID here is the closest master's ID. The ID is not the grandmaster's ID. Because this switch is receiving the grandmaster's data from two different ports, one of the ports became passive.

Showing the Counters

The following example shows the counters of a master port:

```
f2-leaf1# show ptp counters int e1/1

PTP Packet Counters of Interface Eth1/1:
------------------------------------------------------------------
Packet Type                  TX                      RX
---------------    --------------------    --------------------
Announce                           4                       0
Sync                              59                       0
FollowUp                          59                       0
Delay Request                      0                      30
Delay Response                    30                       0
PDelay Request                     0                       0
PDelay Response                    0                       0
PDelay Followup                    0                       0
Management                         0                       0
```

A master port should send the following messages:

- Announce
- Sync
- FollowUp
- Delay Response

A master port should receive the following message:

- Delay Request

The following example shows the counters of a client port:

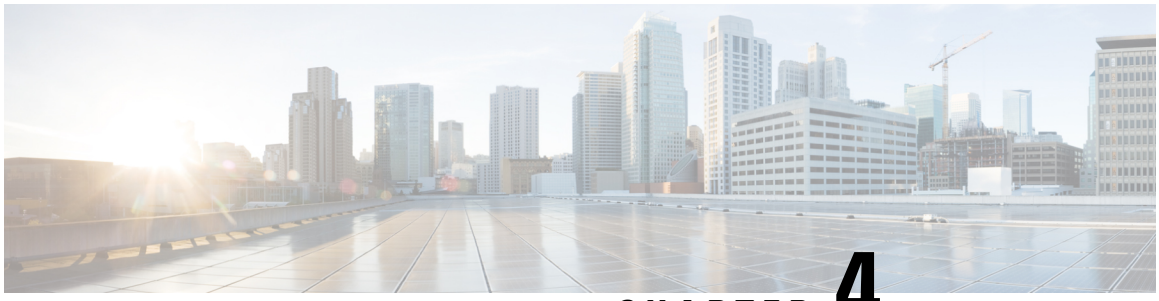
```
f2-leaf1# show ptp counters int e1/52

PTP Packet Counters of Interface Eth1/52:
------------------------------------------------------------------
Packet Type                  TX                      RX
---------------    --------------------    --------------------
Announce                           0                       4
Sync                               0                      59
FollowUp                           0                      59
Delay Request                     30                       0
Delay Response                     0                      30
PDelay Request                     0                       0
PDelay Response                    0                       0
PDelay Followup                    0                       0
Management                         0                       0
```

The sent and received messages are the opposite of a master port. For example, if the Rx of `Delay Request` and the Tx of `Delay Response` are zero on a master port, the other side is not configured or not working as a client correctly since the client should initiate a `Delay Request` for the E2E delay mechanism.

In a real world, the counter information may not be as clean as the example, because the port state may have changed in the past. In such a case, clear the counters with the following command:

```
f2-leaf1# clear ptp counters all
```

Note The `PDelay_xxx` counter is for the P2P mechanism, which is not supported on Cisco Application Centric Infrastructure (ACI).



CHAPTER 4

Synchronous Ethernet (SyncE)

About Synchronous Ethernet (SyncE)

With Ethernet equipment gradually replacing Synchronous Optical Networking (SONET) and Synchronous Digital Hierarchy (SDH) equipment in service-provider networks, frequency synchronization is required to provide high-quality clock synchronization over Ethernet ports. Frequency or timing synchronization is the ability to distribute precision frequency around a network. In this context, timing refers to precision frequency, not an accurate time of day.

Synchronous Ethernet (SyncE), described in ITU G.781, provides the required synchronization at the physical level. In SyncE, Ethernet links are synchronized by timing their bit clocks from high-quality, stratum-1-traceable clock signals in the same manner as SONET/SDH.

To maintain SyncE links, a set of operational messages are required. These messages ensure that a node is always deriving timing information from the most reliable source and then transfers the timing source quality information to clock the SyncE link. In SONET/SDH networks, these are known as Synchronization Status Messages (SSMs). SyncE uses Ethernet Synchronization Message Channel (ESMC) to provide transport for SSMs.

Customers using a packet network find it difficult to provide timing to multiple remote network elements (NEs) through an external time division multiplexed (TDM) circuit. The SyncE feature helps to overcome this problem by providing effective timing to the remote NEs through a packet network. SyncE synchronizes clock frequency over an Ethernet port, leveraging the physical layer of the Ethernet to transmit frequency to the remote sites. SyncE's functionality and accuracy resemble the SONET/SDH network because of its physical layer characteristic.

SONET/SDH use 4 bits from the two S bytes in the SONET/SDH overhead frame for message transmission. Ethernet relies on ESMC that is based on an IEEE 802.3 organization-specific slow protocol for message transmission. Each NE along the synchronization path supports SyncE, and SyncE effectively delivers frequency in the path. SyncE does not support relative time (for example, phase alignment) or absolute time (Time of Day).

SyncE provides the Ethernet physical layer network (ETY) level frequency distribution of known common precision frequency references. Clocks for use in SyncE are compatible with the clocks used in the SONET/SDH

synchronization network. To achieve network synchronization, synchronization information is transmitted through the network via synchronous network connections with performance of egress clock.

ESMC carries a Quality Level (QL) identifier that identifies the timing quality of the synchronization trail. QL values in QL-TLV are the same as QL values defined for SONET and SDH SSM. Information provided by SSM QLs during the network transmission helps a node derive timing from the most reliable source and prevents timing loops. ESMC is used with the synchronization selection algorithms. Because Ethernet networks are not required to be synchronous on all links or in all locations, the ESMC channel provides this service. ESMC, described in G.8264, is composed of the standard Ethernet header for an organization-specific slow protocol; the ITU-T OUI, a specific ITU-T subtype; an ESMC-specific header; a flag field; and a type, length, value (TLV) structure. The use of flags and TLVs improves the management of SyncE links and the associated timing change.

Sources and Selection Points

A Frequency Synchronization implementation involves Sources and Selection Points.

A Source inputs frequency signals into a system or transmits them out of a system. There are four types of sources:

- Line interfaces, including SyncE interfaces.
- Clock interfaces. These are external connectors for connecting other timing signals, such as BITS, UTI and GPS.
- PTP clock. If IEEE 1588 version 2 is configured on the router, a PTP clock may be available to frequency synchronization as a source of the time-of-day and frequency.
- Internal oscillator. This is a free-running internal oscillator chip.

Each source has an associated Quality Level (QL), which specifies the accuracy of the clock. This QL information is transmitted across the network using SSMs carried by ESMC. The QL information is used to determine the best available source to which the devices in the system can synchronize.

To define a predefined network synchronization flow and to prevent timing loops, you can assign priority values to each source on the switch. When more than one source has the same QL, the user-assigned priority value determines the relative preference among the sources.

A selection point is the process within the switch where a choice is made between several available frequency signals. The combination of QL information and user-assigned priority levels allows each switch to choose a source to synchronize its SyncE interfaces, as described in the ITU standard G.781.

Guidelines and Limitations for SyncE

SyncE has the following configuration guidelines and limitations:

- SyncE is supported on the N9K-C93180YC-FX3 switch.
- SyncE can be enabled only on downstream front panel ports. The interface can be switched, routed, or a subinterface.
- In the 5.2(3) release and earlier, SyncE is not supported on an SVI, port channel, or vPC or on its member interfaces.

- In the 5.2(4) release and later, SyncE is not supported on an SVI, a vPC, or its member interfaces. SyncE is supported on a port channel.
- Starting with the 5.2(4) release, SyncE can be enabled only on downstream front panel ports. The interface can be switched, or a routed physical interface, port channel, or subinterface.
- Starting with the 5.2(4) release, SyncE on non-vPC port channel interfaces are supported. When enabling SyncE on port channel interfaces, SyncE is configured per port channel and enabled on all of its member interfaces. Enabling SyncE per port channel member interface is not supported.
- SyncE is not supported on the fabric ports.
- We do not recommend configuring SyncE on a non-fabric port that is connected to another leaf switch.
- Local distribution of SyncE is supported. This is the case in which both the reference source and the client are on the same leaf switch. The leaf switch can be within a pod or a remote leaf switch.
- SyncE distribution over peer-link between two remote leaf switches is not supported.
- Hybrid mode with Precision Time Protocol (PTP) is supported for telecom profile ITU-T G8275.1.
- The switch can monitor a maximum of four downlink SyncE sources. The switch can lock to one of these sources.
- Each quad port group on the PHY provides one reference clock. For example, a leaf switch can monitor and lock to one source when interfaces 1/1 to 1/4 are connected to four different sources.
- Extended SSM or Extended QL TLV format are not supported.
- GPS and GNSS are not supported.
- SyncE is supported on all qualified optics with the exception of copper Gigabit Ethernet SFPs.

Configuring Synchronous Ethernet

To enable SyncE on a leaf switch, you must create two levels of policy:

- A node level policy enables the SyncE process on the leaf or remote leaf switch. This policy specifies the global Quality Level (QL) option configuration for the SyncE node.
- An interface level policy configures SyncE properties on the interface. This policy can also enable QL level override specific to an interface. The QL option in the interface policy should match the QL option in the node level policy.

Creating a Synchronous Ethernet Node Policy

This procedure creates a node level configuration policy for SyncE.

Procedure

Step 1 On the menu bar, choose **Fabric > Access Policies**.

Step 2 In the Navigation pane, choose **Policies > Switch > Synchronous Ethernet Node**.

Step 3 Right-click **Synchronous Ethernet Node** and choose **Create Synchronous Ethernet Node Policy**.

Step 4 In the **Create Synchronous Ethernet Node Policy** dialog box, complete the following steps:

a) Enter a **Name** for the policy.

b) Enter a **Description** of the policy.

c) Set the **Admin State** control to either **Enabled** to activate the policy or **Disabled** (default) to deactivate the policy.

d) In the **QL Option** drop-down list, choose the quality level.

Choose one of the following ITU-T Quality Level (QL) options:

- **Option 1**: Includes DNU, EEC1, PRC, PRTC, SEC, SSU-A, SSU-B, eEEC and ePRTC.
- **Option 2 generation 1**: Includes DUS, EEC2, PRS, PRTC, RES, SMC, ST2, ST3, ST4, STU, eEEC and ePRTC.
- **Option 2 generation 2**: Includes DUS, EEC2, PROV, PRS, PRTC, SMC, ST2, ST3, ST3E, ST4, STU, TNC, eEEC and ePRTC.

Note
Extended SSM QL options PRTC, eEEC, and ePRTC are not supported.

Stratum 4 freerun (ST4) is not supported on Ethernet line interfaces.

For details of QL mapping for these options, see QL Mapping with ACI Configuration Options, on page 74.

Note
The **Quality Level Option** is typically configured here instead of at the interface level. If configured at the interface level, the QL option there must match the QL selected here.

e) (Optional) Beginning with the 5.2(4) release, enable the **Transmit DNU on Lag Members** feature.

When this option is enabled on the node and one of the port channel member ports is locked as the SyncE source, other member ports send QL-DNU (Do Not Use) using SyncE ESMC messages to prevent potential timing issues in selecting the SyncE input port. This feature enables compliance with 11.1.1 ESMC operation with link aggregation in G.8264.

f) Click **Submit**.

What to do next

Add the policy to an Access Switch Policy Group at **Fabric > Access Policies > Switches > Leaf Switches > Policy Groups**.

Creating a Synchronous Ethernet Interface Policy

This procedure creates an interface level configuration policy for Synchronous Ethernet (SyncE).

A SyncE interface policy allows you to configure an Ethernet interface as a frequency synchronization input and output. Configuring an interface as an input (using **Selection Input**) allows the interface to be passed to the selection algorithm to be considered as a timing source for frequency synchronization.

If the interface is locked to an input, the interface will always transmit synchronized to the selected frequency signal.

Procedure

Step 1 On the menu bar, choose **Fabric > Access Policies**.

Step 2 In the Navigation pane, choose **Policies > Interface > Synchronous Ethernet Interface**.

Step 3 Right-click **Synchronous Ethernet Interface** and choose **Create Synchronous Ethernet Interface Policy**.

Step 4 In the **Create Synchronous Ethernet Interface Policy** dialog box, complete the following steps:

a) Enter a **Name** for the policy.

b) Enter a **Description** of the policy.

c) Set the **Admin State** control to either **Enabled** to activate the policy or **Disabled** (default) to deactivate the policy.

d) Check or uncheck the **Synchronization Status Message** checkbox.

If unchecked, disables sending ESMC packets and also ignores any received ESMC packets. This checkbox is checked by default.

e) Check or uncheck the **Selection Input** checkbox.

If checked, assigns the interface as a timing source to be passed to the selection algorithm. This checkbox is unchecked by default.

f) Click the up or down controls to set the **Source Priority**.

The priority of the frequency source on an interface. This value is used in the clock-selection algorithm to choose between two sources that have the same QL. Values can range from 1 (highest priority) to 254 (lowest priority). The default value is 100.

Note
This setting is active only if **Selection Input** is checked.

g) Click the up or down controls to set the **Wait-To-Restore** time in minutes.

The wait-to-restore time, in minutes, is the amount of time after the interface comes up before it is used for frequency synchronization on an interface. Values can range from 0 to 12 minutes. The default value is 5.

Note
This setting is active only if **Selection Input** is checked.

h) In the **Quality Level Option** drop-down list, select the quality level (QL).

This setting allows you to specify or override the Quality Level (QL) received or transmitted at the interface level. The ITU-T Quality Level options are:

- **No Quality Level configured**: (Default) The QL received from the connected source via ESMC is used for frequency synchronization.
- **Option 1**: Includes DNU, EEC1, PRC, PRTC, SEC, SSU-A, SSU-B, eEEC and ePRTC.
- **Option 2 generation 1**: Includes DUS, EEC2, PRS, PRTC, RES, SMC, ST2, ST3, ST4, STU, eEEC and ePRTC.
- **Option 2 generation 2**: Includes DUS, EEC2, PROV, PRS, PRTC, SMC, ST2, ST3, ST3E, ST4, STU, TNC, eEEC and ePRTC.

Note
Extended SSM QL options PRTC, eEEC, and ePRTC are not supported.

Stratum 4 freerun (ST4) is not supported on Ethernet line interfaces.

For details of QL mapping for these options, see QL Mapping with ACI Configuration Options, on page 74.

i) If you selected a **Quality Level Option**, you can configure either or both of the **Quality Receive** and **Quality Transmit** values.

The Quality Receive values allow you to override the received QL value in the SSM messages, which is used in the selection algorithm. The choices are as follows:

- **Exact Value**: Use the exact QL, regardless of the value received, unless the received value is Do Not Use (DNU).
- **Highest Value**: Sets an upper limit on the received QL. If the received value is higher than this specified QL, this QL is used instead.
- **Lowest Value**: Sets a lower limit on the received QL. If the received value is lower than this specified QL, DNU is used instead.

The Quality Transmit values allow you to override the QL value to be transmitted in the SSM messages. The choices are as follows:

- **Exact Value**: Use the exact QL unless Do Not Use (DNU) would otherwise be sent.
- **Highest Value**: Sets an upper limit on the QL to be sent. If the selected source has a higher QL than the QL specified here, this QL is sent instead.
- **Lowest Value**: Sets a lower limit on the QL to be sent. If the selected source has a lower QL than the QL specified here, DNU is sent instead.

Note
The quality options specified in these settings must match the configured QL option in the Synchronous Ethernet Node Policy for the switch.

Step 5 Click **Submit**.

What to do next

Add the policy to a Leaf Access Port Policy Group at **Fabric > Access Policies > Interfaces > Leaf Interfaces > Policy Groups > Leaf Access Port**.

QL Mapping with ACI Configuration Options

The following tables list the clock-source quality level (QL) value selections in the synchronous Ethernet policy configuration.

For details about these QL options, see *ITU-T G.781, Synchronization layer functions for frequency synchronization based on the physical layer*.

ITU-T Option 1

Quality Transmit/Receive Value	Quality Level
This signal should not be used for synchronization	QL-DNU

Quality Transmit/Receive Value	Quality Level
Quality common failed	QL-FAILED (see Notes)
Quality common invalid	QL-INVx (see Notes)
Quality common none	(see Notes)
ITU-T Option 1: Ethernet equipment clock	QL-SEC/QL-EEC1
ITU-T Option 1: Enhanced Ethernet equipment clock	QL-eEEC not supported Translated to QL-SEC/QL-EEC1 (see Notes)
ITU-T Option 1: Enhanced primary reference timing clock	QL-ePRTC not supported Translated to QL-PRC (see Notes)
ITU-T Option 1: Primary reference clock	QL-PRC
ITU-T Option 1: Primary reference timing clock	QL-PRTC not supported Translated to QL-PRC (see Notes)
ITU-T Option 1: SONET equipment clock	QL-SEC
ITU-T Option 1: Type I or V slave clock	QL-SSU-A
ITU-T Option 1: Type IV slave clock	QL-SSU-B

ITU-T Option 2, Generation 1

Quality Transmit/Receive Value	Quality Level
This signal should not be used for synchronization	QL-DUS
Quality common failed	QL-FAILED (see Notes)
Quality common invalid	QL-INVx (see Notes)
Quality common none	(see Notes)
ITU-T Option 2, Generation 1: Ethernet equipment clock	QL-EEC2

Quality Transmit/Receive Value	Quality Level
ITU-T Option 2, Generation 1: Enhanced Ethernet equipment clock	QL-eEEC not supported Translated to QL-ST3 (see Notes)
ITU-T Option 2, Generation 1: Enhanced primary reference timing clock	QL-ePRTC not supported Translated to QL-PRS (see Notes)
ITU-T Option 2, Generation 1: Primary reference source	QL-PRS
ITU-T Option 2, Generation 1: Primary reference timing clock	QL-PRTC not supported Translated to QL-PRS (see Notes)
ITU-T Option 2, Generation 1: RES	QL-RES
ITU-T Option 2, Generation 1: SONET clock self timed	QL-SMC
ITU-T Option 2, Generation 1: Stratum 2	QL-ST2
ITU-T Option 2, Generation 1: Stratum 3	QL-ST3
ITU-T Option 2, Generation 1: Stratum 4 freerun	(see Notes)
ITU-T Option 2, Generation 1: Synchronized - traceability unknown	QL-STU

ITU-T Option 2, Generation 2

Quality Transmit/Receive Value	ITU Quality Level
This signal should not be used for synchronization	QL-DUS
Quality common failed	QL-FAILED (see Notes)
Quality common invalid	QL-INVx (see Notes)
Quality common none	(see Notes)
ITU-T Option 2, Generation 2: Ethernet equipment clock	QL-EEC2
ITU-T Option 2, Generation 2: Enhanced Ethernet equipment clock	QL-eEEC not supported Translated to QL-ST3 (see Notes)

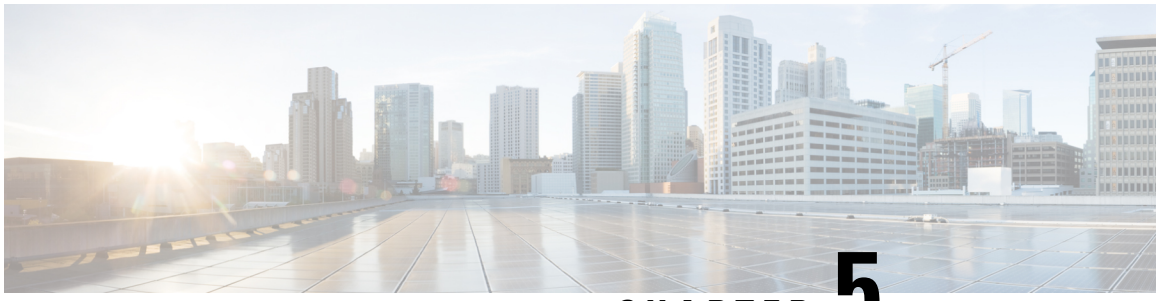
Quality Transmit/Receive Value	ITU Quality Level
ITU-T Option 2, Generation 2: Enhanced primary reference timing clock	QL-ePRTC not supported Translated to QL-PRS (see Notes)
ITU-T Option 2, Generation 2: PROV	QL-PROV
ITU-T Option 2, Generation 2: Primary reference source	QL-PRS
ITU-T Option 2, Generation 2: Primary reference timing clock	QL-PRTC not supported Translated to QL-PRS (see Notes)
ITU-T Option 2, Generation 2: SONET clock self timed	QL-SMC
ITU-T Option 2, Generation 2: Stratum 2	QL-ST2
ITU-T Option 2, Generation 2: Stratum 3	QL-ST3
ITU-T Option 2, Generation 2: Stratum 3E	QL-ST3E
ITU-T Option 2, Generation 2: Stratum 4 freerun	(see Notes)
ITU-T Option 2, Generation 2: Synchronized - traceability unknown	QL-STU
ITU-T Option 2, Generation 2: Transit node clock	QL-TNC

Notes

- The "quality common none" QL is the default when no QL is configured.
- The quality levels "quality common invalid" (QL-INVx) and "quality common failed" (QL-FAILED) are internal quality levels inside the leaf or remote leaf switch and are never generated at an output port.
- ITU-T Option 2, Generation 1 and Generation 2: Stratum 4 freerun (QL-ST4) is not supported on Ethernet line interfaces.
- Extended QL TLV (type-length-value) is not supported. When an extended QL TLV is received in an ESMC frame from the connected frequency source, the leaf or remote leaf switch will process the received ESMC frame but only honor the Standard TLV, ignoring the specified Extended TLV.
- Several QL values are described by combining standard QL TLV and extended QL TLV. These values are translated on ACI leaf nodes to QL values that can be described only with standard QL TLV. The translations are shown in the following tables:

Extended TLV	Description	Translated/Effective QL
ITU-T Option 1		
QL-PRTC	ITU-T Option 1: Primary reference timing clock	QL-PRC

Extended TLV	Description	Translated/Effective QL
QL-eEEC	ITU-T Option 1: Enhanced ethernet equipment clock	QL-SEC/QL-EEC1
QL-ePRTC	ITU-T Option 1: Enhanced primary reference timing clock	QL-PRC
ITU-T Option 2		
QL-PRTC	ITU-T Option 2, Generation 1 and Generation 2: Primary reference timing clock	QL-PRS
QL-eEEC	ITU-T Option 2, Generation 1 and Generation 2: Enhanced ethernet equipment clock	QL-ST3
QL-ePRTC	ITU-T Option 2, Generation 1 and Generation 2: Enhanced primary reference timing clock	QL-PRS



CHAPTER 5

HTTP/HTTPS Proxy Policy

About the HTTP/HTTPS Proxy Policy

Beginning with release 5.2(1), you can configure the HTTP or HTTPS proxy address on Cisco Application Policy Infrastructure Controller (APIC) for features that need Internet access. In addition to Cisco APIC features that automatically use the configured proxy addresses, the surrounding ecosystems of the Cisco APIC can also query the object `proxyServer` on the Cisco APIC so that the ecosystems can use the same proxy server as the Cisco APIC without needing you to configure the proxy information on multiple platforms.

The HTTP/HTTPS proxy policy itself does not control nor alter the management network (out-of-band or in-band) that each Cisco APIC feature uses. You can specify the management network setting in the Cisco APIC Connectivity Preferences. For more information, see the "Adding Management Access" section in the "Management" chapter of the *Cisco APIC Basic Configuration Guide*.

Cisco APIC Features That Use the HTTP/HTTPS Proxy

If you configured an HTTP or HTTPS proxy server, the following Cisco Application Policy Infrastructure Controller (APIC) features send the traffic through the proxy server:

- Cisco Intersight - Device Connector
- Cisco APIC GUI built-in feedback feature

Note Before release 5.2(1), Cisco Intersight - Device Connector had a built-in proxy setting. This functionality now exists in the HTTP/HTTPS proxy policy in Cisco APIC.

Configuring the HTTP/HTTPS Proxy Policy Using the GUI

The following procedure configures the HTTP or HTTPS proxy policy. You can also configure the proxy settings through the First Time Setup wizard. For more information about First Time Setup wizard, see the chapter "First Time Setup Wizard" in the *Cisco APIC Basic Configuration Guide*.

Procedure

Step 1 On the menu bar, choose **System** > **System Settings**.

Step 2 In the Navigation pane, choose **Proxy Policy**.

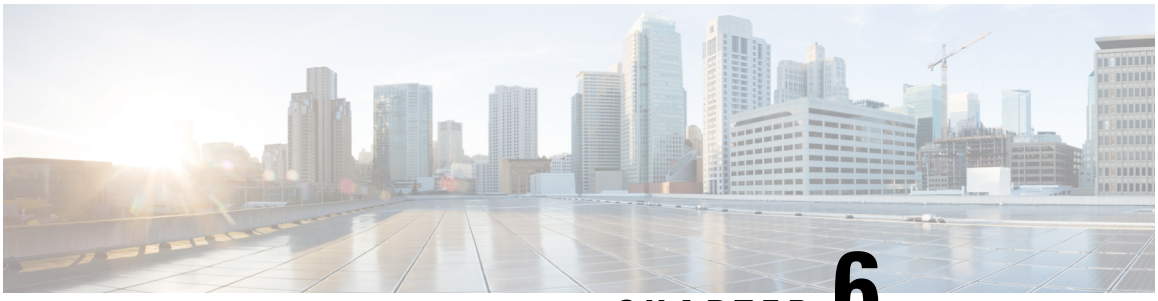
Step 3 In the Work pane, enter a URL in the **HTTP URL** or **HTTPS URL** field as appropriate.

When a proxy server requires authentication, use the following format:

```
http[s]://[username:password]@proxy-server[:proxyport]
```

Step 4 (Optional) In the **Ignore Hosts** table, click **+**, enter the hostname or IP address of a host that should not use the HTTP or HTTPS proxy, and click **Update**.

Repeat this step if you want to add more hosts that should not use the HTTP or HTTPS proxy.



CHAPTER 6

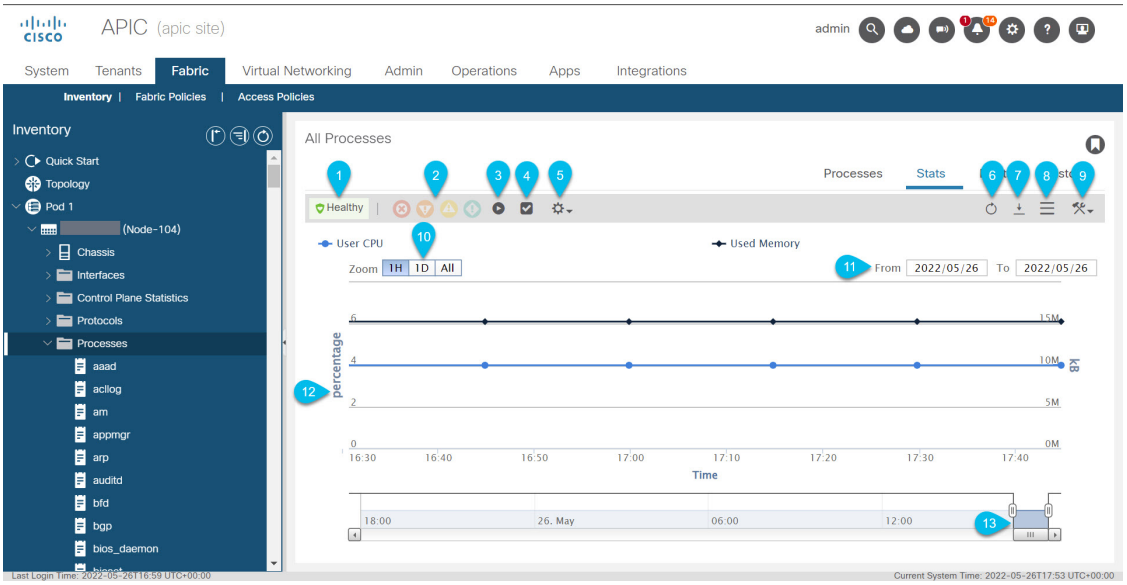
Process Statistics

Viewing the Statistics for Processes Using the GUI

To view the statistics for a process, on the menu bar, choose **Fabric** > **Inventory**. In the Navigation pane, perform one of the following actions:

- For all processes, choose ***pod_ID*** > ***node_name*** > **Processes**.
- For a specific process, choose ***pod_ID*** > ***node_name*** > **Processes** > ***process_name***.

In the Work pane, choose the **Stats** tab. The following screenshot shows an example of all processes, but the view for a specific process is nearly identical:



Callout	Description
1	The overall health of the process. Hover over this to see the health score.
2	The faults. Hover of this to see the number of faults for each severity. Click one of the severities to go to the **Faults** tab to see the faults of that severity.
3	Stops the GUI from displaying updated statistics, which enables you to examine the statistics as they were at the time that you clicked this button. Click the button again for the GUI to resume displaying updated statistics. Even when you stop the GUI from displaying updated statistics, the Cisco Application Policy Infrastructure Controller (APIC) continues to collect the latest statistics.
4	Opens the **Select Stats** dialog, which enables you to choose the sampling interval and select the statistics to view.
5	Enables you to choose the statistics type to view. • **Average**: Shows the average resources that each statistic used during the retention period. • **Min**: Shows the minimum resources that each statistic used during the retention period. • **Max**: Shows the maximum resources that each statistic used during the retention period. • **Trend**: Shows the trend in resource usage for each statistic during the retention period. • **Rate**: Shows the rate at which resources that each statistic used during the retention period. • **default**: Currently, this type shows the same information as the **Average** type.
6	Refreshes the statistics data.
7	Downloads the statistics data to your local system as an XML file. The file gets downloaded to your browser's default download location.
8	Toggles between the table and topology (graph) views.
9	Click this, then choose **Configure Statistics Policy** to open the **Create Stats Target** dialog. In the dialog, you can choose one or more statistics targets and configure the collections. The collections enable you to specify the retention period for each collection granularity, and enable or disable each granularity. For more information, see Configuring the Statistics Policy for All Processes After Configuring the Policy the First Time Using the GUI, on page 84.

Callout	Description
10	Only visible in the topology view, and only for the **15 Minute** and **1 Hour** sampling intervals. This sets the zoom to a preset value. The zoom specifies what time range to display in the topology. • **1H**: Sets the zoom to the last hour. • **1D**: Sets the zoom to the last day (the last 24 hours). • **1M**: Sets the zoom to the last minute. This choice is visible only if you chose the **1 Hour** sampling interval. • **All**: Sets the zoom to display the full range of time, which is slightly over 24 hours for the **15 Minute** sampling interval, and is the same as **1M** for the **1 Hour** sampling interval.
11	Only visible in the topology view, and only for the **15 Minute** and **1 Hour** sampling intervals. The date range of the topology. You can click the dates and change the values. You cannot enter a date that is not visible in the timeline at the bottom of the topology. The **From** date cannot be later than the **To** date.
12	In the topology view, this area shows a graph of the selected statistics. Hover over any of the epochs to see the exact data for all selected statistics at that time. In the table view, this area shows a table of the same statistics. You can sort the table by clicking any of the headers. You can filter the table by clicking the drop-down list arrow at the right side of a header, choosing **Columns**, then putting a check in or removing a check from any of the boxes.
13	Only visible in the topology view. This is the zoom, which specifies what time range to display in the topology. This enables you to set the zoom to an arbitrary amount. Drag the left side to specify the start of the zoom and the right sides specify the end of the zoom, which determines the length of time to display. After you set the start and end, you can use the horizontal scroll bar to change the portion of the time line that you are viewing while retaining the same length of time.

Configuring the Statistics Policy for All Processes for the First Time Using the GUI

This procedure describes how to configure the statistics policy for all processes for the first time since you brought up the Cisco Application Policy Infrastructure Controller (APIC). The GUI dialog is different if you previously configured the policy. In this case, see the Configuring the Statistics Policy for All Processes After Configuring the Policy the First Time Using the GUI, on page 84.

The Cisco APIC creates and stores one stats object whenever the granularity (time interval) of a collection passes. For example, for the 15 minute collection, after 1 hour passes, the Cisco APIC creates and stores 4 stats objects. The Cisco APIC stores up to 1,000 stats objects for each collection, except for the 5 minute granularity, of which the Cisco APIC stores only 12 stats objects.

Procedure

Step 1 On the menu bar, choose **Fabric** > **Inventory**.

Step 2 In the **Navigation** pane, choose ***pod_ID*** > ***node_name*** > **Processes**.

Step 3 In the **Work** pane, choose **Action** > **Configure Statistics Policy**.

The **Create Stats Target** dialog appears.

Step 4 In the **Available** area, select one or more of the stats types, then click the top gray button that is between the **Available** and **Selected** areas.

The selected stats types get moved to the **Selected** area. Any stats type that you do not select uses the default parameters from **Fabric** > **Fabric Policies** > **Policies** > **Monitoring** > **default** > **Stats Collection Policies** > **ALL**.

You can select multiple stats types by holidng Ctrl and clicking the desired stats types. You can also hold Shift and click the first and last stats type to select all of the stats types.

Step 5 Click **Next**.

Step 6 Double-click the row of a granularity to enable or disable that granularity and to change the history retention period, then click **Update**.

Repeat this step for each granularity that you want to modify. These values get applied to all of the selected stats types.

Step 7 Click **OK**.

Configuring the Statistics Policy for All Processes After Configuring the Policy the First Time Using the GUI

This procedure describes how to configure the statistics policy for all processes after you configured the policy for the first time. The GUI dialog is different if you have not previously configured the policy. In this case, see the Configuring the Statistics Policy for All Processes for the First Time Using the GUI, on page 83.

The Cisco Application Policy Infrastructure Controller (APIC) creates and stores one stats object whenever the granularity (time interval) of a collection passes. For example, for the 15 minute collection, after 1 hour passes, the Cisco APIC creates and stores 4 stats objects. The Cisco APIC stores up to 1,000 stats objects for each collection, except for the 5 minute granularity, of which the Cisco APIC stores only 12 stats objects.

Procedure

Step 1 On the menu bar, choose **Fabric** > **Inventory**.

Step 2 In the **Navigation** pane, choose ***pod_ID*** > ***node_name*** > **Processes**.

Step 3 In the **Work** pane, choose **Action** > **Configure Statistics Policy**.

The **Edit Stats Policy default** dialog appears.

Step 4 In the **Collections and Thresholds** tab, expand **System CPU**, **System load**, or **System memory** as desired.

System CPU, **System load**, and **System memory** each appear only if you configured them previously.

Step 5 To edit a collection, click the edit button (pencil icon) to the right of the desired collection interval.

The **Stats Collection and Thresholds** dialog appears for that collection interval. The collections specify whether the Cisco APIC collects the stats for a specific granularity and how long the Cisco APIC retains the collected stats.

a) Under the **Policy** tab, set the properties as desired.

Property	Description
Granularity	The granularity of the collection that you are editing. You cannot change this value.
Admin State	The administrative state of the collection. The possible values are: • **disabled**: Disables this collection, meaning that the Cisco APIC will not collect the stats for this collection granularity. • **enabled**: Enables this collection, meaning that the Cisco APIC collects the stats for this collection granularity. • **inherited**: This collection inherits its administrative state from the default policy. You can view and edit the default policy by navigating to **Fabric** > **Fabric Policies**, then **Policies** > **Monitoring** > **default** > **Stats Collection Policies**.
History Retention Period	The length of time that the Cisco APIC retains a stats object.

b) Under the **Thresholds** tab, you can edit or delete any configured thresholds.
c) Under the **History** tab, you can view the events and audit log.
d) After you finish making changes, click **Submit**.

Step 6 To configure a threshold, click the **+** button to the right of the desired collection interval and choose a property.

The **Create Stats Threshold** dialog appears for that collection interval. The thresholds specify that the Cisco APIC will set a fault when the value of a specific stat reaches or exceeds a specific value.

a) Set the properties as desired.

Property	Description
Normal Value	The baseline value for the thresholds.
Threshold Direction	Specifies whether you can set thresholds for the stats values as they rise or fall, or both. • **Both**: You can configure thresholds for both increases and decreases in the stat's value. • **Rising**: You can configure thresholds for only increases in the stat's value. • **Falling**: You can configure thresholds for only decreases in the stat's value.

Property	Description
Rising Thresholds to Config	This is visible only if you chose **Both** or **Rising** for **Threshold Direction**. Put a check in the box for each of the fault severities that you want the Cisco APIC to set as the stat's value rises.
Falling Thresholds to Config	This is visible only if you chose **Both** or **Falling** for **Threshold Direction**. Put a check in the box for each of the fault severities that you want the Cisco APIC to set as the stat's value falls.
Rising area	This is visible only if you chose **Both** or **Rising** for **Threshold Direction**. In this area, you specify the stat value that sets or resets a fault of the specified severity. The **Set** and **Reset** values can be the same. The **Reset** value cannot be greater than the **Set** value. The values for the different severities can be the same, but the value for a lower severity cannot be greater than the value for a higher severity. For example, if **Critical** has a **Set** value of 70, **Major** can have a **Set** value of 70 or less.
Falling area	This is visible only if you chose **Both** or **Falling** for **Threshold Direction**. In this area, you specify the stat value that sets or resets a fault of the specified severity. The **Set** and **Reset** values can be the same. The **Set** value cannot be greater than the **Reset** value. The values for the different severities can be the same, but the value for a higher severity cannot be greater than the value for a lower severity. For example, if **Minor** has a **Set** value of 50, **Major** can have a **Set** value of 50 or less.

b) Click **Submit**.

Step 7 (Optional) In the **Reportables** tab, you can specify for which stats type you want to configure dedicated parameters for the collections and thresholds.

Any stats type for which you do not configure dedicated parameters use the default parameters from **Fabric** > **Fabric Policies** > **Policies** > **Monitoring** > **default** > **Stats Collection Policies** > **ALL**.

Reportables are referred to as "monitoring objects" in other parts of the GUI.

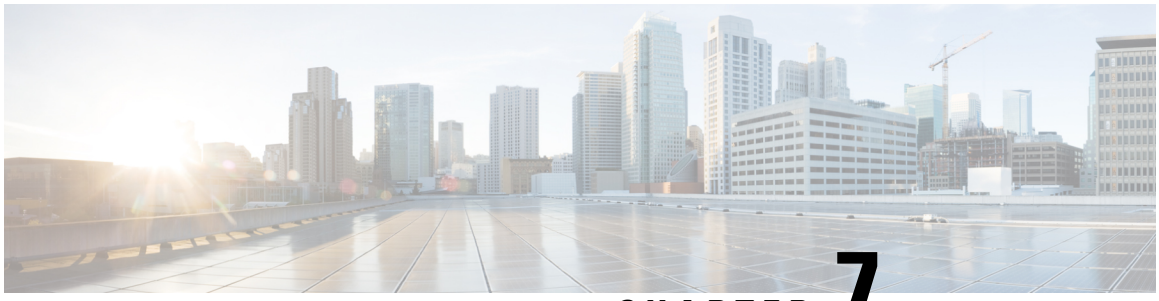
a) In the **Add/Remove Reportables** area, put a check in the box for any stats type for which you want to configure the collection and threshold parameters that are dedicated for that stats type.

After you add the stats type from here, the stats type will appear in the **Collections and Thresholds** tab and you can modify the dedicated parameters from there. If the stats type must use the parameters from the default stats policy, remove the check from the box.

b) In the **Configure Collections for New Reportables** table, you can set the initial collection parameters that are dedicated for the stats type.

However, the parameters from this table will not take effect for a stats type that is already configured with a dedicated set of parameters, as in the stats type already has a check in the box in the **Add/Remove Reportables** area. For those stats types, go to the **Collections and Thresholds** tab and modify the dedicated parameters from there.

Step 8 Click **Submit**.



CHAPTER 7

Basic Operations

Troubleshooting APIC Crash Scenarios

Cluster Troubleshooting Scenarios

The following table summarizes common cluster troubleshooting scenarios for the Cisco APIC.

Problem	Solution
An APIC node fails within the cluster. For example, node 2 of a cluster of 5 APICs fails.	There are two available solutions: • Leave the target size and replace the APIC. • Reduce the cluster size to 4, decommission controller 5, and recommission it as APIC 2.The target size remains 4, and the operational size is 4 when the reconfigured APIC becomes active. **Note** You can add a replacement APIC to the cluster and expand the target and operational size. For instructions on how to add a new APIC, refer to the *Cisco APIC Management, Installation, Upgrade, and Downgrade Guide*..

Problem	Solution
A new APIC connects to the fabric and loses connection to a leaf switch.	Use the following commands to check for an infra (infrastructure) VLAN mismatch: • cat /mit/sys/lldp/inst/if-\[eth1--1\]/ctrlradj/summary—Displays the VLAN configured on the leaf switch. • cat /mit/sys/lldp/inst/if-\[eth1--1\]/ctrlradj/summary—Displays the infra (infrastructure) VLANs advertised by connected APICs. If the output of these commands shows different VLANs, the new APIC is not configured with the correct infra (infrastructure) VLAN. To correct this issue, follow these steps: • Log in to the APIC using rescue-user. **Note** Admin credentials do not work because the APIC is not part of the fabric. • Erase the configuration and reboot the APIC using the **acidiag touch setup** command. • Reconfigure the APIC. Verify that the fabric name, TEP addresses, and infra (infrastructure) VLAN match the APICs in the cluster. • Reload the leaf node.
Two APICs cannot communicate after a reboot.	The issue can occur after the following sequence of events: • APIC1 and APIC2 discover each other. • APIC1 reboots and becomes active with a new ChassisID (APIC1a) • The two APICs no longer communicate. In this scenario, APIC1a discovers APIC2, but APIC2 is unavailable because it is in a cluster with APIC1, which appears to be offline. As a result, APIC1a does not accept messages from APIC2. To resolve the issue, decommission APIC1 on APIC2, and commission APIC1 again.
A decommissioned APIC joins a cluster.	The issue can occur after the following sequence of events: • A member of the cluster becomes unavailable or the cluster splits. • An APIC is decommissioned. • After the cluster recovers, the decommissioned APIC is automatically commissioned. To resolve the issue, decommission the APIC after the cluster recovers.

Problem	Solution
Mismatched ChassisID following reboot.	The issue occurs when an APIC boots with a ChassisID different from the ChassisID registered in the cluster. As a result, messages from this APIC are discarded. To resolve the issue, ensure that you decommission the APIC before rebooting.
The APIC displays faults during changes to cluster size.	A variety of conditions can prevent a cluster from extending the OperationalClusterSize to meet the AdminstrativeClusterSize. For more information, inspect the fault and review the "Cluster Faults" section in the *Cisco APIC Basic Configuration Guide*.
An APIC is unable to join a cluster.	The issue occurs when two APICs are configured with the same ClusterID when a cluster expands. As a result, one of the two APICs cannot join the cluster and displays an expansion-contender-chassis-id-mismatch fault. To resolve the issue, configure the APIC outside the cluster with a new cluster ID.
APIC unreachable in cluster.	Check the following settings to diagnose the issue: • Verify that fabric discovery is complete. • Identify the switch that is missing from the fabric. • Check whether the switch has requested and received an IP address from an APIC. • Verify that the switch has loaded a software image. • Verify how long the switch has been active. • Verify that all processes are running on the switch. For more information, see the "acidiag Command" section in the *Cisco APIC Basic Configuration Guide*. • Confirm that the missing switch has the correct date and time. • Confirm that the switch can communicate with other APICs.

Problem	Solution
Cluster does not expand.	The issue occurs under the following circumstances: • The OperationalClusterSize is smaller than the number of APICs. • No expansion contender (for example, the admin size is 5 and there is not an APIC with a clusterID of 4. • There is no connectivity between the cluster and a new APIC • Heartbeat messages are rejected by the new APIC • System is not healthy. • An unavailable appliance is carrying a data subset that is related to relocation. • Service is down on an appliance with a data subset that is related to relocation. • Unhealthy data subset related to relocation.
An APIC is down.	Check the following: • Connectivity issue—Verify connectivity using ping. • Interface type mismatch—Confirm that all APICs are set to in-band communication. • Fabric connectivity—Confirm that fabric connectivity is normal and that fabric discovery is complete. • Heartbeat rejected—Check the fltInfraIICIMsgSrcOutsider fault. Common errors include operational cluster size, mismatched ChassisID, source ID outside of the operational cluster size, source not commissioned, and fabric domain mismatch.

Cluster Faults

The APIC supports a variety of faults to help diagnose cluster problems. The following sections describe the two major cluster fault types.

Discard Faults

The APIC discards cluster messages that are not from a current cluster peer or cluster expansion candidate. If the APIC discards a message, it raises a fault that contains the originating APIC's serial number, cluster ID, and a timestamp. The following table summarizes the faults for discarded messages:

Fault	Meaning
expansion-contender-chassis-id-mismatch	The ChassisID of the transmitting APIC does not match the ChassisID learned by the cluster for expansion.
expansion-contender-fabric-domain-mismatch	The FabricID of the transmitting APIC does not match the FabricID learned by the cluster for expansion.

Fault	Meaning
expansion-contender-id-is-not-next-to-oper-cluster-size	The transmitting APIC has an inappropriate cluster ID for expansion. The value should be one greater than the current OperationalClusterSize.
expansion-contender-message-is-not-heartbeat	The transmitting APIC does not transmit continuous heartbeat messages.
fabric-domain-mismatch	The FabricID of the transmitting APIC does not match the FabricID of the cluster.
operational-cluster-size-distance-cannot-be-bridged	The transmitting APIC has an OperationalClusterSize that is different from that of the receiving APIC by more than 1. The receiving APIC rejects the request.
source-chassis-id-mismatch	The ChassisID of the transmitting APIC does not match the ChassisID registered with the cluster.
source-cluster-id-illegal	The transmitting APIC has a clusterID value that is not permitted.
source-has-mismatched-target-chassis-id	The target ChassisID of the transmitting APIC does not match the Chassis ID of the receiving APIC.
source-id-is-outside-operational-cluster-size	The transmitting APIC has a cluster ID that is outside of the OperationalClusterSize for the cluster.
source-is-not-commissioned	The transmitting APIC has a cluster ID that is currently decommissioned in the cluster.

Cluster Change Faults

The following faults apply when there is an error during a change to the APIC cluster size.

Fault	Meaning
cluster-is-stuck-at-size-2	This fault is issued if the OperationalClusterSize remains at 2 for an extended period. To resolve the issue, restore the cluster target size.
most-right-appliance-remains-commissioned	The last APIC within a cluster is still in service, which prevents the cluster from shrinking.
no-expansion-contender	The cluster cannot detect an APIC with a higher cluster ID, preventing the cluster from expanding.
service-down-on-appliance-carrying-replica-related-to-relocation	The data subset to be relocated has a copy on a service that is experiencing a failure. Indicates that there are multiple such failures on the APIC.
unavailable-appliance-carrying-replica-related-to-relocation	The data subset to be relocated has a copy on an unavailable APIC. To resolve the fault, restore the unavailable APIC.
unhealthy-replica-related-to-relocation	The data subset to be relocated has a copy on an APIC that is not healthy. To resolve the fault, determine the root cause of the failure.

APIC Unavailable

The following cluster faults can apply when an APIC is unavailable:

Fault	Meaning
fltInfraReplicaReplicaState	The cluster is unable to bring up a data subset.
fltInfraReplicaDatabaseState	Indicates a corruption in the data store service.
fltInfraServiceHealth	Indicates that a data subset is not fully functional.
fltInfraWiNodeHealth	Indicates that an APIC is not fully functional.

Troubleshooting Fabric Node and Process Crash

The ACI switch node has numerous processes which control various functional aspects on the system. If the system has a software failure in a particular process, a core file will be generated and the process will be reloaded.

If the process is a Data Management Engine (DME) process, the DME process will restart automatically. If the process is a non-DME process, it will not restart automatically and the switch will reboot to recover.

This section presents an overview of the various processes, how to detect that a process has cored, and what actions should be taken when this occurs

DME Processes

The essential processes running on an APIC can be found through the CLI. Unlike the APIC, the processes that can be seen via the GUI in **FABRIC** > **INVENTORY** > **Pod 1** > ***node*** shows all processes running on the leaf.

Through the **ps-ef | grep svc_ifc**:

```
rtp_leaf1# ps -ef |grep svc_ifc
root 3990 3087 1 Oct13 ? 00:43:36 /isan/bin/svc_ifc_policyelem --x
root 4039 3087 1 Oct13 ? 00:42:00 /isan/bin/svc_ifc_eventmgr --x
root 4261 3087 1 Oct13 ? 00:40:05 /isan/bin/svc_ifc_opflexelem --x -v
dptcp:8000
root 4271 3087 1 Oct13 ? 00:44:21 /isan/bin/svc_ifc_observerelem --x
root 4277 3087 1 Oct13 ? 00:40:42 /isan/bin/svc_ifc_dbgrelem --x
root 4279 3087 1 Oct13 ? 00:41:02 /isan/bin/svc_ifc_confelem --x
rtp_leaf1#
```

Each of the processes running on the switch writes activity to a log file on the system. These log files are bundled as part of the techsupport file but can be found via CLI access in /tmp/logs/ directory. For example, the Policy Element process log output is written into /tmp/logs/svc_ifc_policyelem.log.

The following is a brief description of the DME processes running on the system. This can help in understanding which log files to reference when troubleshooting a particular process or understand the impact to the system if a process crashed:

Process	Function
policyelem	Policy Element: Process logical MO from APIC and push concrete model to the switch
eventmgr	Event Manager: Processes local faults, events, health score
opflexelem	Opflex Element: Opflex server on switch

Process	Function
observerelem	Observer Element: Process local stats sent to APIC
dbgrelem	Debugger Element: Core handler
nginx	Web server handling traffic between the switch and APIC

Identify When a Process Crashes

When a process crashes and a core file is generated, a fault as well as an event is generated. The fault for the particular process is shown as a "process-crash" as shown in this syslog output from the APIC:

```
Oct 16 03:54:35 apic3 %LOG_LOCAL7-3-SYSTEM_MSG [E4208395][process-crash][major]
[subj-[dbgs/cores/node-102-card-1-svc-policyelem-ts-2014-10-16T03:54:55.000+00:00]/
rec-12884905092]Process policyelem cored
```

When the process on the switch crashes, the core file is compressed and copied to the APIC. The syslog message notification comes from the APIC.

The fault that is generated when the process crashes is cleared when the process is Troubleshooting Cisco Application Centric Infrastructure 275 restarted. The fault can be viewed via the GUI in the fabric history tab at **FABRIC** > **INVENTORY** > **Pod 1**.

Collecting the Core Files

The APIC GUI provides a central location to collect the core files for the fabric nodes.

An export policy can be created from **ADMIN** > **IMPORT/EXPORT** > **Export Policies** > **Core**. However, there is a default core policy where files can be downloaded directly.

The core files can be accessed via SSH/SCP through the APIC at /data/techsupport on the APIC where the core file is located. Note that the core file will be available at /data/ techsupport on one APIC in the cluster, the exact APIC that the core file resides can be found by the Export Location path as shown in the GUI. For example, if the Export Location begins with “files/3/”, the file is located on node 3 (APIC3).

APIC Process Crash Verification and Restart

Symptom 1

Process on switch fabric crashes. Either the process restarts automatically or the switch reloads to recover.

- **Verification**:

  As indicated in the overview section, if a DME process crashes, it should restart automatically without the switch restarting. If a non-DME process crashes, the process will not automatically restart and the switch will reboot to recover.

  Depending on which process crashes, the impact of the process core will vary.

  When a non-DME process crashes, this will typical lead to a HAP reset as seen on the console:

  ```
  [ 1130.593388] nvram_klm wrote rr=16 rr_str=ntp hap reset to nvram
  [ 1130.599990] obfl_klm writing reset reason 16, ntp hap reset
  [ 1130.612558] Collected 8 ext4 filesystems
  ```

- **Check Process Log**:

  The process which crashes should have at some level of log output prior to the crash. The output of the logs on the switch are written into the /tmp/logs directory. The process name will be part of the file name. For example, for the Policy Element process, the file is svc_ifc_policyelem.log

```
rtp_leaf2# ls -l |grep policyelem
-rw-r--r-- 2 root root 13767569 Oct 16 00:37 svc_ifc_policyelem.log
-rw-r--r-- 1 root root 1413246 Oct 14 22:10 svc_ifc_policyelem.log.1.gz
-rw-r--r-- 1 root root 1276434 Oct 14 22:15 svc_ifc_policyelem.log.2.gz
-rw-r--r-- 1 root root 1588816 Oct 14 23:12 svc_ifc_policyelem.log.3.gz
-rw-r--r-- 1 root root 2124876 Oct 15 14:34 svc_ifc_policyelem.log.4.gz
-rw-r--r-- 1 root root 1354160 Oct 15 22:30 svc_ifc_policyelem.log.5.gz
-rw-r--r-- 2 root root 13767569 Oct 16 00:37 svc_ifc_policyelem.log.6
-rw-rw-rw- 1 root root 2 Oct 14 22:06 svc_ifc_policyelem.log.PRESERVED
-rw-rw-rw- 1 root root 209 Oct 14 22:06 svc_ifc_policyelem.log.stderr
rtp_leaf2#
```

  There will be several files for each process located at /tmp/logs. As the log file increases in size, it will be compressed and older log files will be rotated off. Check the core file creation time (as shown in the GUI and the core file name) to understand where to look in the file. Also, when the process first attempts to come up, there be an entry in the log file that indicates “Process is restarting after a crash” that can be used to search backwards as to what might have happened prior to the crash.

- **Check Activity**:

  A process which has been running has had some change which then caused it to crash. In many cases the changes may have been some configuration activity on the system. What activity occurred on the system can be found in the audit log history of the system.

- **Contact TAC**:

  A process crashing should not normally occur. In order to understand better why beyond the above steps it will be necessary to decode the core file. At this point, the file will need to be collected and provided to the TAC for further processing.

  Collect the core file (as indicated above how to do this) and open up a case with the TAC.

Symptom 2

Fabric switch continuously reloads or is stuck at the BIOS loader prompt.

- **Verification**:

  If a DME process crashes, it should restart automatically without the switch restarting. If a non-DME process crashes, the process will not automatically restart and the switch will reboot to recover. However in either case if the process continuously crashes, the switch may get into a continuous reload loop or end up in the BIOS loader prompt.

```
[ 1130.593388] nvram_klm wrote rr=16 rr_str=policyelem hap reset to nvram
[ 1130.599990] obfl_klm writing reset reason 16, policyelem hap reset
[ 1130.612558] Collected 8 ext4 filesystems
```

- **Break the HAP Reset Loop**:

  First step is to attempt to get the switch back into a state where further information can be collected.

  If the switch is continuously rebooting, when the switch is booting up, break into the BIOS loader prompt through the console by typing CTRL C when the switch is first part of the boot cycle.

  Once the switch is at the loader prompt, enter in the following commands:

- cmdline no_hap_reset
- boot

The cmdline command will prevent the switch from reloading with a hap reset is called. The second command will boot the system. Note that the boot command is needed instead of a reload at the loader as a reload will remove the cmdline option entered.

Though the system should now remain up to allow better access to collect data, whatever process is crashing will impact the functionality of the switch.

As in the previous table, check the process log, activity, and contact TAC steps.

Troubleshooting an APIC Process Crash

The APIC has a series of Data Management Engine (DME) processes which control various functional aspects on the system. When the system has a software failure in a particular process, a core file will be generated and the process will be reloaded.

The following sections cover potential issues involving system processes crashes or software failures, beginning with an overview of the various system processes, how to detect that a process has cored, and what actions should be taken when this occurs. The displays taken on a working healthy system can then be used to identify processes that may have terminated abruptly.

DME Processes

The essential processes running on an APIC can be found either through the GUI or the CLI. Using the GUI, the processes and the process ID running is found in **System** > **Controllers** > **Processes**.

Using the CLI, the processes and the process ID are found in the summary file at /aci/system/controllers/1/processes (for APIC1):

```
admin@RTP_Apic1:processes> cat summary
processes:
process-id process-name max-memory-allocated state
---------- ---------------- ------------------- -------------------
0 KERNEL 0 interruptible-sleep
331 dhcpd 108920832 interruptible-sleep
336 vmmmgr 334442496 interruptible-sleep
554 neo 398274560 interruptible-sleep
1034 ae 153690112 interruptible-sleep
1214 eventmgr 514793472 interruptible-sleep
2541 bootmgr 292020224 interruptible-sleep
4390 snoopy 28499968 interruptible-sleep
5832 scripthandler 254308352 interruptible-sleep
19204 dbgr 648941568 interruptible-sleep
21863 nginx 4312199168 interruptible-sleep
32192 appliancedirector 136732672 interruptible-sleep
32197 sshd 1228800 interruptible-sleep
32202 perfwatch 19345408 interruptible-sleep
32203 observer 724484096 interruptible-sleep
32205 lldpad 1200128 interruptible-sleep
32209 topomgr 280576000 interruptible-sleep
32210 xinetd 99258368 interruptible-sleep
32213 policymgr 673251328 interruptible-sleep
32215 reader 258940928 interruptible-sleep
32216 logwatch 266596352 interruptible-sleep
32218 idmgr 246824960 interruptible-sleep
```

```
32416 keyhole 15233024 interruptible-sleep
admin@apic1:processes>
```

Each of the processes running on the APIC writes to a log file on the system. These log files can be bundled as part of the APIC techsupport file but can also be observed through SSH shell access in /var/log/dme/log. For example, the Policy Manager process log output is written into /var/log/dme/log/svc_ifc_policymgr.bin.log.

The following is a brief description of the processes running on the system. This can help in understanding which log files to reference when troubleshooting a particular process or understand the impact to the system if a process crashed:

Process	Function
KERNEL	Linux kernel
dhcpd	DHCP process running for APIC to assign infra addresses
vmmmgr	Handles process between APIC and Hypervisors
neo	Shell CLI Interpreter
ae	Handles the state and inventory of local APIC appliance
eventmgr	Handles all events and faults on the system
bootmgr	Controls boot and firmware updates on fabric nodes
snoopy	Shell CLI help, tab command completion
scripthandler	Handles the L4-L7 device scripts and communication
dbgr	Generates core files when process crashes
nginx	Web service handling GUI and REST API access
appliancedirector	Handles formation and control of APIC cluster
sshd	Enabled SSH access into the APIC
perfwatch	Monitors Linux cgroup resource usage
observer	Monitors the fabric system and data handling of state, stats, health
lldpad	LLDP Agent
topomgr	Maintains fabric topology and inventory

Cisco APIC Troubleshooting Operations

Shutting Down the Cisco APIC System

This procedure shuts down the Cisco Application Policy Infrastructure Controller (APIC) system. After you shut down the system, you will relocate the entire fabric and power it up, then update the time zone and/or NTP servers accordingly.

Before you begin

Ensure cluster health is fully fit.

Procedure

Step 1 On the menu bar, choose **System** > **Controllers**.
Step 2 In the Navigation pane, choose **Controllers** > ***apic_name***.
Step 3 Right-click the Cisco APIC and choose **Shutdown**.
Step 4 Relocate the Cisco APIC, then power it up.
Step 5 Confirm that the cluster has fully converged.
Step 6 Repeat this procedure for the next Cisco APIC.

Shutting Down a Cisco APIC Using the GUI

This procedure shuts down a Cisco Application Policy Infrastructure Controller (APIC). This procedure shuts down only one Cisco APIC, not the entire Cisco APIC system itself. Following this procedure causes the controller to shut down immediately. Use caution in performing a shutdown because the only way to bring the controller back up is to do so from the actual machine. If you need to access the machine, see Controlling the LED Locator Using the GUI, on page 98.

Note If possible, move Cisco APICs one at a time. As long as there are at least two Cisco APICs in the cluster online, there is read/write access. If you need to relocate more than one Cisco APIC at a time, this results in one or no remaining controllers online, and the fabric will go into a read-only mode when they are shut down. During this time, there can be no policy changes including endpoint moves (which includes virtual machine movement).

Procedure

Step 1 On the menu bar, choose **System** > **Controllers**.
Step 2 In the Navigation pane, choose **Controllers** > ***apic_name***.

Step 3 Right-click the Cisco APIC and choose **Shutdown**.

Step 4 Relocate the Cisco APIC, then power it up.

Step 5 Confirm that the cluster has fully converged.

Using the APIC Reload Option Using the GUI

This procedure reloads the Cisco Application Policy Infrastructure Controller (APIC), not the entire Cisco APIC system, using the GUI.

Procedure

Step 1 On the menu bar, choose **System** > **Controllers**.

Step 2 In the Navigation pane, choose **Controllers** > ***apic_name***.

Step 3 Right-click the Cisco APIC and choose **Reload**.

Controlling the LED Locator Using the GUI

This procedure turns on or off the LED locator for the Cisco Application Policy Infrastructure Controller (APIC) using the GUI.

Procedure

Step 1 On the menu bar, choose **System** > **Controllers**.

Step 2 In the Navigation pane, choose **Controllers** > ***apic_name***.

Step 3 Right-click the Cisco APIC and choose **Turn On Locator LED** or **Turn On Locator LED** as appropriate.

Powering Down the Fabric Using the GUI

This procedure powers down the fabric for power maintenance using the Cisco Application Policy Infrastructure Controller (APIC) GUI and Cisco Integrated Management Controller (IMC) GUI.

Procedure

Step 1 Shut down all Cisco APICs except the last one using the Cisco APIC GUI.

a) Log in to a Cisco APIC.

b) On the menu bar, choose **System** > **Controllers**.

c) Choose one of the Cisco APICs. In the Navigation pane, choose **Controllers** > ***apic_name***.

d) Right-click the Cisco APIC and choose **Shutdown**.
e) Repeat steps 1.c, on page 98 and 1.d, on page 99 for all other Cisco APICs, except the last one.

Step 2 Shut down the last Cisco APIC using the Cisco IMC GUI.

a) Log in to the Cisco IMC GUI of the last Cisco APIC.
b) In the Navigation pane, click the **Chassis** menu.
c) In the **Chassis** menu, choose **Summary**.
d) In the toolbar above the work pane, choose **Host Power** > **Shut Down**.

You must shut down the last Cisco APIC using the Cisco IMC GUI because this server will be in the read-only mode and will not process a shutdown request through Cisco APIC GUI.

Step 3 After you shut down all Cisco APICs, power down the switches by turning off their power supply.

Powering Up the Fabric Using the GUI

This procedure powers up the fabric using the Cisco Integrated Management Controller (IMC) GUI.

Procedure

Step 1 Power on Cisco APICs using the Cisco IMC GUI.

a) Log in to the Cisco IMC GUI of a Cisco APIC.
b) In the Navigation pane, click the **Chassis** menu.
c) In the **Chassis** menu, choose **Summary**.
d) In the toolbar above the work pane, choose **Host Power** > **Power On**.
e) Repeat these substeps for all Cisco APICs.

Step 2 Power on the leaf switches that are connected directly to the Cisco APIC.

Step 3 Power on the spine switches approximately a minute after you powered on the leaf switches.

Step 4 Power on the remaining leaf switches in the fabric.

The Cisco APICs discover the leaf switches directly connected to them through LLDP, followed by discovering the spine switches and remaining leaf switches. The discovery is automatic because the Cisco APICs retain the configurations and fabric memberships across reloads and shutdowns. The cluster comes up in the fully fit state after the Cisco APICs discover all leaf switches connected to them and discover the spine switches.

Switch Operations

Manually Removing Disabled Interfaces and Decommissioned Switches from the GUI

In a scenario where a fabric port is shut down then brought back up, it is possible that the port entry will remain disabled in the GUI. If this occurs, no operations can be performed on the port. To resolve this, you must manually remove the port from the GUI.

Procedure

Step 1 From the **Fabric** tab, click **Inventory**.

Step 2 In the **Navigation** pane, click **Disabled Interfaces and Decommissioned Switches**.
The list of disabled interfaces and decommissioned switches appears in a summary table in the **Work** pane.

Step 3 From the **Work** pane, right-click on the interface or switch that you want to remove and choose **Delete**.

Decommissioning and Recommissioning Switches

To decommission and recommission all the nodes in a pod, perform this procedure. One use case for this is to change the node IDs to a more logical, scalable numbering convention.

Procedure

Step 1 Decommission the nodes in the pod by following these steps for each one:

a) Navigate to **Fabric** > **Inventory** and expand the **Pod**.
b) Select the switch, right-click on it, and choose **Remove from Controller**.
c) Confirm the action and click **OK**.

The process takes about 10 minutes. The node is automatically wiped and reloaded. In addition, the node configuration is removed from the controller.

d) If a decommissioned node had the port profile feature deployed on it, some port configurations are not removed with the rest of the configuration. It is necessary to manually delete the configurations after the decommission for the ports to return to the default state. To do this, log on to the switch, run the **setup-clean-config.sh** script, and wait for it to run. Then, enter the **reload** command.

Step 2 When all the switches have been decommissioned from the pod, verify they are all physically connected and booted in the desired configuration.

Step 3 Perform the following actions to recommission each node.

Note
Before recommissioning a node with a port profile configuration as a new node, you must run the **setup-clean-config.sh** script to restore the port configuration to the default settings.

a) Navigate to **Fabric** > **Inventory**, expand **Quick Start**, and click **Node or Pod Setup**.
b) Click **Setup Node**.
c) In the **Pod ID** field, choose the pod ID.
d) Click the + to open the **Nodes** table.
e) Enter the node ID, serial number, Switch name, TEP Pool ID, and Role (**leaf** or **spine**) for the switch.
f) Click **Update**.

Step 4 Verify the nodes are all set up by navigating to **Fabric** > **Inventory** > **Fabric Membership**.

What to do next

If the pod is one of the pods in a multipod topology, reconfigure multipod for this pod and the nodes. For more information, see *Multipod* in the *Cisco APIC Layer 3 Networking Configuration Guide*.

Clean Reloading a Cisco ACI-Mode Switch

This procedure performs a clean reload of Cisco ACI-mode switches. A clean reload erases the configuration on the switch. After the switch boots up, the switch gets its configuration from the Cisco Application Policy Infrastructure Controller (APIC).

Procedure

Step 1 Log into a switch that you want to clean reload.

Step 2 Run the **setup-clean-config.sh** script with the **-k** argument.

Example:

```
switch1# setup-clean-config.sh -k
```

Step 3 Reload the switch.

Example:

```
switch1# reload
```

Recovering a Disconnected Leaf

If all fabric interfaces on a leaf are disabled (interfaces connecting a leaf to the spine) due to a configuration pushed to the leaf, connectivity to the leaf is lost forever and the leaf becomes inactive in the fabric. Trying to push a configuration to the leaf does not work because connectivity has been lost. This chapter describes how to recover a disconnected leaf.

Recovering a Disconnected Leaf Using the NX-OS-Style CLI

This this procedure enables fabric interfaces using the Cisco Application Policy Infrastructure Controller (APIC) NX-OS-style CLI. Use this procedure if you do not have any external tools from which you can make REST API calls.

Note This procedure assumes that 1/31 is one of the leaf switch ports connecting to the spine switch.

Procedure

Step 1 Using Cisco APIC NX-OS-style CLI, remove the block list policy.

Example:

```
apic1# podId='1'
apic1# nodeId='103'
apic1# interface='eth1/31'
apic1# icurl -sX POST 'http://127.0.0.1:7777/api/mo/.json' -d '{"fabricRsOosPath":{"attributes":

{"dn":"uni/fabric/outofsvc/rsoosPath-[topology/pod-'$podId'/paths-'$nodeId'/pathep-['$interface']]","status":"deleted"}}}'
```

Step 2 Using the CLI of a leaf or spine switch, set the port in service to bring up the port on the leaf switch.

Example:

```
switch1# podId='1'
switch1# nodeId='103'
switch1# interface='eth1/31'
switch1# icurl -X POST
'http://127.0.0.1:7777/api/node/mo/topology/pod-'$podId'/node-'$nodeId'/sys/action.json'
  -d
'{"actionLSubj":{"attributes":{"oDn":"sys/phys-['$interface']"},"children":[{"l1EthIfSetInServiceLTask":

  {"attributes":{"adminSt":"start"}}}]}}'
```

Recovering a Disconnected Leaf Using the REST API

To recover a disconnected leaf switch, you must enable at least one of the fabric interfaces using this procedure. You can enable the remaining interfaces using the GUI, REST API, or CLI.

To enable the first interface, post a policy using the REST API to delete the policy posted and bring the fabric ports Out-of-Service. You can post a policy to the leaf switch to bring the port that is Out-of-Service to In-Service as follows:

Note This procedure assumes that 1/49 is one of the leaf switch ports connecting to the spine switch.

Procedure

Step 1 Clear the block list policy from the Cisco APIC using the REST API.

Example:

```
$APIC_Address/api/policymgr/mo/.xml
<polUni>
```

```
    <fabricInst>
        <fabricOOServicePol>
            <fabricRsOosPath tDn="topology/pod-1/paths-$LEAF_Id/pathep-[eth1/49]" lc="blacklist"
status ="deleted"/>
        </fabricOOServicePol>
    </fabricInst>
</polUni>
```

Step 2 Post a local task to the node itself to bring up the interfaces you want using **l1EthIfSetInServiceLTask**.

Example:

```
$LEAF_Address/api/node/mo/topology/pod-1/node-$LEAF_Id/sys/action.xml
<actionLSubj oDn="sys/phys-[eth1/49]">
    <l1EthIfSetInServiceLTask adminSt='start'/>
</actionLSubj>
```

Performing a Rebuild of the Fabric

Rebuilding the Fabric

Caution This procedure is extremely disruptive. It eliminates the existing fabric and recreates a new one.

This procedure allows you to rebuild (reinitialize) your fabric, which you may need to do for any of the following reasons:

- To change the TEP IPs
- To change the Infra VLAN
- To change the fabric name
- To perform TAC troubleshooting tasks

Deleting the APICs erases the configuration on them and brings them up in the startup script. Performing this on the APICs can be done in any order, but ensure that you perform the procedure on all of them (every leaf and spine in the fabric).

Before you begin

Ensure that the following is in place:

- Regularly scheduled backups of the configuration
- Console access to the leaves and spines
- A configured and reachable CIMC, which is necessary for KVM console access
- No Java issues

Procedure

Step 1 If you would like to retain your current configuration, you can perform a configuration export. For more information, see the *Cisco ACI Configuration Files: Import and Export* document: https://www.cisco.com/c/en/us/support/cloud-systems-management/application-policy-infrastructure-controller-apic/tsd-products-support-series-home.html

Step 2 Erase the configuration on the APICs by connecting to the KVM console and entering the following commands:

a) **>acidiag touch clean**
b) **>acidiag touch setup**
c) **>acidiag reboot**

Ensure that each node boots up in fabric discovery mode and is not part of the previously configured fabric.

Note
The **acidiag touch** command alone is not useful for this procedure, because it does not bring the APIC up in the startup script.

Caution
It is extremely important that you ensure that all previous fabric configurations have been removed. If any previous fabric configuration exists on even a single node, the fabric cannot be rebuilt.

Step 3 When all previous configurations have been removed, run the startup script for all APICs. At this point, you can change any of the above values, TEP, TEP Vlan, and/or Fabric Name. Ensure that these are consistent across all APICs. For more information, refer to the *Cisco APIC Getting Started Guide*: https://www.cisco.com/c/en/us/support/cloud-systems-management/application-policy-infrastructure-controller-apic/tsd-products-support-series-home.html.

Step 4 To clean reboot the fabric nodes, log in to each fabric node and execute the following:

a) **>setup-clean-config.sh**
b) **>reload**

Step 5 Log in to apic1 and perform a configuration import. For more information, see the *Cisco ACI Configuration Files: Import and Export* document: https://www.cisco.com/c/en/us/support/cloud-systems-management/application-policy-infrastructure-controller-apic/tsd-products-support-series-home.html.

Step 6 Wait for a few minutes as the fabric now uses the previous fabric registration policies to rebuild the fabric over the nodes. (Depending on the size of the fabric, this step may take awhile.)

Troubleshooting a Loopback Failure

Identifying a Failed Line Card

This section expains how to identify a failed line card when getting a loopback failure.

Before you begin

You should have created a On-Demand TechSupport policy for the fabric node. If you have not already created an On-Demand TechSupport policy, see the “Sending an On-Demand Tech Support File Using the GUI” section in the *Cisco APIC Basic Configuration Guide*.

Procedure

Step 1 Collect the Logs Location file of the On-Demand TechSupport policy for the fabric node. To initiate the collection:

a) In the menu bar, click **Admin**.
b) In the submenu bar, click **Import/Export**.
c) In the **Navigation** pane, expand **Export Policies** and right-click the On-Demand TechSupport policy for the fabric node.
A list of options appears.
d) Choose **Collect Tech Supports**.
The **Collect Tech Supports** dialog box appears.
e) In the **Collect Tech Supports** dialog box, click **Yes** to begin collecting tech support information.

Step 2 Download the the Logs Location file of the On-Demand TechSupport policy for the fabric node. To download the Logs Location file:

a) From the On-Demand TechSupport policy window in the **Work** pane, click the **Operational** tab.
A summary table appears in the On-Demand TechSupport policy window with several columns, including the **Logs Location** column.
b) Click the URL in the **Logs Location** column.

Step 3 Inside the Logs Location file, go to the /var/sysmgr/tmp_logs/ directory and unzip the svc_ifc_techsup_nxos.tar file.

```
-bash-4.1$ tar xopf svc_ifc_techsup_nxos.tar
```

The show_tech_info directory is created.

Step 4 Run **zgrep "fclc-conn failed" show-tech-sup-output.gz | less**.

```
-bash-4.1$ zgrep "fclc-conn failed" show-tech-sup-output.gz | less
[103] diag_port_lb_fail_module: Bringing down the module 25 for Loopback test failed. Packets possibly
 lost on the switch SPINE or LC fabric (fclc-conn failed)
[103] diag_port_lb_fail_module: Bringing down the module 24 for Loopback test failed. Packets possibly
 lost on the switch SPINE or LC fabric (fclc-conn failed)
```

Note
The **fclc-conn failed** message indicates a failed line card.

Step 5 Power cycle the currently failed fabric cards and ensure the fabric cards come online.

Step 6 If the fabric cards fail to come online, or after the fabric cards go offline again, immediately collect the diag_port_lb.log file and send the file to the TAC team. The diag_port_lb.log file is located in the /var/sysmgr/tmp_logs/ directory of the Logs Location file.

Removing Unwanted _ui_ Objects

Caution Changes made through the APIC Basic GUI can be seen, but cannot be modified in the Advanced GUI, and changes made in the Advanced GUI cannot be rendered in the Basic GUI. The Basic GUI is kept synchronized with the NX-OS style CLI, so that if you make a change from the NX-OS style CLI, these changes are rendered in the Basic GUI, and changes made in the Basic GUI are rendered in the NX-OS style CLI, but the same synchronization does not occur between the Advanced GUI and the NX-OS style CLI. See the following examples:

- Do not mix Basic and Advanced GUI modes. If you apply an interface policy to two ports using Advanced mode and then change the settings of one port using Basic mode, your changes might be applied to both ports.
- Do not mix the Advanced GUI and the CLI, when doing per-interface configuration on APIC. Configurations performed in the GUI, may only partially work in the NX-OS CLI.

  For example, if you configure a switch port in the GUI at **Tenants** > ***tenant-name*** > **Application Profiles** > ***application-profile-name*** > **Application EPGs** > ***EPG-name*** > **Static Ports** > **Deploy Static EPG on PC, VPC, or Interface**

  Then you use the show running-config command in the NX-OS style CLI, you receive output such as:

  ```
  leaf 102
  interface ethernet 1/15
  switchport trunk allowed vlan 201 tenant t1 application ap1 epg ep1
   exit
   exit
  ```

  If you use these commands to configure a static port in the NX-OS style CLI, the following error occurs:

  ```
  apic1(config)# leaf 102
  apic1(config-leaf)# interface ethernet 1/15
  apic1(config-leaf-if)# switchport trunk allowed vlan 201 tenant t1 application ap1 epg
   ep1
  No vlan-domain associated to node 102 interface ethernet1/15 encap vlan-201
  ```

  This occurs because the CLI has validations that are not performed by the APIC GUI. For the commands from the show running-config command to function in the NX-OS CLI, a vlan-domain must have been previously configured. The order of configuration is not enforced in the GUI.
- Do not make changes with the Basic GUI or the NX-OS CLI before using the Advanced GUI. This may also inadvertantly cause objects to be created (with names prepended with `_ui_`) which cannot be changed or deleted in the Advanced GUI.

If you make changes with the Basic GUI or the NX-OS CLI before using the Advanced GUI, this may inadvertently cause objects to be created (with names prepended with `_ui_`) which cannot be changed or deleted in the Advanced GUI.

For the steps to remove such objects, see Removing Unwanted _ui_ Objects Using the REST API, on page 107.

Removing Unwanted _ui_ Objects Using the REST API

If you make changes with the Cisco NX-OS-Style CLI before using the Cisco APIC GUI, and objects appear in the Cisco APIC GUI (with names prepended with `_ui_`), these objects can be removed by performing a REST API request to the API, containing the following:

- The Class name, for example **infraAccPortGrp**
- The Dn attribute, for example **dn="uni/infra/funcprof/accportgrp-__ui_l101_eth1--31"**
- The Status attribute set to **status="deleted"**

Perform the POST to the API with the following steps:

Procedure

Step 1 Log on to a user account with write access to the object to be removed.

Step 2 Send a POST to the API such as the following example:

```
POST https://192.168.20.123/api/mo/uni.xml
Payload:<infraAccPortGrp dn="uni/infra/funcprof/accportgrp-__ui_l101_eth1--31" status="deleted"/>
```

Cisco APIC SSD Replacement

Use this procedure to replace the Solid-State Drive (SSD) in Cisco APIC.

Note This procedure should only be performed when there is at least one APIC with a healthy SSD in the cluster, that is fully fit. If all the APIC controllers in the cluster have SSDs that have failed, open a case with the Cisco Technical Assistance Center (TAC).

Replacing the Solid-State Drive in Cisco APIC

Before you begin

- If your Cisco IMC release is earlier than 2.0(9c), you must upgrade the Cisco IMC software before replacing the solid-state drive (SSD). Refer to the release notes of the target Cisco IMC release to determine the recommended upgrade path from your currrent release to the target release. Follow the instructions in the current version of the *Cisco Host Upgrade Utility (HUU) User Guide* at this link to perform the upgrade.
- In the Cisco IMC BIOS, verify that the Trusted Platform Module (TPM) state is set to "Enabled." Using the KVM console to access the BIOS settings, you can view and configure the TPM state under **Advanced > Trusted Computing > TPM State**.

Note APIC will fail to boot if the TPM state is "Disabled."

- Obtain an APIC `.iso` image from the Cisco Software Download site.

Note The release version of the APIC `.iso` image must be the same version as the other APIC controllers in the cluster.

Procedure

Step 1 From another APIC in the cluster, decommission the APIC whose SSD is to be replaced.

a) On the menu bar, choose **System** > **Controllers**.
b) In the **Navigation** pane, expand **Controllers** > **apic_controller_name** > **Cluster as Seen by Node**. For the **apic_controller_name**, specify an APIC controller that is not being decommissioned.
c) In the **Work** pane, verify that the **Health State** in the **Active Controllers** summary table indicates the cluster is **Fully Fit** before continuing.
d) In the same **Work** pane, select the controller to be decommissioned and click **Actions** > **Decommission**.
e) Click **Yes**.
The decommissioned controller displays **Unregistered** in the **Operational State** column. The controller is then taken out of service and is no longer visible in the **Work** pane.

Step 2 Physically remove the old SSD, if any, and add the new SSD.

Step 3 In the Cisco IMC, create a RAID volume using the newly installed SSD.

Refer to the *Cisco UCS C-Series Integrated Management Controller GUI Configuration Guide* for your Cisco IMC release. In the "Managing Storage Adapters" chapter, follow the instructions in the procedure "Creating Virtual Drive from Unused Physical Drives" to create and initialize a RAID 0 virtual drive.

Step 4 In the Cisco IMC, install the APIC image using the virtual media. In this step, the SSD is partitioned and the APIC software is installed on the HDD.

Note
For a fresh install of Cisco APIC Release 4.x or later, see the *Cisco APIC Installation, Upgrade, and Downgrade Guide*.

a) Mount the APIC `.iso` image using the Cisco IMC vMedia functionality.
b) Boot or power cycle the APIC controller.
c) During the boot process press **F6** to select the **Cisco vKVM-Mapped vDVD** as the one-time boot device. You may be required to enter the BIOS password. The default password is 'password'.
d) During the initial bringup, a configuration script runs. Follow the onscreen instructions to configure the initial settings of the APIC software.
e) After the installation is completed, un-map the virtual media mount.

Step 5 From an APIC in the cluster, commission the decommissioned APIC.

a) Select any other APIC that is part of the cluster. From the menu bar, choose **System** > **Controllers**.
b) In the **Navigation** pane, expand **Controllers** > **apic_controller_name** > **Cluster as Seen by Node**. For the **apic_controller_name**, specify any active controller that is part of the cluster.

c) From the **Work** pane, click the decommissioned controller that displays **Unregistered** in the **Operational State** column.
d) From the **Work** pane, click **Actions** > **Commission**.
e) In the **Confirmation** dialog box, click **Yes**.

The commissioned controller displays the Health state as **Fully-fit** and the operational state as **Available**. The controller should now be visible in the **Work** pane.

Viewing CRC Error Counters

Viewing CRC and Stomped CRC Error Counters

Beginning in Cisco APIC Release 4.2(3), CRC errors are split into two categories: CRC errors and stomped CRC errors. CRC errors are corrupted frames that were dropped locally and stomped CRC errors are corrupted frames that were cut-through switched. This differentiation can make it easier to identify the actual interface impacted by CRC errors and troubleshoot physical layer issues within the fabric.

This section demonstrates how to view the CRC and stomped CRC errors.

Viewing CRC Errors Using the GUI

This section demonstrates how to view CRC and stomped CRC error counters using the GUI.

SUMMARY STEPS

1. On the menu bar, choose **Fabric > Inventory**.
2. In the **Navigation** pane, click to expand a pod.
3. Click to expand **Interfaces**.
4. Click to choose an interface.
5. In the **Work** pane, click the **Error Counters** tab.

DETAILED STEPS

Procedure

Step 1 On the menu bar, choose **Fabric > Inventory**.

Step 2 In the **Navigation** pane, click to expand a pod.

Step 3 Click to expand **Interfaces**.
A list of interfaces appear in the **Navigation** pane.

Step 4 Click to choose an interface.
A list of tabs appear in the **Work** pane.

Step 5 In the **Work** pane, click the **Error Counters** tab.

A list of error categories appears including **CRC Errors (FCS Errors)** and **Stomped CRC Errors (packets)**.

Viewing CRC Errors Using the CLI

This section demonstrates how to view CRC and stomped CRC error counters using the CLI

Procedure

To view CRC and stomped CRC errors:

Example:

```
Switch# show interface ethernet 1/1
Ethernet1/1 is up
admin state is up, Dedicated Interface
  Belongs to po4
  Hardware: 100/1000/10000/25000/auto Ethernet, address: 00a6.cab6.bda5 (bia 00a6.cab6.bda5)
  MTU 9000 bytes, BW 10000000 Kbit, DLY 1 usec
  reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation ARPA, medium is broadcast
  Port mode is trunk
  full-duplex, 10 Gb/s, media type is 10G
  FEC (forward-error-correction) : disable-fec
^[[B  Beacon is turned off
  Auto-Negotiation is turned on
  Input flow-control is off, output flow-control is off
  Auto-mdix is turned off
  Rate mode is dedicated
  Switchport monitor is off
  EtherType is 0x8100
  EEE (efficient-ethernet) : n/a
  Last link flapped 3d02h
  Last clearing of "show interface" counters never
  1 interface resets
  30 seconds input rate 0 bits/sec, 0 packets/sec
  30 seconds output rate 4992 bits/sec, 8 packets/sec
  Load-Interval #2: 5 minute (300 seconds)
    input rate 0 bps, 0 pps; output rate 4536 bps, 8 pps
  RX
    0 unicast packets  200563 multicast packets  0 broadcast packets
    200563 input packets  27949761 bytes
    0 jumbo packets  0 storm suppression bytes
    0 runts  0 giants  0 CRC  0 Stomped CRC  0 no buffer
    0 input error  0 short frame  0 overrun   0 underrun  0 ignored
    0 watchdog  0 bad etype drop  0 bad proto drop  0 if down drop
    0 input with dribble  0 input discard
    0 input buffer drop  0 input total drop
    0 Rx pause
  TX
    0 unicast packets  2156812 multicast packets  0 broadcast packets
    2156812 output packets  151413837 bytes
    0 jumbo packets
    0 output error  0 collision  0 deferred  0 late collision
    0 lost carrier  0 no carrier  0 babble  0 output discard
    0 output buffer drops  0 output total drops
    0 Tx pause
```