PTP Webinar 2023 Follow-up

Introduction

Greetings M&E aficionados. We ran an M&E webinar entitled "Reliable PTP: Lessons from the Field" on 30th March 2023. The webinar recording can be found here if you missed it, or would like to re-watch, or share with others.

During the webinar, we ran a live demo of configuring PTP Boundary Clocks, and we took a lot of very interesting questions. This document aims to answer in more detail, as many of those questions as possible, as well as documenting the config steps undertaken in the demo.

This document contains many links - some require visitor registration / login, others may be limited to access by customer login. If you have problems accessing any links, or if you have other questions, please reach out to your Account Team for assistance, or at the Arista Community Forum.

Table of Contents

Webinar Questions Answered

Q: Will the recording be available for others who can't make this webinar?

A: Yes, the recording is available here. If there are any other questions you'd like to see answered, please send an email to your account team, to sales@arista.com, or at the Arista Community Forum.

Q: Is it common to have customers use g8275.2? What is the most common profile and why are they more popular than others?

A: The use of the g8275.2 PTP profile can vary depending on the specific industry and application. Generally speaking, this profile is used in applications where high-precision time synchronization is required, such as in power utilities or telecommunications networks, however the most common profile used in PTP is the "default" profile defined in the IEEE 1588 standard. The default profile provides basic functionality for time synchronization, and is widely used in many different types of networks. SMPTE ST2059-2 builds on the default profile to provide synchronization for SMPTE ST2110 infrastructure.

Q: Is unicast more common than multicast from deployment perspective? What are the trade-offs in terms of scale?

A: IEEE1588, PTPv2 is a very flexible protocol set, which allows for unicast, multicast, or mixed message transmission. Which transport is allowed is defined by the appropriate profile. For example, the SMPTE ST2059 profile used with SMPTE ST2110 for live media production recommends the use of multicast sync and announce messages, along with unicast delay-request and delay-response. In a typical all-Boundary-Clock ST2059 installation, since all the GrandMaster (GM) to Boundary Clock (BC), BC to BC, and BC to host connections are direct, whether unicast or multicast is used does not really affect scale. The use of boundary clocks offloads message scale from the GrandMasters as well. Unicast delay-request / delay-response can make a big difference in achievable scale in Transparent Clock (TC) PTP infrastructure.

The g8275.2 profile is used in the Telco world, and allows for PTP synchronisation over unicast networks with partial PTP support.

Q: Can Arista switches be a Boundary Clock for more than one domain?

A: Currently, a Boundary Clock supports one domain. A Transparent Clock can support multiple domains as we just pass multicast traffic and adjust the correction field.

Q: Since you are recommending the use of BC, what is the holdover performance in ppm or time period of the oscillator in your recommended platforms?

A: We strongly recommend that the switch holdover functionality is not relied on to drive the system timing for any periods except the normal BMCA change over events. Typical systems have at least two GrandMasters for resilience, and if more resilience is required, we recommend designing in more GrandMasters. Good GrandMasters add many features that are not efficient to do in datacenter grade switches: controlled drift back to external reference after recovery, generation of application specific timing signals (Black and Burst, Audio Word Clock, PPS), etc. The specific ppm will vary from platform to platform, but in general we use OCXOs to give improved performance.

Q: What is the layer 1 connection of a GM usually seen as? Fibre, copper, other?

A: We see GrandMasters support both Fiber and Copper. Higher interface speeds can be argued to improve performance, as the intrinsic link jitter is lower. We typically see GrandMaster interface speeds as 1G or 10G.

Q: I work in finance, using 7150's today for PTP distribution.

A: Yes, the 7150 was a great switch for PTP distribution, and many are still hard at work today. The high precision "CL" versions allowed for longer holdover times, in the case of system GrandMaster failure. Most current Arista switches will provide similar performance (with the exception of the precision clock for holdover), 7050X3, 7280R2/3, 7020R, 7800R3, 7500R/R2/R3 series for example.

Q: I see 7050SX3-48C8 in your presentation. Which arista switches are best suited for PTP (other than 7050SX3-48C8). I want optimal(low) PTP jitter

A: Almost all Arista switches that support PTP have excellent PTP jitter performance. The only switch I would call out as not being quite as good as all the others is the 720XP. The performance here is very good, but not quite as good as others.

Q: Does Arista have special interfaces to connect to the GrandMaster or through ethernet?

A: We do not. Any dataplane interface on the switch is suitable.

Q: Do you see more multicast or unicast PTP deployments? We have been seeing more unicast asks for PTP

A: In a directly connected Boundary Clock environment, the transport doesn't really matter. The transport method may be defined by the PTP profile, which is in turn defined by the use-case. However, in Media and Entertainment, we typically see PTP deployed as multicast. In environments where not all devices in the path support PTP, the g8275.2 profile is generally considered well suited.

Q: What kind of sync/announce interval rates do you recommend ?

A: Message rates will very much depend on your application, and the desired accuracy. The PTP profile in use may well define the message rate ranges, and their defaults. For example, in the Live Production environment, the industry has settled on a set of rates described in AES-R16-2016, which provide a good compromise between accuracy and lock speed, without overloading lower capacity devices:

- 1 Announce message per second
- 8 Sync messages per second
- 8 Delay-Request messages per second

Q: What would be the impact on accuracy during speed translation, say from GrandMaster on 10G towards Spine (as BC) distributing time on a 400G interface?

A: On all current PTP supporting Arista switches, interface speed changes are accounted for, and there is no appreciable impact on accuracy.

Q: What if we want to run both PTPv1 and PTPv2 in the same environment?

A: In a Transparent Clock environment, then PTPv1 and PTPv2 can operate quite happily together, but the PTPv1 messages will be treated as normal multicast, with no PTP awareness. In Boundary Clock mode, the switch will need to be configured to "forward PTPv1", this enables PTPv1 and PTPv2 to pass through the same interfaces, assuming the interfaces in question have PTP enabled. The TOI for "forward PTPv1" can be found here.

Q: For the link from spine to leaf, wouldn't it be better to configure ptp-role master, so the leaf wouldn't take over as the GM by accident if someone introduces a new switch ?

A: For most installations, we recommend not enabling "ptp role master" on switch-to-switch links, as this maximises the resilience that is built into the system. The switch-to-switch links will only result in leafs driving spines in extreme and highly failed states. Automation techniques can also be applied to ensure that the leaf switches all have their P1/P2 values following a tiered approach to prevent this from occurring.

Q: What are Red and Blue used for? Do they have different traffic passing through?

A: Yes, in the Media Live Production industry, a resilience architecture described by SMPTE ST2022-7 provides highly reliable transmission of multicast traffic, using a technique called "hitless merge", where each multicast flow is delivered by both a Red and Blue network.

Q: For Dante (PTPv1), are there any use cases/customer studies available on the website

A: We have hundreds of customers running Dante successfully, however, I don't think we have any case-studies that cover Dante specifically. If Dante equipment can't be configured to use PTPv2, or if this is not desirable, then we recommend the use of "ptp forwardv1", which allows PTPv1 and PTPv2 to be present in the same PTPv2 aware fabric. More details can be found in this TOI

Q: Do all Arista platforms support PTP, the Campus portfolio specifically for campus-wide audio/video?

A: Almost all Arista switches support PTPv2. You can do a quick check on the "Feature Matrix"; selecting the switch range and "Management Features"

The 7280R3, 7050X3 or 720XP ranges would be a good place to start, depending on your requirements. For more help, contact your Account Team, or email sales@arista.com for support.

Q: Is it possible to sync PTP across multiple sites via a private media WAN?

A: It would very much depend what accuracy you need, and how the private media WAN was structured. If this is a private, PTP aware network, then this is very likely to be possible with PTP transmission in the underlay. If this is a VPN, with PTP carried in the overlay, then end accuracy will be significantly impacted by the lack of PTP aware processing at each hop. An alternative would be to install GrandMasters at the various sites and ensure they are locked to GPS. Depending on your needs, this could be a good use-case for unicast based PTP profiles like g8275.2

Q: Will PTP work on LACP connections?

A: Yes, PTP on LACP between Arista switches will just work.

PTP running over LACP connections from a host to Arista MLAG switch pair will need MPASS to be configured. More details can be found here.

Q: What are the limits scale-wise on PTP boundary clock interfaces on leaf switches facing environments that do not support PTP Boundary Clock. For example, if an Arista switch interface is facing a downstream hypervisor running numerous VMs consuming PTP. The vSwitch is not likely to support PTP. A second example might be an Arista PTP boundary clock upstream of a blade chassis/some other fixed switch implementation that does not support PTP Boundary Clock.

A: The answer here is very much "it depends". There are many factors at play - the Arista switch platform, CPU capability, other CPU tasks being performed etc. This would be a design question best brought up with your Arista Account team.

Q: Could I use an Evertz as GrandMaster1 and a Meinberg as GrandMaster2, or do they need to be the same vendor?

A: As long as your GrandMasters are standards (and profile) compliant, you can use any vendor, or mix of vendors that you prefer. We have plenty of customers doing exactly this right now.

Q: Why are you using "speed forced" on the interface configuration to the GrandMaster?

A: Reducing any possibility of error is good practice for the highly available PTP service, and so we recommend specifying the interface speed explicitly.

Q: Is it normal to only have a single interface to GrandMaster from spine1 or are there situations where you'd want two interfaces from GM1 to spine1 and the other to spine2?

A: Some GrandMasters have multiple interfaces. It's important to understand how these interfaces work. In some cases, these interfaces are simply multiple ports presenting the same MAC / IP Address. If this is the case, this arrangement does not typically add any extra resilience. If however the multiple interfaces are individual GrandMasters (each with their own IP and MAC addresses), then clearly this adds more GrandMaster capability, and so more resilience.

Q: Didn't your GrandMaster just change to an Arista switch instead of either the Meinberg or the Tek?

A: During the demo, when we caused a BMCA event - for example adding a connection that allowed PTP traffic between the two PTP distribution Leafs, its likely that for a very short period, the Arista may have considered itself the best available clock. This is completely normal "BMCA Churn", and will also be seen during a real GrandMaster BMCA event.

Q: Can you PTP through MLAG'd switches?

A: Yes, PTP can be distributed through MLAG'd switches. Enabling this is the same as non-MLAG'd switches when these switches are attached to other Arista switches. Support for PTP Boundary Clock delivery to non-Arista devices, through MLAG'd architectures is also supported from EOS 4.26.2 onwards. More details can be found here.

Q: Does the programmable ClockID need to be the MAC address or can it be a "friendly" name?

A: The user can configure the ClockID to be anything they like, as long as it follows the appropriate MAC like format - the format is defined by the IEEE 1588 standard.

Q: Do you support other PTP profiles? Which ones?

A: Clock modes: Boundary (one-step and 2-step), Transparent (one-step and two-step, but recommend one-step always), e2e Transparent, 2-step p2p Transparent, gptp

Profiles: SMPTE/AES67 (Based on messaging rates), g8275.1, g8275.2 (check S/W and H/W dependencies). Note that g8275.2 will interoperate with g8265.1 configured on the client side.

Q: Is Arista compliant with PTPv2.1?

A: No, at the moment we do not support PTPv2.1 Customer demand has not yet been high enough. If you need PTPv2.1, please raise that with your account team, and they can make the request clear to the Product Management teams.

Q: What would determine the interval selection? Just specific application use cases?

A: Yes, the message intervals are selected to fit with the application needs - lockup time, accuracy etc.

Q: Are these configuration commands available in a doc? Would it be possible to get a copy of the example configs Gerard is working with for review?

A: Yes, these configuration steps are included in the next section of this document.

Q: If PTP multicast is enabled. What multicast address will it be used? Or is it just between boundary clock to the next device as the TTL is 1. (link local multicast ?)

A: For L2 multicast transport, the destination addresses is 01-1B-19-00-00-00 for all messages except p2p, which use 01-80-C2-00-00-0E

For IPv4 transport, the destination address is 224.0.1.129 for all messages except p2p, which use 224.0.0.107

For IPV6 transport, the destination address is FF0X:0:0:0:0:0:0:181 for all messages except p2p, which use FF02:0:0:0:0:0:0:6B

The UDP ports are 319 (event messages) and 320 (general messages). Boundary Clock message TTL can be increased from 1, in non direct BC <-> BC designs.

Q: Do you need an RP (Rendezvous Port) defined for PTP?

A: No RP is needed for direct Boundary Clock to Boundary Clock. However standard multicast config and support is needed for non-PTP aware and Transparent Clock implementations.

Q: Can an Arista switch belong to multiple PTP domains?

A: Currently a Boundary Clock can only belong to one PTP domain. A Transparent Clock can belong to multiple domains.

Q: Can you please elaborate more about priority1 and 2, and preferred values?

A: In the default profile BMCA the attributes used in the Announce Message to elect a GrandMaster are as follows: P1, Clock Class, Clock Accuracy, Clock Variance, P2, GMID. In systems with multiple GrandMasters we typically see them have the same P1 value (a low value perhaps 1 or 2), and then use P2 as the tiebreaker assuming that clock quality is consistent. Then we simply will increase the P1 values in a tiered approach so that the election process in the absence of GrandMasters is consistent. For example, the PTP distribution layer will have a P1 of 10 and P2 of 1 and 2. The spines will have a P1 of 20, and P2 of 1-N, and so on.

In the alternative BMCAs, it is possible that P1 is not used and is set to a constant value 128. G8275.2 is an example of a profile using such BMCA.

Q: In boundary clock mode, do you specify what VLAN needs to be the PTP VLAN or can all VLANs participate in boundary clock mode?

A: On a trunk interface, by default we send PTP over the native VLAN. Using the "ptp vlan<id>" command you can enable PTP on whatever VLANs you need, but by default we will not send PTP traffic tagged across all VLANs. Many PTP profiles allow operation on a routed port, where the traffic will be untagged.

Q: How to debug a rogue device is challenging the master clock or flip flopping?

A: The "show ptp int XXX counters" will help you here, this command will show that a rogue host is sending unexpected sync and/or announce messages. "Show ptp" will show the alternative GMs that we are locking to, and if enabled, logging and syslog will also provide pointers to the GMs we are seeing, and their interface locations. CloudVision can provide very useful timelines of GM activity.

Q:What about overlap with ntp and ptp on same switch: bad idea?

A: We typically see PTP and NTP configured on the same switch, but there is no direct relationship between the two - they are not synchronized together.

Q: Will the AQL for PTP Monitoring be built into CVP?

A: We're not sure if this is planned, but we do intend to publish the AQL once we've had a chance to document for customer use.

Q: If we configured an Arista as a Transparent Clock, how do you map different switch ports to different PTP domains?

A: In Transparent Clock mode, the switch is unaware of the PTP domains that are being used. It becomes the job of the GrandMasters and Slaves to only act on appropriately domained messages. If domains can be limited to geographic scopes, then standard multicast design can be used to limit the scope of multicast PTP delivery. Other segmentation techniques may be available, depending on the network design and requirements; VRFs, VLAN, Multicast boundaries, separated RPs etc.

Q: We find many production appliances that use PTP do not participate in the PTP management protocol. Is there a good way of monitoring these devices from a PTP point of view from the switch?

A: The switch can help to provide visibility of the number of Management Messages being sent and received on a port, but we do not look deeper into these messages to determine host health.

Q: Can I get a link to look up AQL protocol?

A: Details of the AQL Tool can be found in the TOI here.

Q: Does the PTP state "passive" derived from Spanning Tree blocking state of the port?

A: No. The PTP Topology is determined independently, by the passing of PTP Announce messages. A "passive" port will be instantiated if there are more than one valid paths from this switch back to a valid GM.

Q: Is "ptp clock-identity MAC-ADDR" the mac of the GM? Is that the same throughout the topology, even if it crosses boundary clocks?

A: The GMID (GrandMasterID) is constant throughout the network. Each switch also has a unique ClockID, this can be manually set, or automatically derived from the switch system MAC address. The GMID is useful to verify that everybody is locked to the same, valid GM. Local ClockID is useful to determine who the next hop up the Boundary Clock chain is.

Q: Can an Arista switch running in Boundary Clock be used as the GrandMaster in a system?

A: Yes, it could, but there are some reasons why this may not be suitable. Without external synchronisation, the PTP would be asynchronous to any external system, it would always be drifting relative to other clocks. The switch oscillator, while an OCXO on many platforms, is still going to have a lower stability than most external clock sources.

Q: Can you share more details of the automation options that were mentioned?

A: We highlighted "AVD" the Arista specific collection of Ansible roles and modules that can be used to automate the configuration and deployment of network infrastructure, encapsulating Arista best design practices (Arista Validated Designs) This collection includes support for ST2059 profile Boundary Clock configuration. More details on AVD can be found here.

Q: Can you share any details or links for MPASS? Excellent feature btw!

A: Yes, there is a TOI here, you'll need an Arista registration to access.

Q: What platforms are you getting this kind of jitter on?

A: The platforms in the webinar were 7020TR-48 (PTP distribution), 7280CR2A-30 (Spine), and 7280SR2-48YC6 (Leaf), running the most current version of EOS (4.29.3M at time of writing)

Q: What's the maximum number of hops you recommend between the end device and the GM?

A: The answer here will very much depend on the application, and PTP profile used. For Live Production ST2110, with ST2059-2 profile, we typically see between 1 and 4 switches running Boundary Clock between the GM and hosts, with no adverse effects. If you think you'd need significantly more in your application, I would suggest reaching out to your Arista Account Team for a chat.

Q: So to go back to the dante, if our network is a domain of say 20, and dante expects 0, the v1 setting uses domain id 0?

A: PTP-v2 and PTP-v1 domains do not need to be the same. Since the ptpv1 forwarding function relies on normal multicast forwarding, it is not aware of the domain, so running PTPv2 Boundary Clock for domain 20, and PTPv1 with domain 0 is perfectly fine.

Q: What is the best cli command to show the switch is in holdover?

A: You can run 'show ptp' and then see that the number of slave ports = 0. If there are no slave ports, the boundary clock is in holdover mode. Also, we will send syslog messages indicating that a boundary clock state has transitioned to Holdover mode.

Q: How can a leaf detect when the Arista switch is in holdover mode?

A: There is a configurable field called the Announce Timeout. The default is 3. This means that if the switch does not see an Announce Message for 3 Announce Intervals, the switch will enter into holdover mode. This holdover state will be temporary, assuming there is another GrandMaster in the network that can take over. However, if there are no more valid GrandMasters, one of the switches in Boundary Clock mode will ultimately become the new GrandMaster. Leafs don't need to know if they, or the switches above them are in holdover mode, BUT telemetry can easily determine this state, by looking at the "show ptp" output, which will identify the ClockID of the current GrandMaster.

Q: Does PTP deployment need multicast or can run over unicast ? Looking for deployment best practices in VxLAN EVPN environment with multiple VTEPs.

A: Arista does not support PTP aware VXLAN, so the best practice right now would be to run PTP in the underlay, and then distribute to hosts at VTEPs. If this model does not fit your use case, I suggest contacting your Arista Account Team to discuss further. Unicast based PTP profiles like g8275.2 could probably be run over VXLAN tunnels, but the lack of PTP aware processing in the tunnel will impact performance.

Q: When a GrandMaster reappears in holdover mode, how does the switches resync to that GM? Hard cut or via clock jitter?

A: When the Boundary Clock network loses all GrandMasters, one of the remaining switches will become elected as the new GrandMaster, and will enter holdover mode, and continue to serve time to the whole system. If the original (or new) GrandMaster becomes available again, the switch will drift if under 1us Offset-From-Master, or will snap if over 1us.

Q: Is holdover configurable or is 8 hours just a standard that can't be changed?

A: Holdover can be adjusted up to about 48 hours (172000 seconds). 8 hours is the default value.

Q: The SMPTE 2110 standard calls for PTP time accuracy of 1µ but I hear a good amount of talk that that number is more aspirational.. What is the true accuracy ceiling for a 2110 network?

A: We typically see clock accuracy way better than 1us at the end devices. We'd expect to get to typically +/- 100ns offset from master, in a well designed Boundary Clock network, with the architecture outlined in the presentation.

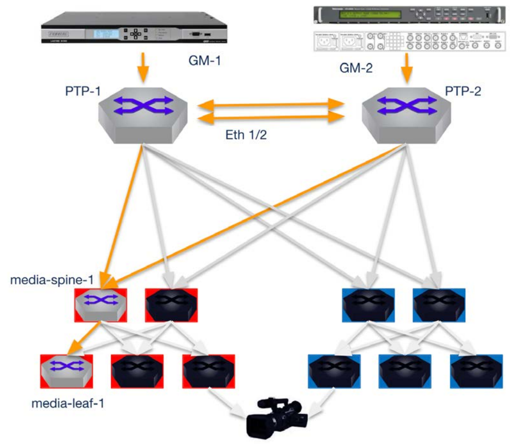
Q: DSCP marking or not?

A: In the recommended point-to-point Boundary Clock environment, DSCP markings are not needed. PTP traffic is naturally generated with high egress priority, and likewise, at ingress, PTP traffic is redirected to the control plane very early in the pipeline. DSCP can certainly be used to influence traffic priority if connecting the Boundary Clocked network to a non-PTP aware, or Transparent Clock PTP network.

PTP Configuration Example

In the Webinar configuration demo, we partially configured our M&E Lab for ST2059-2 Boundary Clocked PTP, following appropriate best practices.

Many webinar attendees asked for the configuration steps to be made available, and so the following section will present those configuration stages, with a light commentary. If you have any questions that remain unanswered, please reach out to your Account Team, or consider asking your question at the Arista Community Forum.



The Diagram above is lifted directly from the Webinar slides. It shows a pair of GrandMasters (GM1 and GM2) attached to a pair of PTP distribution leafs (PTP-1 and PTP-2). These leafs are directly connected together to provide resilience, and lowest latency BMCA between the GrandMasters.

Each PTP distribution switch then cascades PTP to each spine, and each spine cascades PTP to its associated leafs.

All switches in the architecture are to be running Boundary Clock for simplicity, visibility, scale and security.

The demo showed the configuration of PTP-1, PTP-2 and media-leaf-1. In this section, we will also show the appropriate config on media-spine-1 for completeness.

It should be noted that these are partial config snippets that cover PTP. Other config items are very likely to be needed to ensure good unicast reachability, routing, or media multicast operation. All of these non-PTP related config items are omitted for clarity.

Initial Configuration, PTP-1

PTP-1 is one of the PTP distribution switches, it connects to a GrandMaster (GM1), and is used primarily to:

- Convert interface speeds between GM1 and the Spine layer
- Provide distribution of PTP to all spine layer switches

This layer can be omitted in single-spine per side networks if the spine switch can accept the interface speed of the GrandMaster. However, this tends to become more difficult as systems grow, and the spine layer is running at 100G and above. This layer is also difficult to avoid if there are more than a single spine per Red/Blue.

It is possible to merge the PTP distribution functionality with a media leaf pair, or OOB (Out Of Band) monitoring and control switches, but care is needed to ensure the same resilient connectivity to all spines is provided.

When configuring Boundary Clock, there are two stages:

- Global PTP config, to enable boundary clock
- Local interface config, to enable PTP as needed per interface

PTP-1 Global Config

The following global configuration will enable boundary clock mode on the switch, and will define that boundary clock to be locking to and serving PTP Domain 127. All other devices connected to this switch that need to be locked to, or locking to this boundary clock, need to be running in the same domain 127.

The priority settings define how this boundary clock will participate in BMCA for domain 127.

The "ptp source ip" forces PTP packets originated at this switch to be sent with a source IP address of 10.10.10.1. These packets are the announce, sync, follow-up and delay response packets for "master" ports on this switch, and delay-request for the "slave" port on this switch.

The clock-identity is not mandatory, but it does make it easier to identify upstream devices when monitoring or debugging environments. If not set, this is automatically populated by the switch with a system MAC address.

```
ptp mode boundary
ptp domain 127
ptp priority1 10
ptp priority2 1
ptp source ip 10.10.10.1
ptp clock-identity 00:1c:73:ff:ff:0a:00:01
```

Now that we have a global config in place, we need to configure the interfaces on which we want PTP to be active.

PTP-1 Interface Config

An interface config is needed on every interface that we want PTP to be active on.

For this demo, we need to configure interface Ethernet 48 facing the GrandMaster GM1:

```
interface Ethernet48
   description Meinberg ESI_02 Cu
   speed forced 1000full
   switchport access vlan 300
   ptp enable
   ptp announce interval 0
   ptp sync-message interval -3
   ptp delay-req interval -3
```

Adding a description and forcing the speed are not mandatory but are good practice.

In this case the port is an access port on VLAN 300 allowing us to communicate with the GrandMaster for management purposes.

To configure PTP, we use "ptp enable" - now PTP is running on this port, but it would only be running with default message rates if this was the only PTP command we used.

To configure for the recommended message rates for AES-R16-2016, which provide valid rates for BOTH SMPTE 2059-2 and AES67, we need to set the announce, delay-request and sync message rates.

The message intervals are configured as log^2 seconds.

The recommended rates are:

- 1 Announce message per second, the interval is 1 second, so
    - `"ptp announce interval 0"`
- 8 Sync messages per second, the interval is 0.125 seconds, so
    - `"ptp sync-message interval -3"`
- 8 Delay-Request messages per second, the interval is 0.125 seconds, so
    - `"ptp delay-req interval -3"`

With the above config in place and assuming the GrandMaster is appropriately configured we will see the switch lock to the upstream GrandMaster. We can check that with a number of commands:

```
PTP-1#show ptp

PTP Mode: Boundary Clock

PTP Profile: Default ( IEEE1588 )

Clock Identity: 0x00:1c:73:ff:ff:0a:00:01

Grandmaster Clock Identity: 0xec:46:70:ff:fe:00:ff:a8

Number of slave ports: 1

Number of master ports: 0

Slave port: Ethernet48

Offset From Master (nanoseconds): 5

Mean Path Delay (nanoseconds): 503

Steps Removed: 1

Skew (estimated local-to-master clock frequency ratio): 1.0000068869892105

Last Sync Time: 09:24:14 UTC Apr 24 2023

Current PTP System Time: 09:24:14 UTC Apr 24 2023
   Interface       State          Transport    Delay
                                               Mechanism
--------------- -------------- --------------- ---------
   Et48            Slave          ipv4         e2e
```

The "show ptp" status command provides useful information about the local clock identity which matches what we configured in the global section. It also shows the upstream GrandMaster Clock Identity which matches the Meinberg GrandMaster, as expected.

We can see that we are locked through Ethernet48, and by "Steps Removed: 1" we know that there are no other boundary clocks between this switch and the upstream GrandMaster.

We can determine the quality of the lock with "show ptp monitor" (truncated for visibility), which shows the offset-from-master, and mean-path-delay to the upstream master for the last 100 sync messages.

```
PTP-1#show ptp monitor
PTP Mode: Boundary Clock
Ptp monitoring: enabled
Number of entries: 100
Offset from master threshold: 250
Mean path delay threshold: 1500
Skew threshold: not configured
Interface  Time                          Offset from   Mean Path     Skew          Seq Id
                                         Master (ns)   Delay (ns)
---------- ----------------------------- ------------ ------------- ------------- ------
Et48       09:28:07.413 UTC Apr 24 2023  -7            503           1.000006887   32714
Et48       09:28:07.288 UTC Apr 24 2023  18            503           1.000006887   32713
Et48       09:28:07.163 UTC Apr 24 2023  -7            503           1.000006887   32712
Et48       09:28:07.038 UTC Apr 24 2023  1             503           1.000006887   32711
Et48       09:28:06.913 UTC Apr 24 2023  0             503           1.000006895   32710
Et48       09:28:06.788 UTC Apr 24 2023  3             503           1.000006887   32709
Et48       09:28:06.663 UTC Apr 24 2023  -6            503           1.000006887   32708
Et48       09:28:06.538 UTC Apr 24 2023  10            503           1.000006887   32707
Et48       09:28:06.413 UTC Apr 24 2023  3             503           1.000006887   32706
Et48       09:28:06.288 UTC Apr 24 2023  -11           503           1.000006887   32705
Et48       09:28:06.163 UTC Apr 24 2023  10            503           1.000006895   32704
Et48       09:28:06.038 UTC Apr 24 2023  -4            503           1.000006887   32703
Et48       09:28:05.913 UTC Apr 24 2023  4             503           1.000006895   32702
```

Here we see a very good lock - the offset-from-master values are in the 10's of nanoseconds, and the mean-path-delay is stable. A mean-path-delay figure that varies by more than a handful of nanoseconds would indicate something to be investigated if you are directly connected to the upstream master. A mean-path-delay of 0 would suggest that we are not receiving delay-responses from the upstream master.

Now that we are locked to the upstream GrandMaster we can add the other interface configurations that allow us to provide PTP connectivity to the PTP-2 switch and to media-spine-1. These configurations will be similar to those already used for the GrandMaster connection.

First, let's configure the two resilient links that connect PTP-1 to PTP-2, where the backup GrandMaster (GM-2) is located.

Here we'll provide an example of how to **only** enable PTP connectivity. The "no switchport" command creates a routed port and when used without specifying an IP address allows us to prevent any user data connectivity except for PTP.

```
interface Ethernet 1
   description P2P_LINK_TO_media-PTP-2_Ethernet1

interface Ethernet 2
   description P2P_LINK_TO_media-PTP-2_Ethernet2

interface Ethernet 1,2
   no switchport
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3
```

For both interfaces we have provided a useful interface description and configured PTP to run with the same message intervals as before.

We use a very similar approach when connecting PTP-1 to media-spine-1:

```
interface Ethernet 51
   description P2P_LINK_TO_media-spine-1_Ethernet27_1
   no switchport
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3
```

If we now re-check the "show ptp" status, we see the new interfaces are present.

```
media-PTP-1#show ptp
PTP Mode: Boundary Clock
PTP Profile: Default ( IEEE1588 )
Clock Identity: 0x00:1c:73:ff:ff:0a:00:01
Grandmaster Clock Identity: 0xec:46:70:ff:fe:00:ff:a8
Number of slave ports: 1
Number of master ports: 3
Slave port: Ethernet48
Offset From Master (nanoseconds): 6
Mean Path Delay (nanoseconds): 503
Steps Removed: 1
Skew (estimated local-to-master clock frequency ratio): 1.0000068949655345
Last Sync Time: 09:43:43 UTC Apr 24 2023
Current PTP System Time: 09:43:43 UTC Apr 24 2023
   Interface       State          Transport    Delay
                                               Mechanism
--------------- -------------- --------------- ---------
   Et1             Master         ipv4         e2e
   Et2             Master         ipv4         e2e
   Et48            Slave          ipv4         e2e
   Et51            Master         ipv4         e2e
```

Note that we have **NOT** enabled "ptp role master" on any of these interfaces - we want these interfaces to be able to change state thus providing maximum resilience, in the case of maintenance / failure of link, failure of a GrandMaster, or failure of a switch.

Initial Configuration, PTP-2

PTP-2 provides the same services as PTP-1 but it's connected to the backup GrandMaster (GM-2). Therefore the config looks very similar to the config on PTP-1.

The important differences are:

- Clock-ID identifies this switch
- Priority2 makes this switch less preferred than PTP-1
- PTP Source IP is local to this switch

Below we see the configuration details, but for the detailed commentary, please see that for PTP-1.

PTP-2 Global Config

The global config is very similar with the exception of the localised changes to priority2, PTP source IP and clock identity. The "show ptp" command can be used to verify that the switch is now running the PTP service. Without connectivity to its local GrandMaster (GM-2), nor PTP-1, this switch would be an isolated island of PTP running in local holdover mode.

```
ptp mode boundary
ptp domain 127
ptp priority1 10
ptp priority2 2
ptp source ip 10.10.10.2
ptp clock-identity 00:1c:73:ff:ff:0a:00:02
```

PTP-2 Interface Config

The connection to the local backup GrandMaster (GM-2) is the same as on PTP-1.

```
interface Ethernet54
   description Meinberg MRI1 Optical
   speed forced 1000full
   switchport access vlan 2222
   no switchport
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3
```

The links to PTP-1, which allow for fast and resilient BMCA operations, are also similar to the PTP-1 configs on the other side of these links.

```
interface Ethernet1
   description P2P_LINK_TO_media-PTP-1_Ethernet1
   no switchport
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3

interface Ethernet2
   description P2P_LINK_TO_media-PTP-1_Ethernet2
   no switchport
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3
```

And finally the config from PTP-2 down to media-spine-1:

```
interface Ethernet51
   description P2P_LINK_TO_MEDIA-SPINE-1_Ethernet28/1
   no switchport
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3
```

At this point if we were to use "show ptp" or "show ptp monitor" on either PTP-1 or PTP-2, we would see that both switches are locked to the same GrandMaster (GM1) and that PTP-2 is locked via one of either Ethernet1 or Ethernet2, to the upstream PTP-1 switch:

```
PTP-2(config-if-Et51)#show ptp
PTP Mode: Boundary Clock
PTP Profile: Default ( IEEE1588 )
Clock Identity: 0x00:1c:73:ff:ff:0a:00:02
Grandmaster Clock Identity: 0xec:46:70:ff:fe:00:ff:a8
Number of slave ports: 1
Number of master ports: 3
Slave port: Ethernet1
Offset From Master (nanoseconds): 11
Mean Path Delay (nanoseconds): 1009
Steps Removed: 2
Skew (estimated local-to-master clock frequency ratio): 1.0000062128205105
Last Sync Time: 10:04:46 UTC Apr 24 2023
Current PTP System Time: 10:04:46 UTC Apr 24 2023
   Interface       State          Transport    Delay
                                               Mechanism
--------------- -------------- --------------- ---------
   Et1             Slave          ipv4         e2e
   Et2             Passive        ipv4         e2e
   Et51            Master         ipv4         e2e
   Et54            Master         ipv4         e2e
```

We can see that we are locked to the upstream GrandMaster (0xec:46:70:ff:fe:00:ff:a8) via PTP-1 on interface Ethernet1. Since there are two direct paths to PTP-1 we see that the other path (Ethernet2) has been put into "passive" mode by the BMCA process. This port is available to be active again if something were to happen to the Ethernet1 path.

This output also shows that there is a boundary clock between this switch and the GrandMaster (Steps Removed: 2), and it shows that we are mastering PTP towards the backup GrandMaster (GM-2) and media-spine-1. The backup GrandMaster is usually configured to not lock to other potential masters and to be ready to become the active GrandMaster should the primary GrandMaster (GM-1) become a less good clock candidate.

While we are mastering towards the media-spine-1 device we can see from media-spine-1's perspective that this link is in "passive" mode. Media-spine-1 will be locked directly to PTP-1 instead and therefore also to GM1. This path is preferred because it has fewer Boundary Clock hops than the media-spine-1 -> PTP-2 -> PTP-1 -> GM1 path.

media-spine-1 Global Config

The spine global config shows nothing unexpected. The priority1 setting has been bumped to a new number, to ensure that this tier of switches are not part of the BMCA between GM1/2, or PTP-1 and PTP-2. The source IP address and Clock Identity have been localised too.

```
ptp mode boundary
ptp domain 127
ptp priority1 20
ptp priority2 1
ptp source ip 10.10.10.3
ptp clock-identity 00:1c:73:ff:ff:14:00:01
```

media-spine-1 Interface Config

The only difference in interface configurations that we see at this level is the addition of IP addresses on the routed ports from spine to leaf.

```
! Connection from PTP-1 to media-spine-1
interface Ethernet27/1
   description P2P_LINK_TO_MEDIA-PTP-1_Ethernet51
   no switchport
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3

! Connection from PTP-2 to media-spine-1
interface Ethernet28/1
   description P2P_LINK_TO_MEDIA-PTP-2_Ethernet51
   no switchport
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3

! Connection from media-spine-1 to media-leaf-1
interface Ethernet1/1
   description P2P_LINK_TO_MEDIA-LEAF-1_Ethernet49/1
   no switchport
   ip address 192.168.102.64/31
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3
```

media-leaf-1 Global Config

There is a subtle difference for the media-leaf-1 global config - there is no "ptp source ip" address configured. If no "ptp source ip" address is configured the switch will use the routed port IP or the SVI address for access ports, as appropriate. If preferred though the "ptp source ip" construct can be used and this will force the use of the defined address.

If the IP address of routed ports is not set, or there are VLANs being used without SVIs configured, then the "ptp source ip" config is very strongly recommended.

Again we see that the priority1 has been set to the next tier on from the spine.

```
ptp mode boundary
ptp domain 127
ptp priority1 30
ptp priority2 5
ptp clock-identity 00:1c:73:ff:ff:1e:00:05
```

media-leaf-1 Interface Config

The spine facing link follows a similar pattern:

```
interface Ethernet49/1
   description P2P_LINK_TO_MEDIA-SPINE-1_Ethernet1/1
   no switchport
   ip address 192.168.102.65/31
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp sync-message interval -3
```

We do see a difference when we look at the host facing ports however. Here we see that the "ptp role master" config item has been applied. This config prevents a host from participating in BMCA with the switch. It prevents hosts that we know are not candidates for becoming the GrandMaster from accidentally or nefariously perverting the BMCA process. A large number of potential media end-points have the capability to provide PTP GrandMaster services, so designing against unwanted GrandMasters with the "ptp role master" command is a very strong recommendation.

Unless there is an atypical design situation, any "ptp enable" ports that are not fabric ports nor ports that face a GrandMaster candidate should be configured as "ptp role master".

```
interface Ethernet54/1
   description Imagine SNP-2118212563-Data-A
   speed forced 100gfull
   no switchport
   ip address 172.24.151.1/30
   ptp enable
   ptp announce interval 0
   ptp delay-req interval -3
   ptp role master
   ptp sync-message interval -3
```

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

India—R&D Office
Global Tech Park, Tower A, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989