ARISTA
Technical Bulletin

Arista Precision Time Protocol

Overview

Precision timing has become increasingly important with the proliferation of low latency and high performance applications. It is especially critical in HPC or grid environments where there is a desire to correlate or synchronize events within microseconds, or measure utilization or latency with the highest accuracy. In financial networks and applications there is a need to instrument transactions among several devices or participants with nanosecond precision, and capture data with precision for application back testing and performance validation.

Traditionally when precision timing and synchronization is required between individual hosts or networks a coaxial connection to a grand master clock with GPS synchronization is implemented. This requires dedicated cabling and timing cards in every end point, along with GPS antennas for GPS synchronization and atomic clock options for holdover stability, as well as a stratum based hierarchy.

PTP was designed to provide precise time distribution over an Ethernet or IP network, as opposed to other timing solutions that require a discrete physical infrastructure within the data center. PTP provides a standardized, end to end precision timing implementation that can be deployed on a standard Ethernet network, either in band or out of band of the standard data plane. This removes the requirement for a separate, dedicated timing infrastructure while providing similar or increased precision.

Until now, data center class switches have not provided a hardware based PTP implementation. This has limited the scale and precision of PTP implementations by requiring
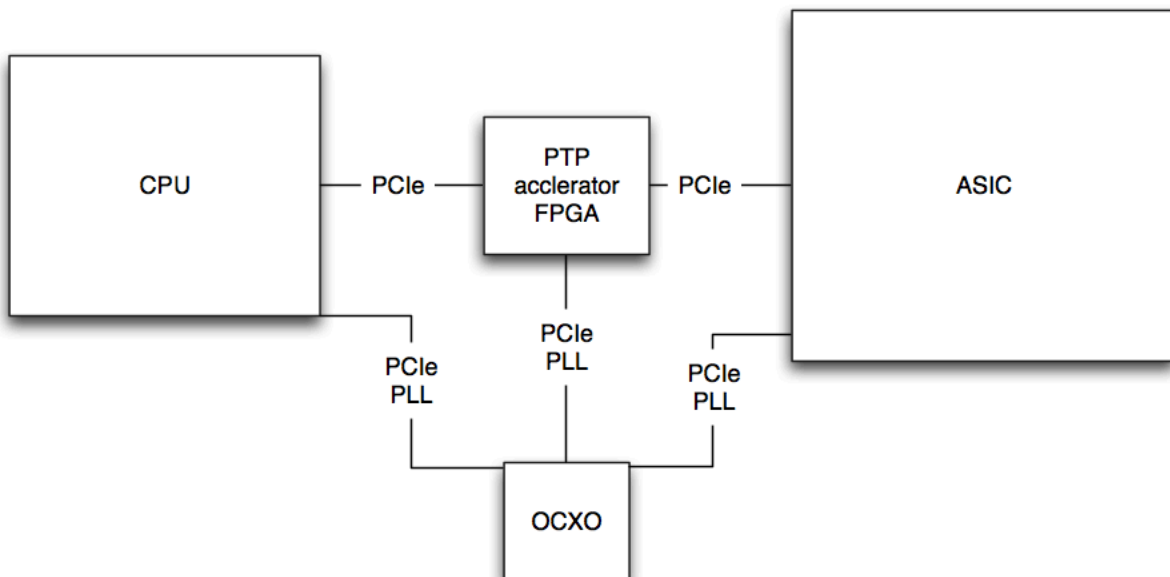
dedicated PTP hardware at every point, essentially minimizing the benefit of a common infrastructure for both data forwarding and time synchronizations. This document will review how the Arista 7150 data center switch enables high precision time distribution directly in the data path, while improving the scale, precision, and cost of a precise time infrastructure.

PTP Baseline requirements: A Solid Hardware Foundation

The Arista 7150S delivers robust PTP functionality in a data center class Ethernet switching platform. The PTP implementation on the 7150 is a two-step process that is hardware-assisted for the highest accuracy possible. On ingress or egress of prescribed PTP event messages, the 7150 ASIC records the arrival or transmission time with ~10ns resolution using a high resolution 31-bit hardware timestamp. This enables a high precision boundary clock and tightly integrated transparent clocking functionality that provides nanosecond level time synchronization performance.

Preamble
Packet
Time [30:23]
Time [22:15]
Time [14:7]
0 | Time [6:0]
New CRC

The 7150S also has an onboard high precision "Oven-Controlled Crystal Oscillator" (OXCO) that provides exponentially better cycle-to-cycle jitter and holdover performance than a standard system clock. This clock is the time base for the system and disciplines the CPU, ASIC, and a system FPGA that provides hardware acceleration for PTP, LANZ, and other unique EOS features.
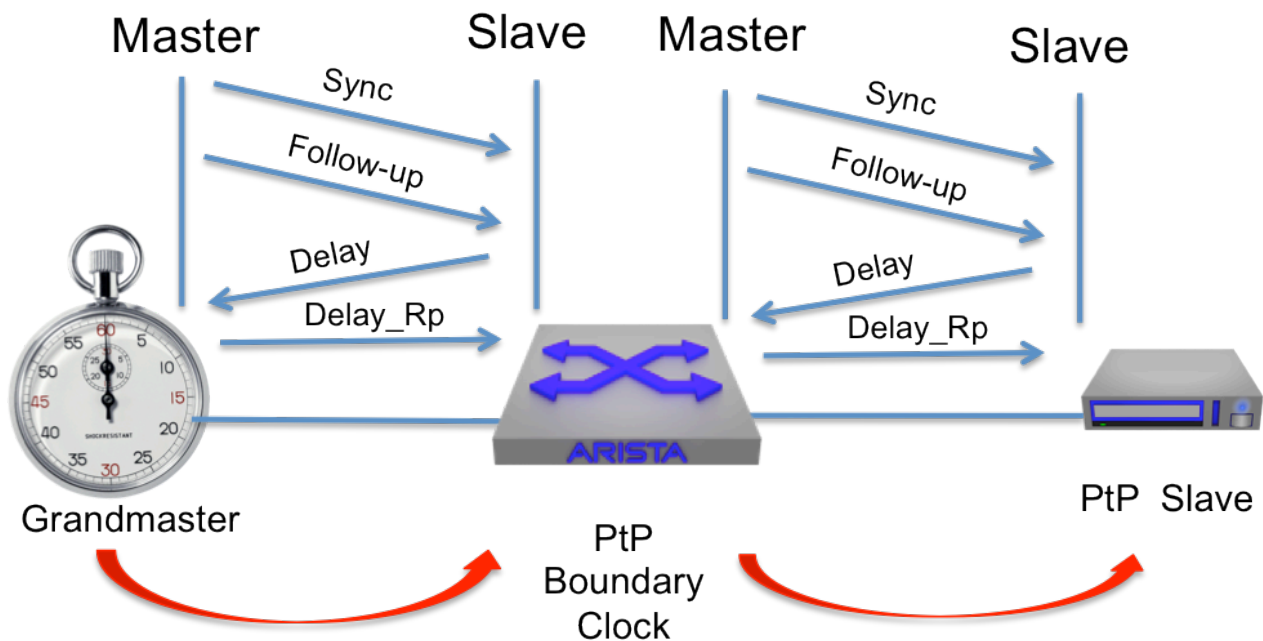
The 7150S hardware-based functionality enables the switch to participate and enhance a PTP implementation in two ways: by improving the scalability of the PTP grandmaster and delivery of PTP messages, and by providing an accurate calculation and correction of the time taken by PTP packets to traverse devices or the entire network.

The boundary clock function allows end devices (ordinary clocks or "slaves") to be served by the local switch, offloading processing from the grand master, and distributing time across VLAN or routed boundaries. Transparent clocking instruments how long PTP packets spent in the network or device so that slaves can synchronize quickly and accurately with the time being distributed by the master clock. Both of these use models are discussed in more detail below.

The 7150 Boundary Clock : Scaling the Grandmaster

One factor to consider in the design of a PTP system is how many slaves or network segments a grandmaster clock can service while maintaining a set level of accuracy. This is typically limited by the amount of PTP sync packets a master can generate, delay messages it can receive, or interfaces it supports without delaying PTP messages and introducing unintentional clock skew. The PTP boundary clock function addresses this issue as an intermediate clock that acts as a slave to the master and performs master functions for multiple slaves downstream, while maintaining a high level of accuracy.
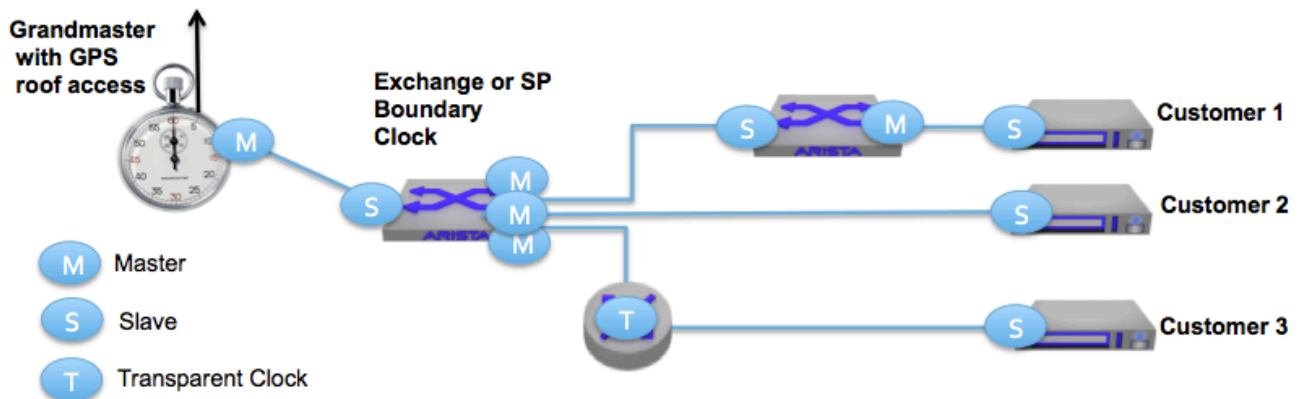

As an example, if a master clock is capable of supporting 100 devices, and each boundary clock can support an additional 100 devices, the PTP implementation can scale by 100X by implementing a single tier of boundary clocks in the network, achieving a much larger network than with the grand master clock alone. This also prevents having to provision numerous grand masters at different points in the network. As the boundary clock becomes the authoritative

source of time to slaves downstream, the on-board high precision clock is absolutely critical to maintain precision in the case where synchronization with the master clock is lost.

The boundary function is useful in deployments where a grandmaster is installed at a fixed location with GPS access and the time distribution network encompasses multiple network devices with many slaves, or even multiple sites. Deploying multiple boundary clocks in a remote site is useful for providing redundancy in a PTP deployment. Via the Best Master Clock (BMC) election process, the boundary clock ensures continued synchronization for downstream slaves if a single master fails.
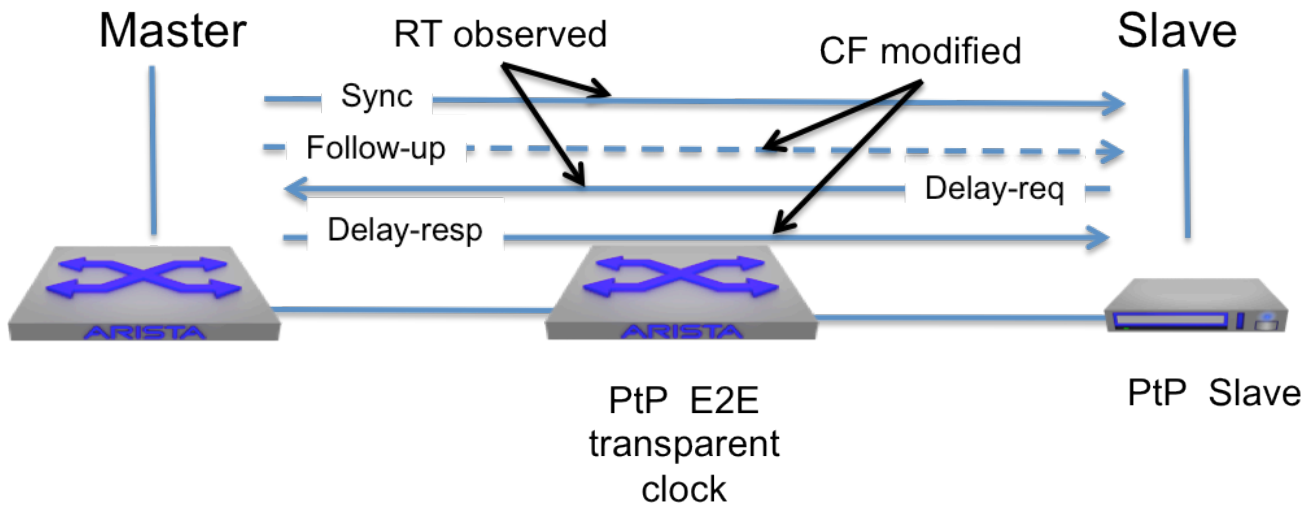
The 7150 Boundary Clock : Multi-tenancy Without Compromise

The 7150S boundary clock implementation addresses a problem seen in many of the popular financial colocation data centers, which is a limitation on provisioning unique GPS antennas for every customer. As opposed to dedicated roof access and serial cabling infrastructure, data center providers would like to be able to deliver precision timing in-band with lower cost and better scalability. The consumers of the time signal want an accurate, precise sync without having to buy additional infrastructure or manage off-premise wiring. The 7150S boundary clock implementation addresses these desires and provides a means of distributing precise time over the Ethernet cross connect to each participant, while providing network isolation and guaranteed services via routing and data plane filtering. When deployed with a transparent clock enabled Arista switch this network can scale both in number of slaves and physical distance from the grandmaster.
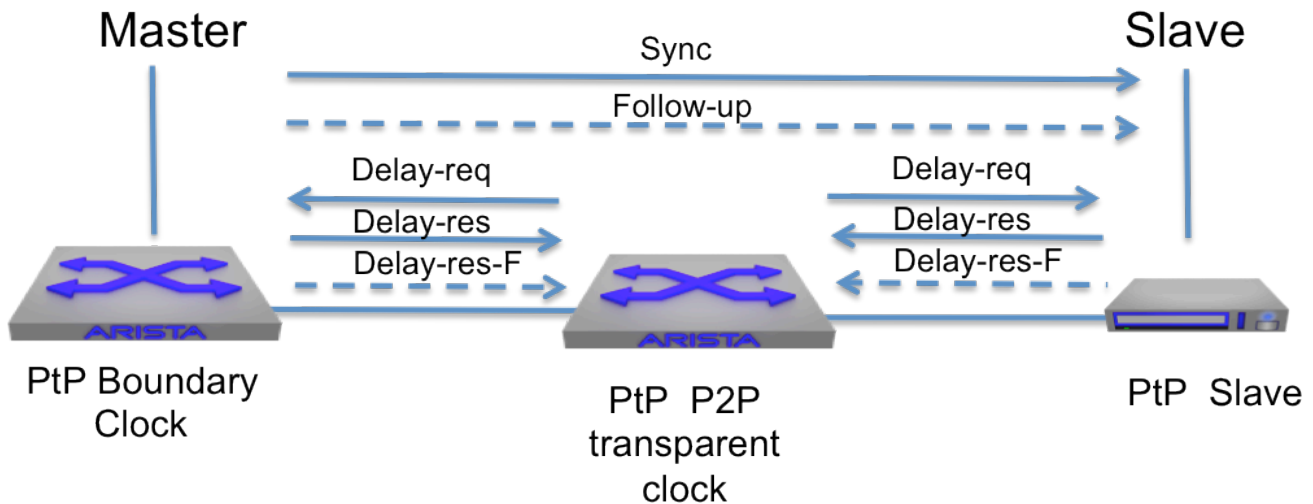

The 7150 Transparent Clock: Enabling the End to End Precision

PTP relies on instrumenting the delay between the master and slave to correctly instrument offset. The less variability in the delay, the more accurate the synchronization will be maintained. In some PTP deployments, the slaves may be a number of network hops away from the master clock. Each network hop has the potential to introduce a non-deterministic amount of latency based on queuing or congestion of the interface at the time the PTP packet is transmitted.

Master   RT observed   CF modified   Slave

Sync
Follow-up
Delay-req
Delay-resp

PtP E2E transparent clock

PtP Slave

The Arista PTP transparent clock functionality provides 2 modes to eliminate queuing processing, and propagation delay for PTP messages. In transparent E2E (End to End) mode each Arista 7150 in the path to the slave measures the residence time (RT) of a PTP packet in its queue and modifies the RT field in the PTP messaging to indicate if the message was delayed. The slave receives the PTP message and uses the RT data to calculate and remove the jitter the network introduced, thereby maintaining lock with the master based on a consistent delay.

Master   Sync   Slave

Follow-up
Delay-req
Delay-res
Delay-res-F

Delay-req
Delay-res
Delay-res-F

PtP Boundary Clock

PtP P2P transparent clock

PtP Slave

In transparent P2P (Peer to Peer) mode, an independent set of PTP messages operating on a point to point basis between each transparent clock calculates peer delay versus path delay, in addition to correcting the RT of the PTP packet. This per hop delay analysis provides a mechanism to remove propagation delay variance, which may be especially useful in implementations where there are multiple links that may vary in latency between the slave and master. In large scale data center environments the transparent clock is key in mitigating queuing or link variability.

Conclusion

The 7150S is the first product in the data center switch segment to offer complete hardware support for PTP boundary and transparent clock functions with the highest levels of precision and scalability in the market. These capabilities unlock the limits on the scale and precision of PTP implementations that no longer require dedicated PTP hardware at every point and provide a common infrastructure for both data forwarding and time synchronization.