AN NTP STRATUM-ONE SERVER FARM FED BY IEEE-1588

Richard E. Schmidt and Blair Fonville
Time Service Department, U.S. Naval Observatory
3450 Massachusetts Ave. NW, Washington, D.C. 20392, USA
rich.schmidt@usno.navy.mil, blair.fonville@usno.navy.mil

Abstract

For the past 16 years, USNO's Network Time Protocol (NTP) stratum-1 servers have been synchronized to its Master Clocks via IRIG-B time code on a low-frequency RF distribution system. The availability of Precise Time Protocol (PTP, IEEE-1588) host-based interfaces will enable the deployment of NTP servers fed by IEEE-1588 over conventional Ethernet networks. This paper describes a PTP GrandMaster/Slave network which maintains synchronization of internal (stratum-0) NTP reference clocks to within tens of nanoseconds. The configuration of a PTP network and the effects of various PTP configuration options are demonstrated.

USNO NETWORK TIME SERVICE

The U.S. Naval Observatory has provided public stratum-1 NTP network time service continuously since 1994 **[1]**. At present, the Washington, D.C., NTP servers receive more than one-half billion hits per day (with on-the-hour peaks in excess of 20,000 packets per second) from millions of unique IP addresses. In order to handle this load, and to provide redundancy, USNO operates an NTP "server farm" with Ethernet traffic controlled by a pair of CAI Networks 590SG load balancers (Figure 1). All traffic to these NTP servers passes first through the one active member of the pair of load balancers (the inactive being a hot stand-by). It is the load balancer's duty to assign each incoming NTP request to one of the available servers, balancing the load by round-robin, weighted round-robin, least active connections, or other algorithm. Each NTP server returns packets to the load balancer for forwarding back to the requestor.

The farm NTP servers are synchronized to the USNO Master Clocks using IRIG-B time code. The current standard NTP computers are Hewlett-Packard rx2620 Itanium2 servers running HP-UX 11.31. Each server has two Symmetricom bc635U-PCI bus clocks which synchronize to IRIG-B and provide memory-mappable BCD time clocks. The standard NTP distribution is augmented with a USNO-developed NTP reference clock driver for the PCI bus clocks. Dual low-frequency RF distribution systems provide IRIG-B from both Master Clocks 1 and 2. This mode of NTP timing is simple and robust, but limited to the microsecond accuracy of our IRIG-B123 analog time code. It constrains the physical distribution of the server farm to proximity with the RF distribution system.

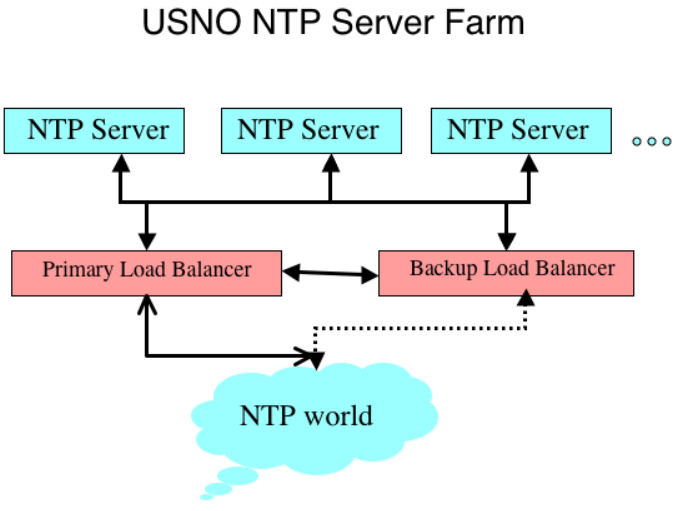

Figure 1. Conceptualized NTP server farm.

IEEE-1588 PRECISE TIME PROTOCOL

The Precise Time Protocol (PTP) is specified by IEEE Std 1588™-2008, *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems* **[2].** PTP is a hierarchical master-slave timing protocol that can achieve network component synchronization at the tens-of-nanoseconds level using a combination of software and augmenting hardware for precision time-stamping of messages. With the advent of commercially-available PTP slaves in the form of PCI bus clocks, it becomes feasible to replace IRIG-B timing of NTP servers with PTP. In 2010, USNO configured a PTP network designed to synchronize its NTP servers over an Ethernet LAN.

PTP GRANDMASTER

The PTP GrandMaster represents the top stratum of the IEEE-1588 timing hierarchy. For our purposes, it must be synchronized to a USNO Master Clock reference. We chose an existing FEI-Zyfer Gsync Model 391 Time and Frequency System **[3]**, adding FEI-Zyfer PTP-enabled Ethernet module 385-4097-02. The Gsync GrandMaster synchronizes to a USNO Master Clock using either a 1PPS or 5 MHz signal. Time of day is loaded into the Gsync from an NTP-synchronized computer, which opens a telnet socket and sends Gsync commands to set the current time of day. When the Gsync has completed a frequency lock to the external reference, it enables PTP GrandMaster operation over the Ethernet port on its PTP module.

PTP SLAVES

For this experiment, USNO has designed a new generation of stratum-1 NTP servers using PTP bus clocks or a combination of PTP and IRIG-B bus clocks. The NTP servers are built on fairly generic Hewlett-Packard ProLiant DL120 G6 servers using Intel® Pentium® G6950 2.8 GHz processors with 4GB RAM **[4].** The operating systems tested were CentOS 5.5 **[5]** and FreeBSD 8.1 **[6]**. To minimize power consumption, heat dissipation, and cost, we selected these systems without Dual-Core or Quad-Core processors and without redundant power supplies (which require additional fans) nor hot-pluggable disk drives. The Oregano Syn1588PCIe PTP bus clock **[7]** was chosen as NTP reference clock because of its support for both 3.3v and 5v PCI and PCI-express formats, as well as its support for FreeBSD UNIX and LINUX operating systems (including CentOS). Oregano has developed an NTP reference clock driver for the Syn1588PCI that was integrated by USNO into the latest public release of NTP,

which was version 4.2.6p2 **[8]**.

NTP Server Considerations

Use of IEEE-1588 PTP as an NTP reference clock entails some design differences from IRIG-B bus clocks. IEEE-1588 requires a software implementation of the PTP control stack which controls PTP options and assignment of timing hierarchy roles, as well as binding to Ethernet NIC devices. In the case of GrandMasters, the PTP stack controls the frequency of *Announce*, *Sync*, and *Delay_Resp* messages. The Gsync GrandMaster handles this in firmware with control options, whereas the Syn1588PCI slave utilizes a PTP stack process running on the slave. (Incidentally, there is a software-only implementation of the PTP protocol stack which is available from SourceForge **[9].**) The PTP protocol propagates the current year, month, and day numbers, which IRIG-B provides only in versions B124-127 **[10]**. Unlike IRIG-B time code, PTP has provisioned a means of leap-second propagation through the *currentUtcOffset* field of the Announce message.

PTP Accuracy and System Design

In 2005, John Eidson, Agilent Techologies, discussed a number of design issues affecting timing accuracy of oscillators on a PTP network **[11]**. Eidson concludes that the achievement of timing accuracy with a precision below 20 nanoseconds requires (among other considerations) careful control of environment, especially temperature; better oscillators than those in stock equipment; and faster sampling (rate of PTP message exchange) than the PTP default 2-second interval. We looked at each of these factors in our PTP network design.

PTP Hardware Phase Measurement

Figure 2 provides an overview of USNO's test facility for PTP hardware.

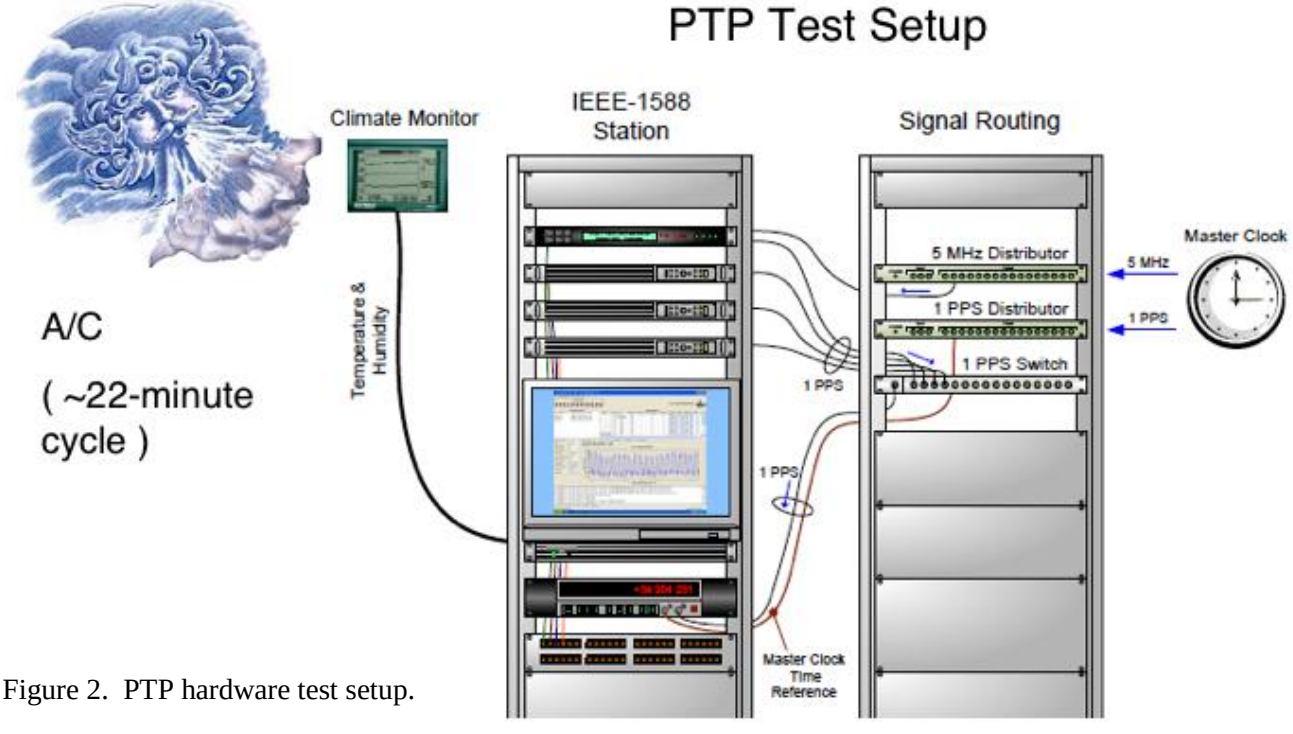

Figure 2. PTP hardware test setup.

The USNO "VDAS" data acquisition system was used to monitor 1PPS phase measurements of the Gsync GrandMaster and the Oregano syn1588 slaves, each of which provide hardware 1PPS outputs. The USNO Master Clock provides a 5 MHz reference and 1PPS distribution. VDAS records phase measurements of PTP components with respect to the Master Clock at 1-minute intervals, and plots phase offsets in real time. VDAS also records ambient temperature and humidity. This facility was set up in a common room without special temperature and humidity control. A forced-air cooling system about 1 meter away cycled cool air with a period of about 22 minutes.

AMBIENT TEMPERATURE AND PTP

Our Oregano syn1588 PCI-express bus clocks were purchased with standard crystal oscillators. A strong correlation of the 1PPS phase offset of these cards with respect to the USNO Master Clock (the Gsync having an ovenized crystal oscillator and so not contributing) is seen in Figure 3 below, where the 22-minute periodicity of the room air cooling is mirrored in the observed phase offsets of three Oregano Syn1588PCIe PTP cards, which were hosted in three independent ProLiant servers. Measurements were made using the 1PPS outputs of the Syn1588PCIe cards. The timing stability of the Gsync GrandMaster, as measured at its 1PPS output, is shown in Figure 4 below.

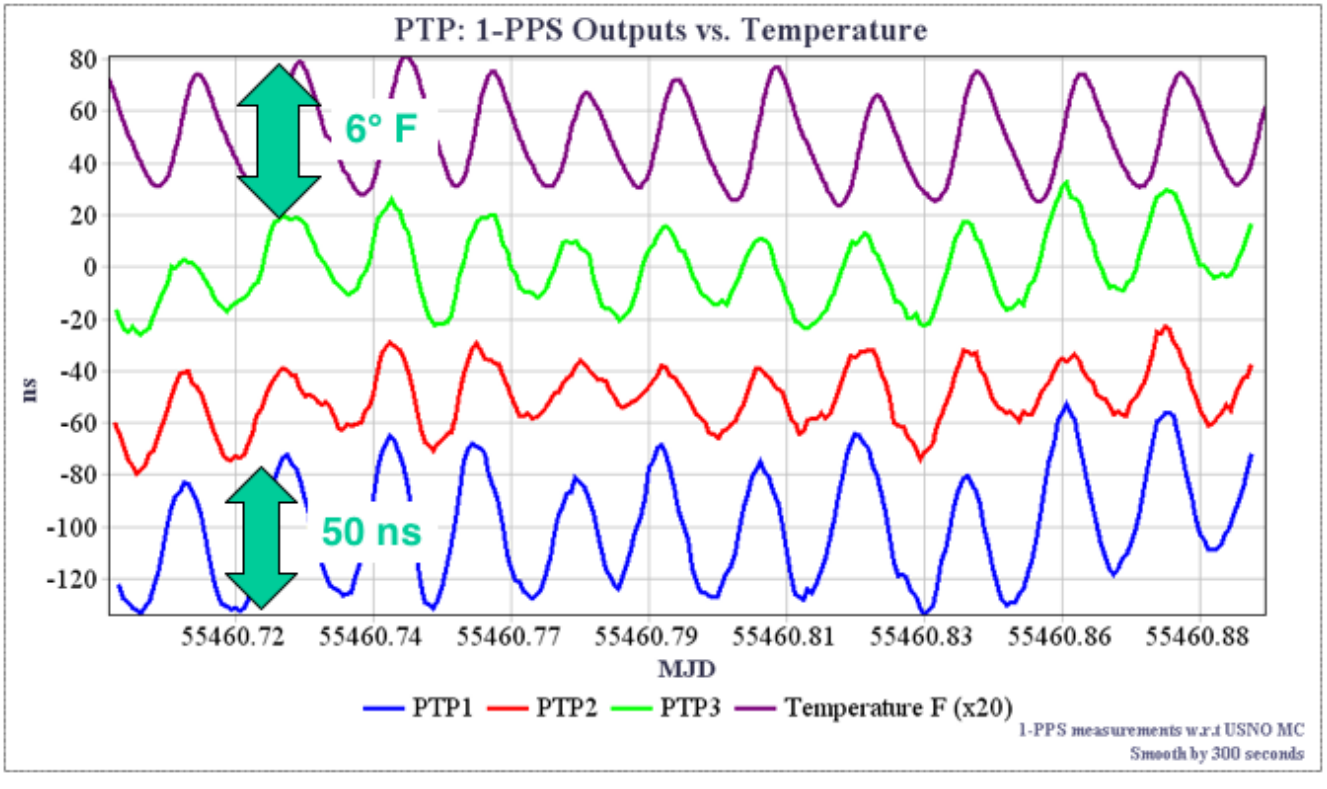

Figure 3. Comparison of ambient temperature (violet, scaled × 20) with 1PPS phase offsets of three independent Syn1588PCIe PTP slaves (green, red, blue).

Oregano provided an upgrade to its standard Syn1588PCIe, which utilizes the KVG O.60.71992-LF 25.000 MHz OCXO said to provide less than ±0.5 ppm frequency stability over a temperature range of 0 - 70° C **[12]**. Figures 5 and 6 show measured improvement of phase error with this OCXO, and overlapping Allan deviation of the same.

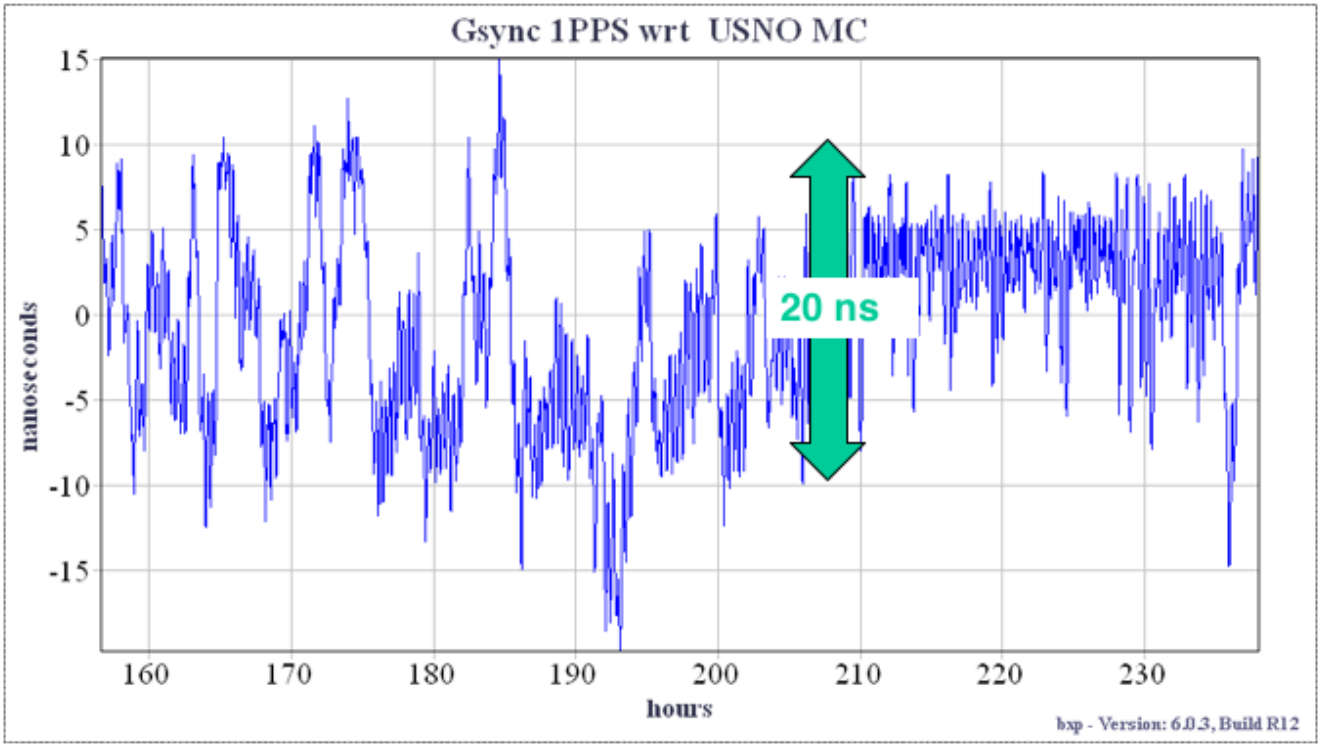

Figure 4. Typical Gsync phase stability (with respect to the USNO Master Clock).

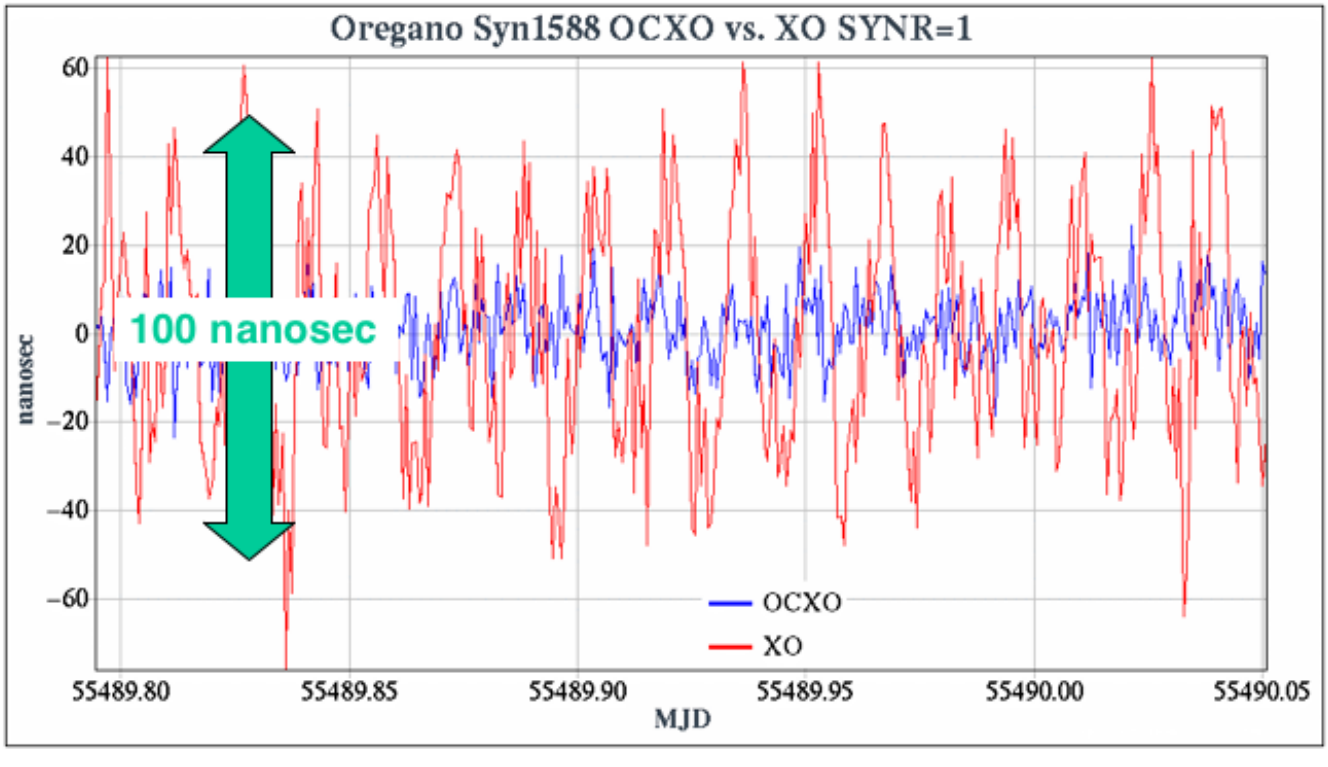

Figure 5. Comparison of phase error of the standard Syn1588PCIe (red) with the OCXO version (blue).

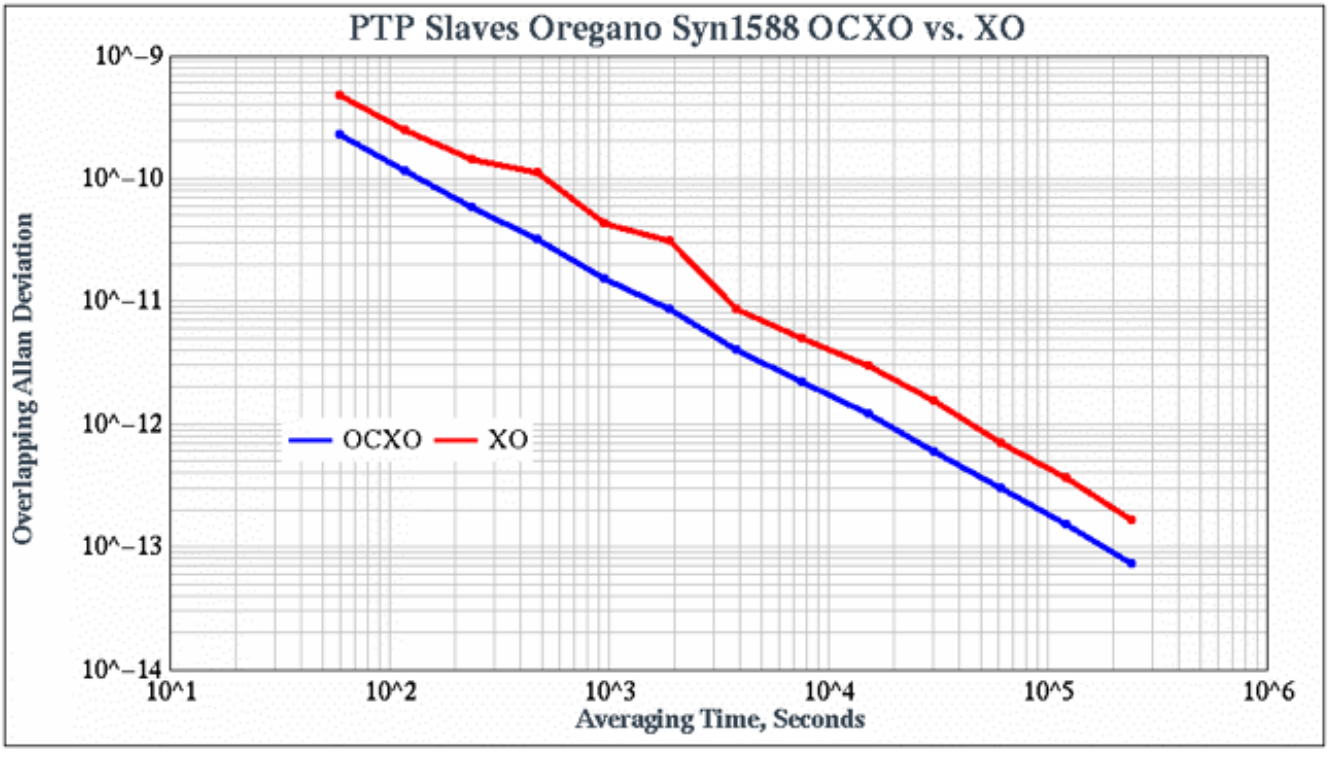

Figure 6. Overlapping Allan deviation, Oregano Syn1588PCIe OCXO vs. XO.

The advantage of using the ovenized crystal option for the Syn1588PCIe becomes immediately evident when we stop the PTP stack, which stops steering the PTP slaves to the GrandMaster. Figure 7 below shows 1PPS outputs from a freewheeling OCXO Syn1588PCIe and a freewheeling XO Syn1588PCIe. The latter shows 100 microseconds of response to temperature variations, a factor of ten worse than the OXCO version.

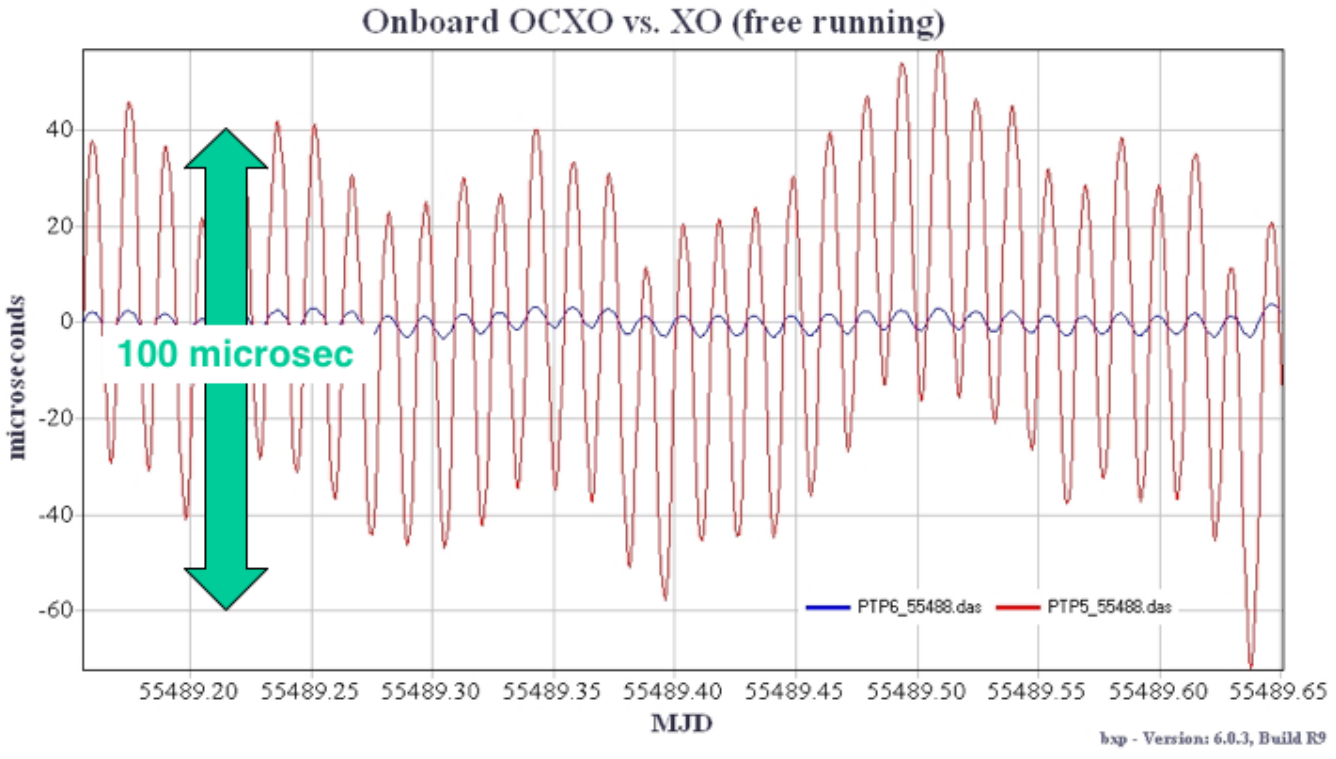

Figure 7. Effect of freewheeling on OCXO (blue) and XO (red) Syn1588PCIe cards.

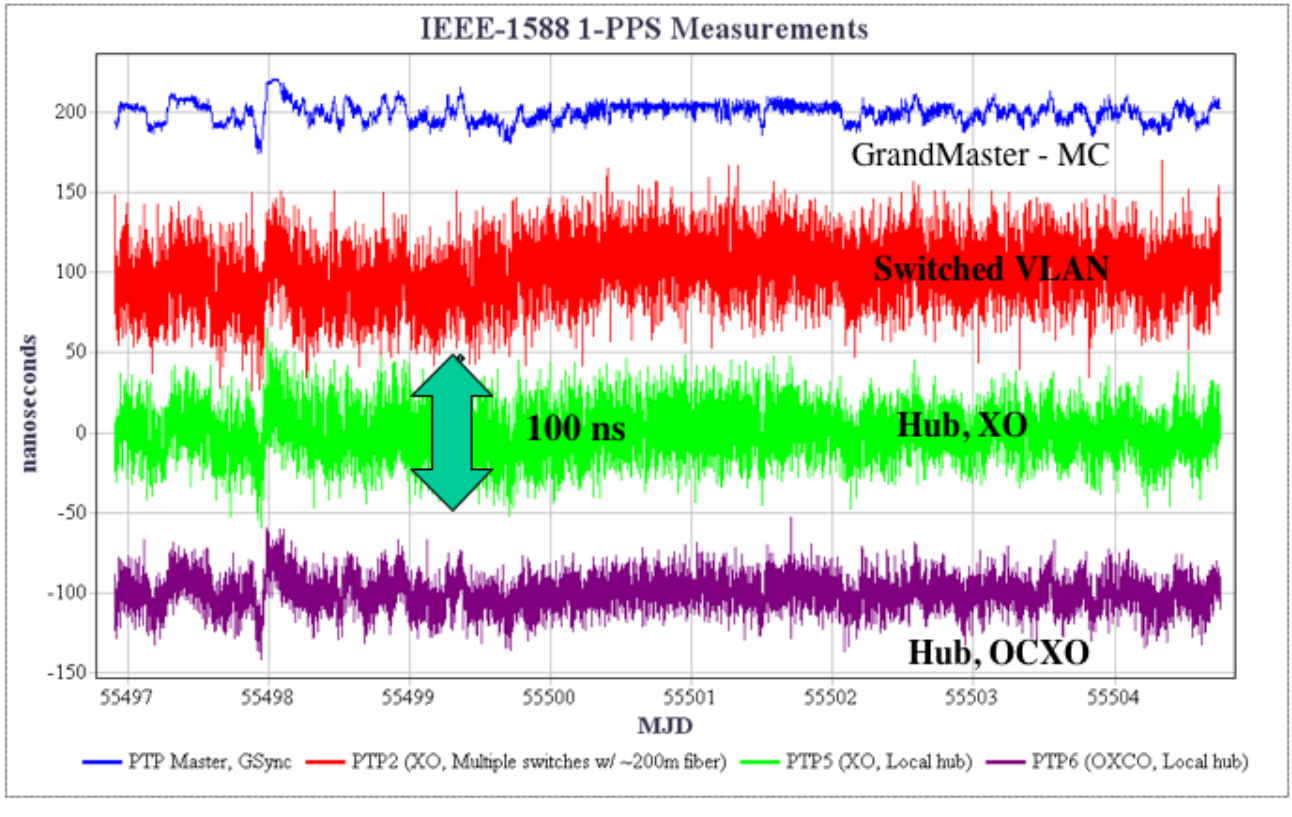

Figure 8. Comparing operation through a switch vs. a hub, and with OCXO.

ETHERNET SWITCHES AND HUBS

PTP accuracies improve when operating on hubs (repeaters) rather than on switches, which can store and forward traffic. We compared operation of the Gsync GrandMaster and three Syn1588PCIe slaves on a Hewlett-Packard Procurve hub with operation through a pair of Juniper EX4200 hubs, which were separated by about 200 meters of fiber optics. Figure 8 shows measurements of phase offsets of these configurations, as well as use of an OXCO Syn1588PCIe slave over the hub. Figure 9 shows overlapping Allan deviations for this data.

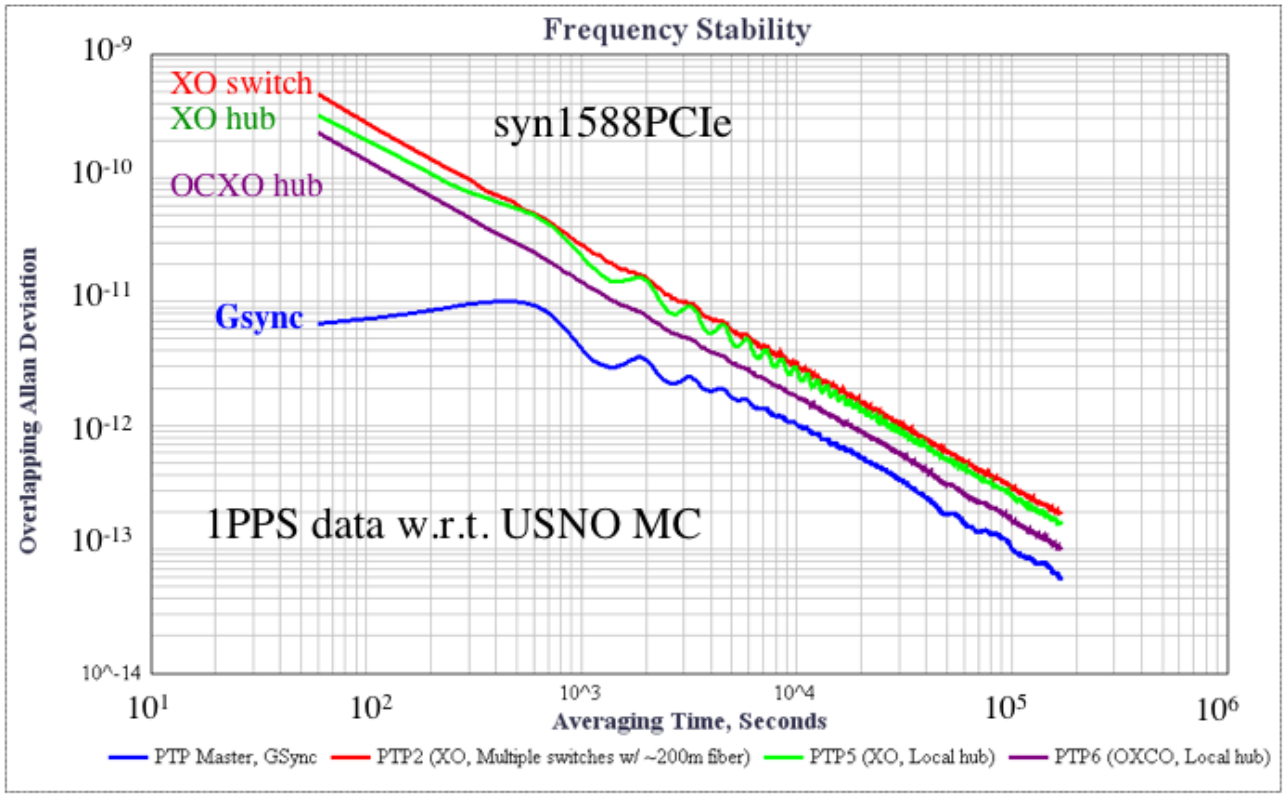

Figure 9. Overlapping Allan deviations, comparing operation through a switch vs. a hub, and with OCXO.

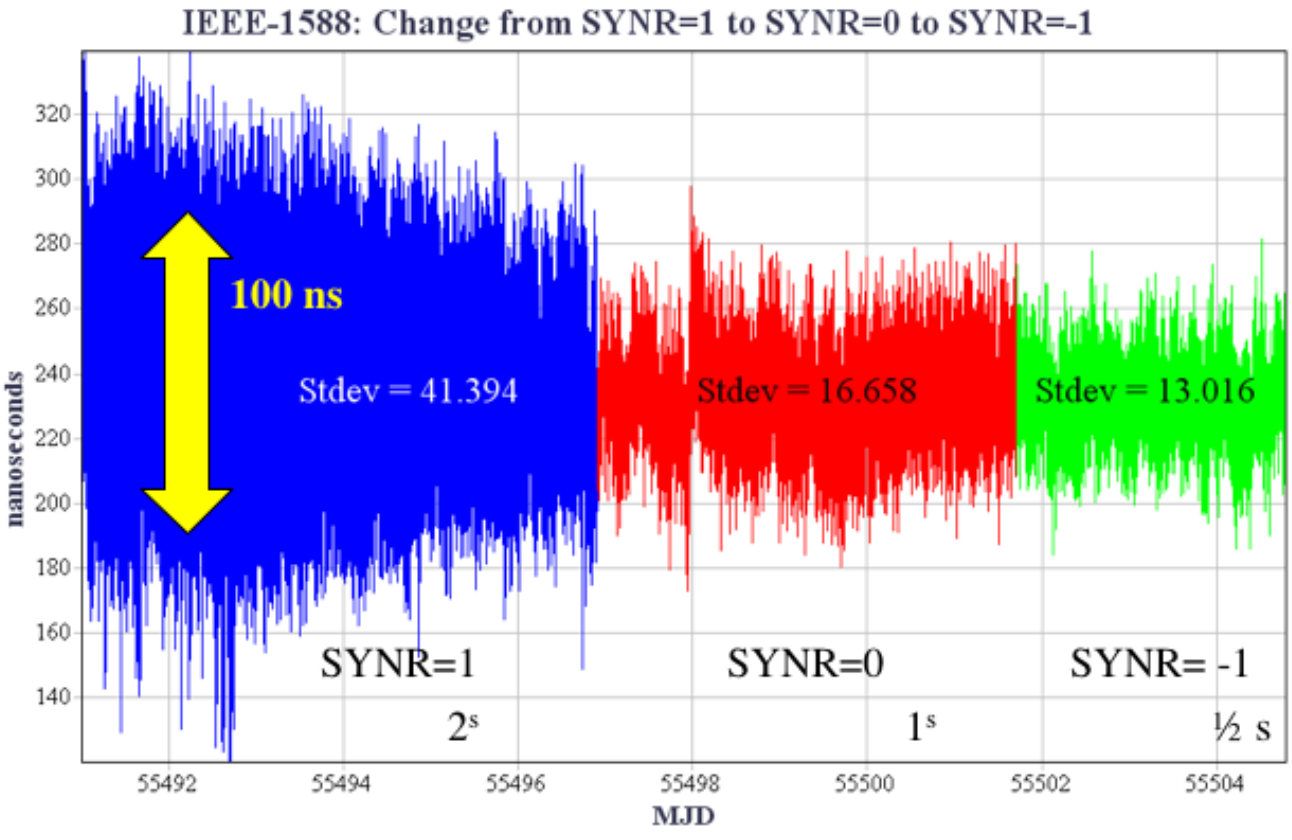

Figure 10. Effect of varying PTP synchronization messaging interval.

EFFECT OF MODIFYING DEFAULT SYNCHRONIZATION MESSAGE RATE

In the IEEE-1588 protocol, the rate at which the GrandMaster sends synchronization messages, SYNR, is expressed as $\log_2$(sync interval) in seconds. The default is SYNR=1, a synchronization interval of 2 seconds. Where network bandwidth is not an issue, decreasing SYNR can tighten the synchronization feedback loop, as seen in Figure 10, where standard deviations in nanoseconds are shown. A significant increase in precision is obtained at the expense of doubling the messaging rate, as shown in Figure 11 with overlapping Allan deviations of the above data, but beyond SYNR=0 only modest improvement was found.

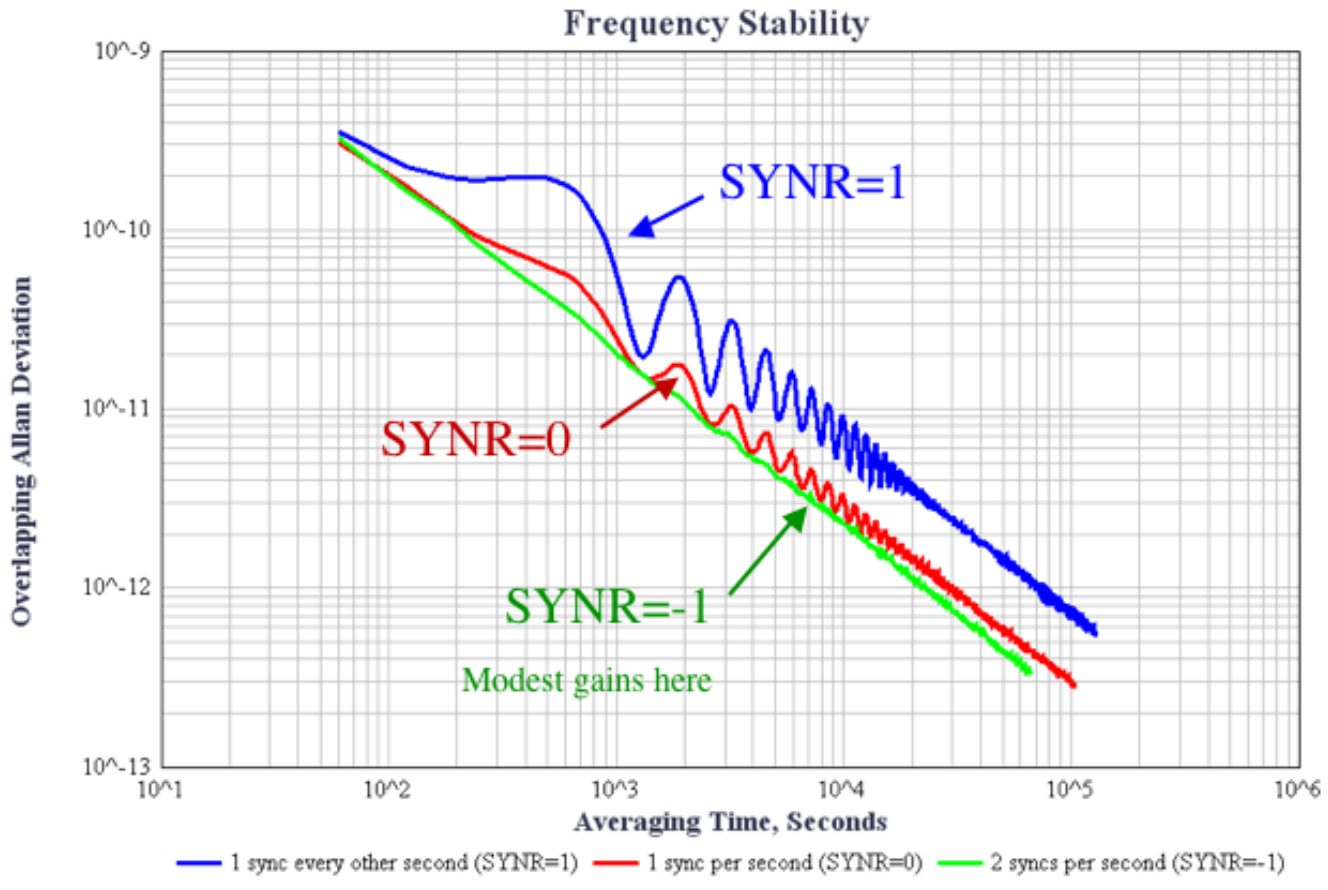

Figure 11. Overlapping Allan deviations, effect of varying PTP synchronization messaging interval.

EFFECT OF TRAFFIC ON A SWITCHED ETHERNET LAN

When PTP messaging takes place on a busy switched Ethernet network, delays in packet processing at the NIC interface can appear as significant phase offsets between the GrandMaster and its slaves. In order to quantify this effect, we introduced a network traffic generator, a Mac OS X laptop, which generated traffic broadcasts of DNS, arp, and ICMP packets at a high rate. Figure 12 shows 1PPS phase offsets of up to 600 nanoseconds during periods of bursty traffic generation.

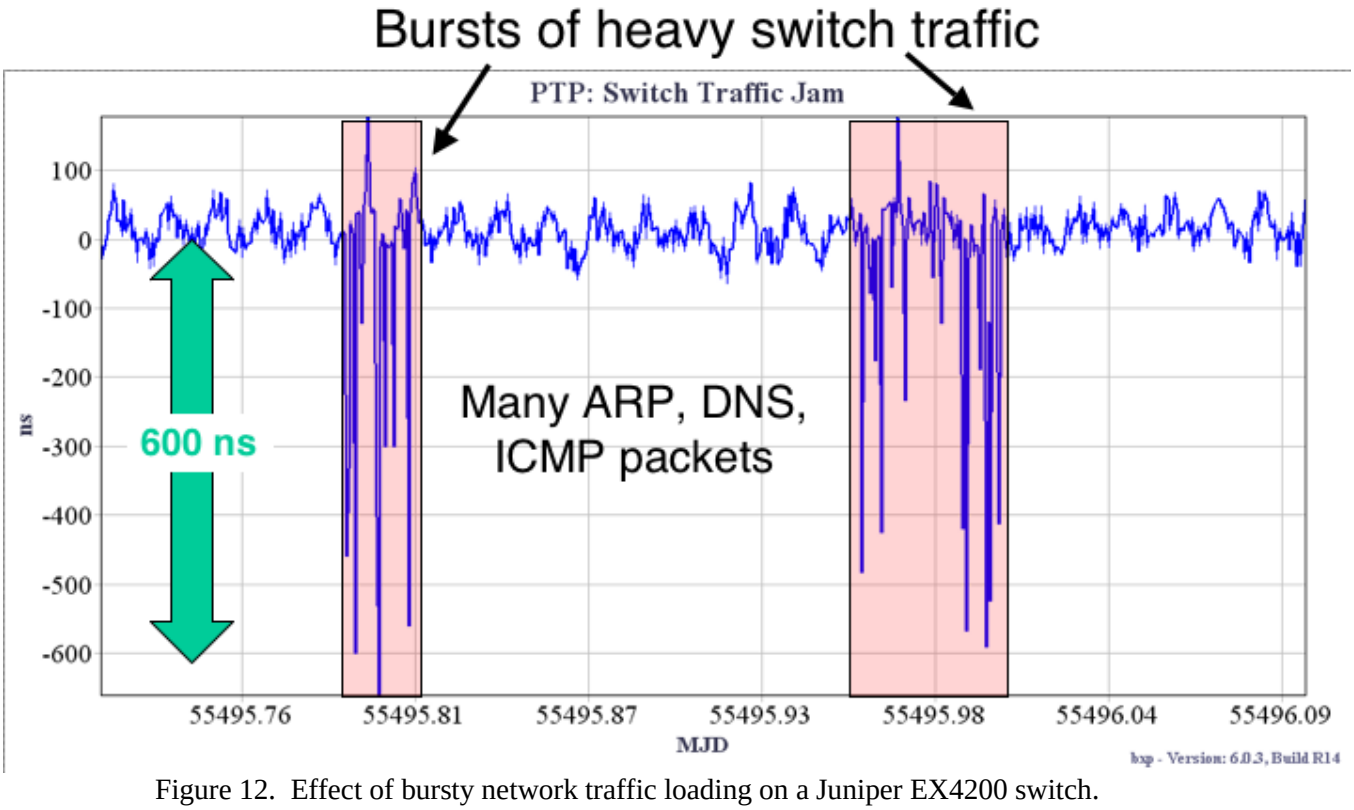

Figure 12. Effect of bursty network traffic loading on a Juniper EX4200 switch.

On an Ethernet switch, traffic can be segmented onto Layer 2 virtual LANs, or VLANs. We created a PTP VLAN on the switches which effectively isolated PTP from the network traffic generator. Figure 13 shows operation of the VLAN during a repeat of heavy traffic generation; the effect upon PTP is no longer perceptible.

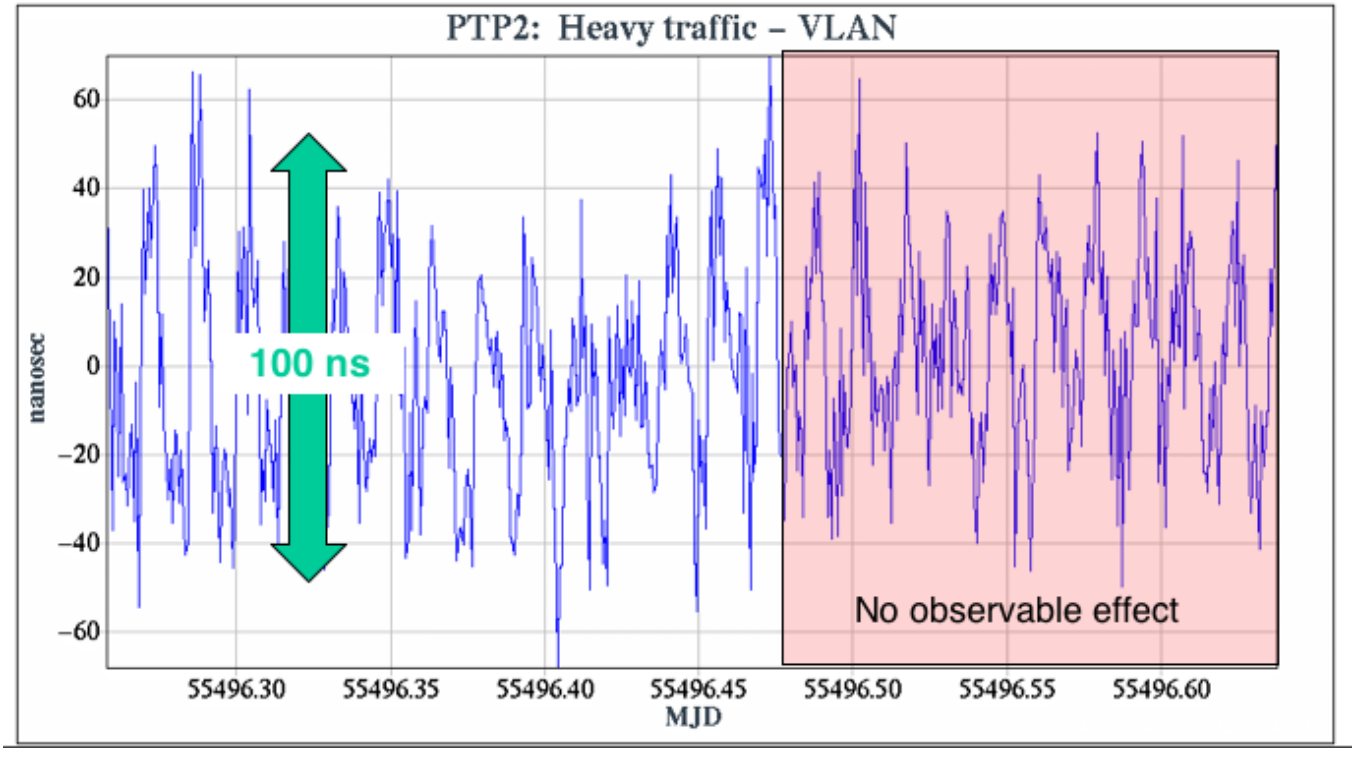

Figure 13. Switch segmentation using a VLAN isolates PTP traffic from other traffic.

NTP Reference Clocks and the System Clock

The PTP and IRIG-B synchronizable clocks on an NTP server's PCI bus can be made to function as NTP *stratum zero servers* with the addition of an NTP reference clock driver to the NTP software. Then it is possible to log the phase and frequency offsets of each with respect to the system clock (NTP peerstats). We configured an HP Proliant CentOS server with NTP drivers for the Oregano syn1588PCIe and the Symmetricom bc635PCIe synchronizable clocks, and logged their peerstats.

Figure 14 shows the NTP phase offsets of both bus clocks with respect to the system clock. Not surprisingly, the two synchronized PCI clocks are in phase and have large (±40 μsec) excursions from the system clock, which is responding to the 22-minute ambient temperature variations with the amplitude of a very temperature-sensitive crystal clock (shown in Figure 15 below). The typical $0.99 metal can oscillators have a frequency stability of ±100 ppm. As all of the NTP timestamps that our server will generate rely upon this poor crystal, which is our NTP flywheel, we regret that we could not replace it with an external, conditioned reference frequency.

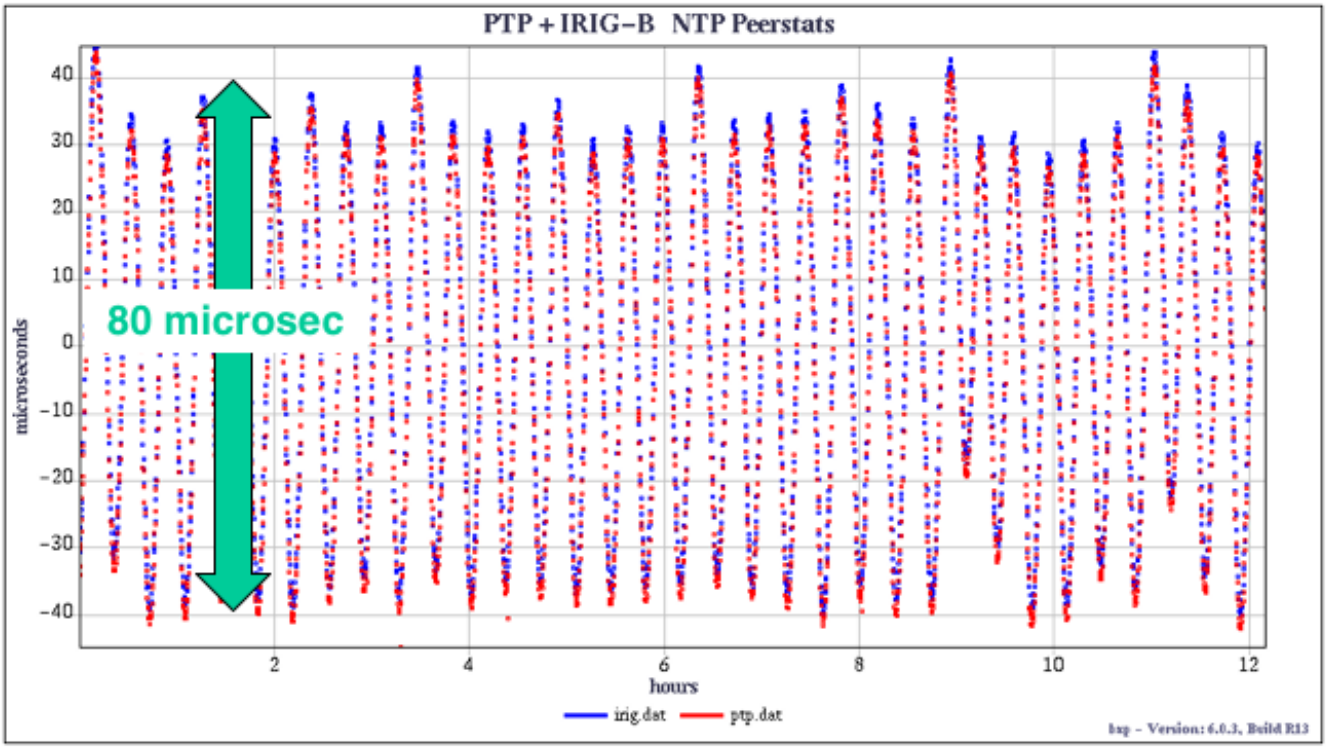

Figure 14. NTP peerstats phase offsets of the PCI clocks, with respect to the system clock.

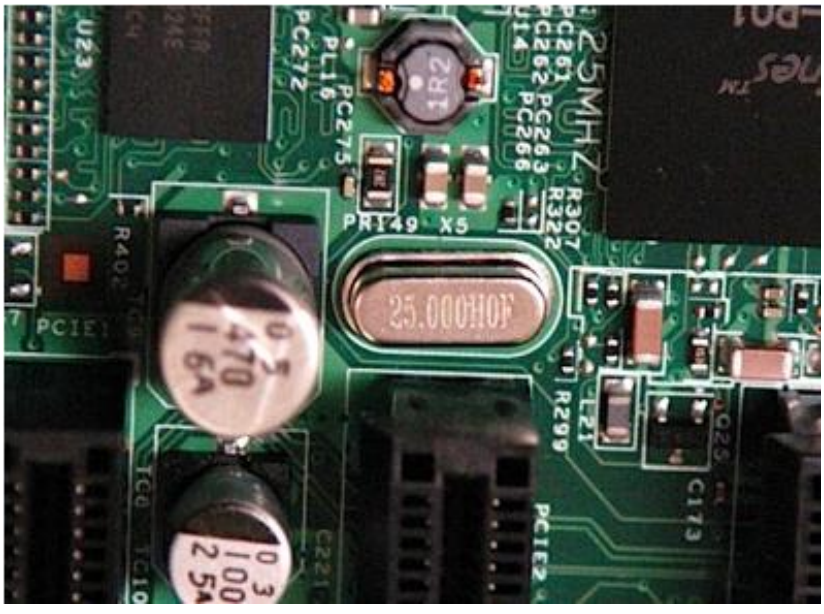

Figure 15. A metal can crystal oscillator in the Hewlett-Packard Proliant 120DL server.

Differencing the two PCI clock peerstat phase offsets, we see a mean of just 2.0 μsec ±0.7 μsec s.d. (Figure 16).

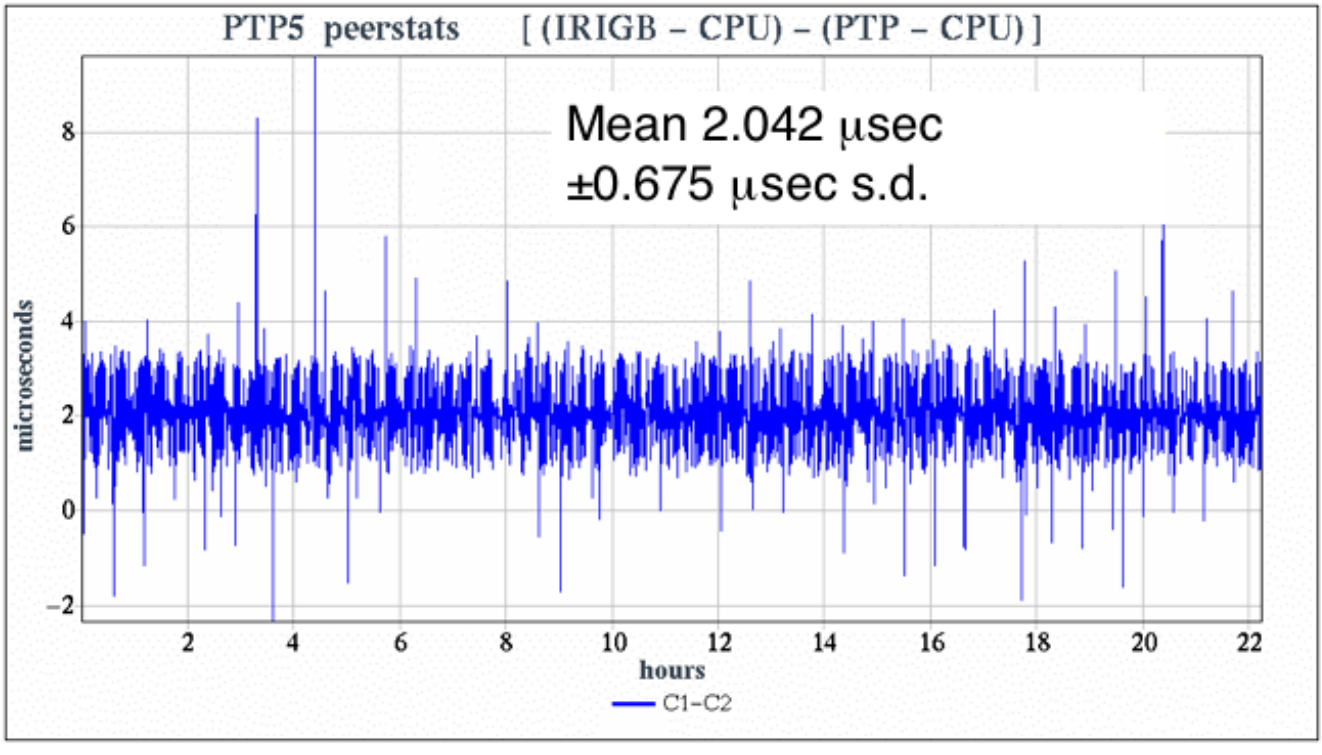

Figure 16. Differences of peerstats phase offsets, IRIG-B and PTP PCI synchronizable clocks.

Another interesting NTP measurement is the *loopstats frequency offset*, that is the current frequency offset of the system clock with respect to NTP. It represents the remaining error in the system clock frequency that NTP is striving to steer away (using the adjtime() system call, for example). In Figure 17, we show the remarkable NTP loopstats frequency offsets measured on two separate Proliant servers, "ptp2" and "ptp5", units ppm. Over 18 days, the loopstats frequency offsets are completely in sync, and have standard deviations of 0.06-0.07 ppm, quite respectable from an NTP point of view.

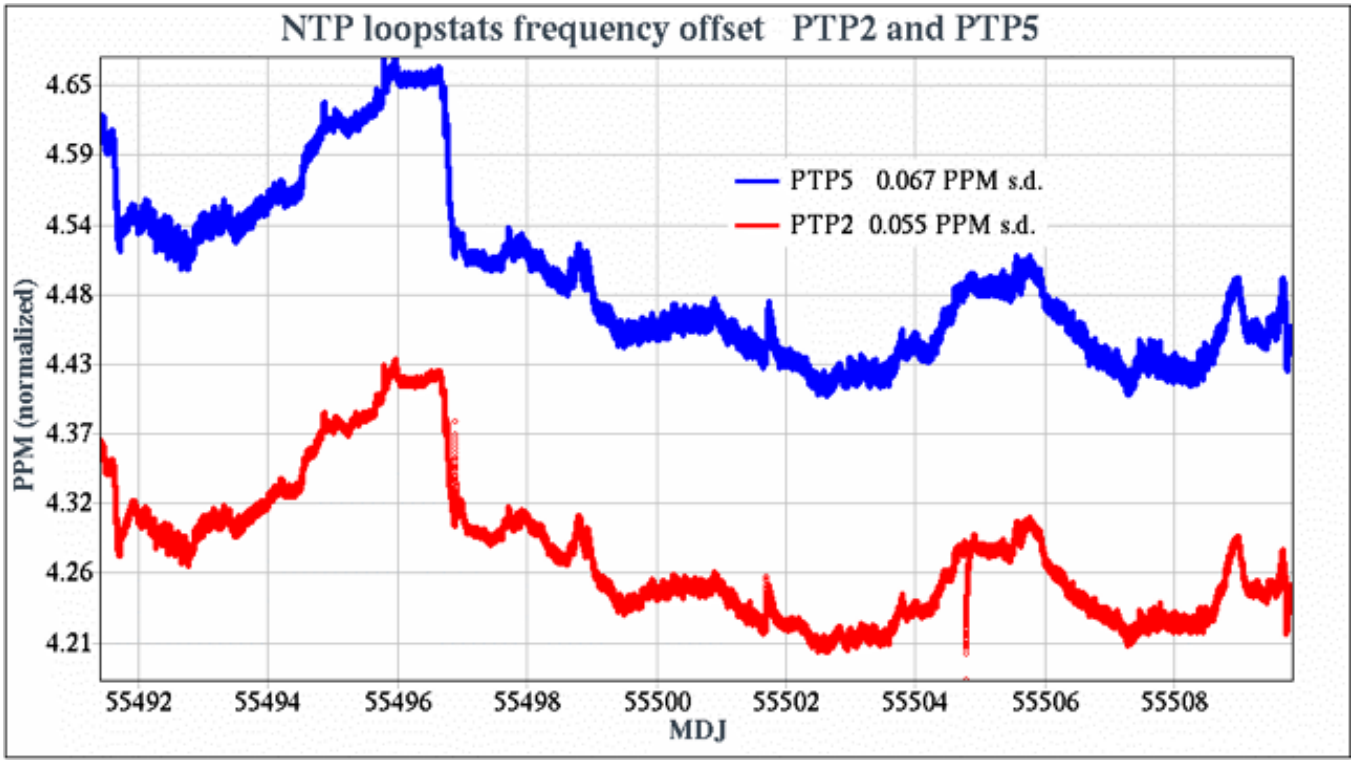

Figure 17. NTP loopstats frequency offsets of two NTP servers in the same computer rack.

GENERAL CONCLUSIONS

We conclude that IEEE-1588 synchronizable clocks in a PCI-express format provide a cost-effective timing reference for NTP stratum-1 servers. With use of Virtual LANs these servers can be dispersed across a campus network. We find that a modest increase in the PTP Sync packet rate can improve timing accuracy. We see measured improvement in timing accuracy with the use of OCXOs in place of ordinary crystal oscillators, and lament that PC manufacturers do not use them on current system boards.

Though particular vendor products are mentioned, neither official USNO endorsement nor recommendation of any product is herein implied.

ACKNOWLEGMENTS

The authors acknowledge the valuable assistance of John Cosand, *FEI-Zyfer*, Jeffrey McDonald, *Jtime! Meinberg USA*, Martin Schimandl, *Oregano*, and Aaron Feickert, Elizabeth Goldberg, Phillip Helfenbein, Warren Walls, Demtrios Matsakis, *U. S. Naval Observatory*.

REFERENCES

[1] R. E. Schmidt, 2005, *"Reflections on Ten Years of Network Time Service,"* in Proceedings of the 36th Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 7-9 December 2004, Washington, D.C., USA (U.S. Naval Observatory, Washington, D.C.), pp. 123-137.

[2] IEEE Instrumentation and Measurement Society, 2008, *"IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems"* (Institute of Electrical and Electronic Engineers, Piscataway, N.J.), *http://www.ieee.org*.

[3] FEI-Zyfer GPS Time and Frequency Systems, Garden Grove, Cal., *http://www.zyfer.com*.

[4] Hewlett-Packard Company, Palo Alto, Cal., *http://www.hp.com*.

[5] The CentOS Project, *http://www.centos.org*.

[6] The FreeBSD Project, *http://www.FreeBSD.org*.

[7] Oregano Systems Design & Consulting Ltd, Vienna, Austria, *http://www.oregano.at*.

[8] The NTP Public Services Project, *http://www.NTP.org*.

[9] Geeknet, Inc., Mountain View, Cal., *http://www.sourceforge.net*.

[10] Timing Committee, Telecommunications and Timing Group, Range Commanders Council, 1998, *"IRIG Serial Time Code Formats,"* U.S. Army White Sands Missile Range, N.M.

[11] J. Eidson, 2005, "*IEEE-1588 Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*," Tutorial on IEEE-1588, Agilent Technologies.

[12] KVG Quartz Crystal Technology GmbH, "*O.60.71992-LF Spec Sheet.*"