NAVSEA WARFARE CENTERS DAHLGREN

# Network Time Mechanisms for Improving Computer Clock Accuracy

**draft-marlow-tictoc-computer-clock-accuracy-00.txt**

## Brief to TICTOC Working Group at IETF 80

## March 31, 2011


**D. Marlow, S. Knickerbocker, and T. Plunkett**

**Naval Surface Warfare Center, Dahlgren Division**

## Introduction

- **New mechanisms beyond those identified in the Network Time Protocol Version 4 (NTPv4) Standard (RFC5905) provide increased time synchronization accuracy for operating system clocks' time and frequency**
  - **By providing improved estimates for when a packet is put on the network, transferred across a network, or taken from the network**
- **The internet draft identifies a set of candidate mechanisms for experimentation**
- **Standardization considerations are described for each mechanism**

# Purpose

- **Examine methods for improving NTPv4 time synchronization performance**
- **Solicit comments and contributions on the mechanisms described and on additional mechanisms that should be considered**
- **Discuss conclusions to motivate experimentation**
    - **Leading to standardization actions to enable better accuracy for future NTP specification**

# Motivation for Increased Performance

- Reasons to improve upon current NTP time synchronization performance:
  - Increased performance for existing product designs
  - New uses not currently available
    - Similar to how network speeds are increased every several years, and the uses for the increased network bandwidth soon follow
- Current methods to increase time synchronization performance involve
  - Use of a technology separate from the existing computer network (e.g., Inter-Range Instrumentation Group [IRIG] technology)
  - Use of Precision Time Protocol (PTP) technology as defined in the Institute of Electrical and Electronics Engineers (IEEE) 1588 standard
- With PTP, applications must interface with the installed PTP hardware in order to read time from its oscillator
- NTP resiliency does not exist in the PTP hardware oscillator
- Unknown results to the time provided by the PTP hardware when a network switch in the network path to the time source is temporarily unavailable
- Benefits of pairing NTP algorithm resiliency with the highly accurate PTP hardware-based time distribution

# NTP/PTP Commonalities and Differences

- **Commonalities**
  - Both are packet-based protocols for exchanging time with a time server over a computer network
  - Both determine the offset between two independent clocks
  - Both implement a hierarchical tree structure for obtaining time from a master time source through intermediary time sources to time clients
  - Both assume symmetric network paths
  - Both have unique methods for addressing asymmetric network delays
- **Differences**
  - PTP uses hardware to measure delays as packets traverse intermediate network devices and corrects its received time information based upon those measured delays
  - Because PTP has the ability to measure actual packet delays and to correct for them, PTP can provide the most accurate measurement of clock offset between two clocks
  - PTP does not define the method for synchronizing that clock once the highly accurate time measurements have been obtained
  - PTP is normally used to synchronize a hardware clock located on an interface card and does not synchronize the operating system clock
    - NTP possesses the ability to synchronize the operating system clock based upon received clock offset measurements
  - The NTP algorithms have a substantial resiliency so that operating system clocks remain stable despite the conditions on the network
  - NTP uses its algorithms to determine which of several consecutive time measurements are most accurate and uses that measurement

# Use Case Targeted

- A dense concentration of computing elements connected by a network
- A satellite-based time source (i.e., GPS) is used for synchronizing primary time servers
- Secondary time servers and leaf computing elements are synchronized to the primary time servers via the network
- In this use case, there are
  - Approximately 150 or so total computers where there are 3 to 4 levels of time servers
  - Time servers may have to communicate to each other through layer 2 and layer 3 network switches (could be 10 to 20 different layer 2 subnetworks)
  - All of the computers are connected together through gigabit or faster network connections
  - Some groups of computers will need to synchronize to each other to within a microsecond
  - Other groups of computers only have to be synchronized to each other to within a millisecond
- There is one interconnected time synchronization scheme where NTP, PTP, or a combination of both is used to meet all time synchronization needs

## Approach

- Determining what the current accuracy capabilities are with NTPv4
- Investigating additional mechanisms that may provide improvements in accuracy
  - Through experiments of those additional mechanisms
    - Estimations of improvements can be calculated
- Depending on the standardization difficulty and potential benefits offered, potentially recommending more than one standardization action

# Mechanisms Considered

- **NTP Interleave**
  - An extension of the NTPv4 protocol included in the current NTP distribution
  - Uses an IEEE 1588 PTP-like feature that provides a follow-up packet with a better estimate of when a previous NTP packet was sent on the network and a message exchange sequence to determine network mean path delay
  - Designed to be backward compatible (i.e., not affecting NTP implementations that do not use the interleave extensions)
  - Uses the same NTP packet format as the current standard NTPv4
- **Use of IEEE 1588 PTP and 802.1AS Mechanisms in the Underlying Network Service (e.g., Network Interface Controller [NIC])**
  - Determine if using special capabilities in the underlying network service can improve the timestamp estimates when NTP packets are put on the network, transferred across networks, or taken from the network
- **Use of IEEE 1588 PTP to Synchronize Computer Clocks**
  - Considers bringing the IEEE 1588 synchronization all the way to the computer clock through a standardized clock discipline algorithm
  - Computing elements synchronized by IEEE 1588 are candidates to be time servers (by the use of NTP) for computing elements not synchronized by IEEE 1588
  - Based on their respective strengths, the natural way to merge NTP and PTP would be to use PTP as the means of obtaining extremely accurate time information from across the network and to let the NTP algorithms use that time to keep local clocks synchronized

# Early Experimentation

- **Preliminary experiments tested the new Interleave mode available in NTP version 4.2.6.**
  - **The stratum 2 server and four of the six other workstations were upgraded to use NTP version 4.2.6**
  - **The remaining two workstations ran NTP version 4.2.2 included with RHEL 5**
  - **Interleave mode was achieved by using the "xleave" option when either the broadcast mode or peer mode was used**
  - **Offset measurements were obtained between the six clients and the stratum 2 server using the ntpdate command with the "-q" option**
  - **Measurements were taken every minute over a period of approximately 4 days**
  - **No discernable network load**
  - **All workstations were connected through the same Virtual Local Area Network (LAN) on the same network switch**
  - **All of the network connections were 100 Mbit/sec Ethernet**
- **Some results were obtained from the experiments where the average and the standard deviation of the absolute value of clock offset were measured**
  - **The worst behaved NTP Interleave client was able to stay synchronized with an average clock offset of 9 microseconds with a standard deviation of 8 microseconds**
  - **The worst behaved computer that synchronized using client/server mode was able to maintain an average clock offset of 11 microseconds with a standard deviation of 10 microseconds**
  - **The worst behaved broadband client (without NTP Interleave) stayed synchronized with an average clock offset of 49 microseconds and with a standard deviation of 58 microseconds**
- **These results illustrate that broadcast with NTP Interleave provides results that are better than having every client poll the server via unicast**
  - **However, the result is not significantly better (e.g., not an order of magnitude better)**
- **Preliminary experiments with hardware-based PTP have been performed in the past where the average offsets between PTP NICs and the PTP Grandmaster clock are in the hundreds of nanoseconds with standard deviations in the tens of nanoseconds**

# Future Experimentation

- **Further work is needed to ensure valid offset measurements**
  - One concern is that the clock offsets are in the microsecond range
  - The use of ntpdate may not be a valid way to accurately measure clock offsets at this level since ntpdate makes measurements across the network and is susceptible to errors caused by variations in network delay
  - An out-of-band measurement technique, which is not affected by variations in network delay, needs to be investigated for use in future experiments
- **Load applied to either the processor or the network carrying the NTP packets**
- **Perform future experiments to measure the resiliency of the NTP Interleave while under network and processor load**
  - Compare results to the other time synchronization methods
- **Perform experiments to determine how performance is affected when a more complex network configuration is used**
  - This experiment was performed in a configuration in which the test workstations were connected to the same network switch
- **Develop a standard to define time synchronization performance metrics that allows different experimental efforts be performed in a way that the results are comparable**

# Analysis of Results

- In 2008, the synchronization of computer clocks using NTPv4 over a LAN approached values on the order of ~145us (mean plus 3 sigma)
- Recently, using newer hardware and a newer version of NTP, time synchronization values were measured on the order of ~40us (mean plus 3 sigma)
  - The variation in the time synchronization comprises the majority of the 40us value
- If this 40us synchronization can be maintained from stratum to stratum for all subsequent tiers in a well-engineered network with 3 to 5 stratums, a maximum offset on the order of 120 to 200us could be achieved
  - A 50 percent improvement in clock synchronization variability from stratum to stratum would reduce this value to less than 125us
  - This does not imply that stratum-to-stratum accuracy should not be improved
  - This needs to be accomplished without overloading the computer or the network
- A 50 percent improvement appears to be a reasonable goal; however, if the stratum-to-stratum synchronization variability could be improved by an order of magnitude, it is reasonable to anticipate maximum theoretical time synchronization offsets of 50us or less in a stable LAN and potentially in the hundreds of microseconds for a Wide Area Network

IETF TICTOC Working Group discussions could define what level of improvement in accuracy warrants a standardization action to add an Interleave option to the NTP standard

# Conclusions

- Conclude Interleave capability can provide a modest improvement to the accuracy achieved with NTPv4
  - Recommend conducting further testing to determine whether this should be added as an option to the NTPv4 standard
- Recommend that the two mechanisms described in the draft be pursued in parallel since NTP Interleave cannot provide accuracies in the range that PTP with hardware assists can
- Recommend the TICTOC Working Group initiate work on a standards-track targeted working group Internet-Draft
- Recommend the TICTOC Working Group standardize the performance metrics used in describing the behavior of clock synchronization
  - Such a standard would enable better comparisons between the mechanisms considered
  - In addition to the same definitions of the metrics used, better agreement should be obtained in experiments being performed by different organizations
- Solicit contributions on the mechanisms described in this draft as well as on additional mechanisms that may improve the accuracy of computer clocks synchronized via a network