White Paper

Cisco IP Telephony Clock Synchronization: Best Practices

More and more enterprises are using IP telephony solutions based on Cisco Systems® networking solutions. But many customers are overlooking an important part of the system architecture: the need to synchronize the infrastructure design with an external network time server.

OVERVIEW

Correct clock synchronization is to good performance, effective troubleshooting, and accurate call detail records (CDRs) for billing.

Cisco® Unified CallManager systems and gateways provide multiple ways to synchronize time. Using lab test results, this document compares the advantages of the two primary options, explains how to mitigate risks, and describes the benefits of synchronization with an external network time server such as the Symmetricom Network Time Protocol (NTP) SyncServer S200. The test findings also apply to other IP telephony platforms, such as Cisco Conference Connection, Cisco Customer Response Solutions, Cisco Unity® Voice Connector, Cisco Personal Assistant, and Cisco IP Phone Productivity Services (PPS).

IMPORTANCE OF NETWORK TIME SYNCHRONIZATION

Improved Event Tracking with Accurate Log Files

Network time synchronization plays an important role in keeping IP telephony networks operational and performing well. Although most people reluctantly accept that a data network may occasionally fail, no one is prepared to lose a voice network for even a minute. Problems with IP telephony must be avoided or, in the worst case, minimized to keep business-critical voice systems operating.

Accurate server and router log files are essential to IP telephony reliability. Every log file entry is time-stamped to establish the time of events and facilitate the sequencing of events. Log file data and subsequent reports allow administrators to identify the root causes of problems in the network. Because server logs are a compilation of information from different hosts, as shown in Figure 1, it is essential that the time stamps be accurate within milliseconds. If they are not, sequencing events becomes problematic, troubleshooting root-cause problems becomes much more difficult, and keeping the IP telephony network operational becomes nearly impossible.

Figure 1.   High-Level IP Telephony Topology
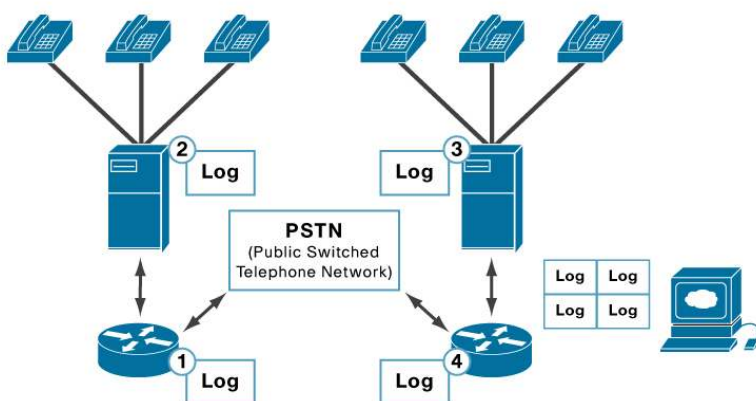
More Useful Traffic Reports with Accurate Metrics

In addition, comprehensive traffic reporting provides information about the grade of service, peak hours, and call volumes and assists in determining whether the network is over- or underused. These same metrics can be used to estimate agent staffing requirements, dimension trunk groups, and calculate IP telephony (IPT) bandwidth. Effective actions based on this information cannot be taken, however, if traffic analysis relies on faulty timing or if call rerouting algorithms use time as the trigger and that time is inaccurate.

Fewer Billing Errors with Accurate Call Detail Records

Network time synchronization also plays an important role in billing. Call Detail Records (CDRs) are the primary source of billing information in an IP telephony environment; they provide information about call origination, destination, and duration. CDR duration information includes the time stamp specifying when the call was initiated and either the length of the call or the time that the call was terminated. When calls cross multiple networks, gateways, and IP telephony servers, CDRs are created by combining data from the multiple call legs, as shown in Figure 2. The integrity of IP telephony billing relies on the accuracy of the time stamps, which are allocated in milliseconds. Without correct network time synchronization, accuracy will quickly decline, and the billing system will become questionable. Synchronization is particularly important when CDR information is shared among carriers and billing discrepancies require time-consuming mediation. Furthermore, many IP telephony networks now include unified messaging, video conferencing, bandwidth on demand, and other real-time provisioning services. Accurate time-stamping—based on network time synchronization—is essential to the billing schemes for all these technologies (refer to Symmetricom, "Synchronization Essentials of VoIP," p. 4, at http://www.symmttm.com/pdf/Network_Timing/wp_voip.pdf).

In addition, communication facilities are the lifeline of successful businesses and often are responsible for significant expenditures. But CDRs and communication facility use reports lose their validity if the timing on which they are based is inaccurate—consider such call center metrics as peak-hour traffic (PHT), busy-hour traffic (BHT), average handling time (AHT), and average speed of answer (ASA) to understand the important role that the time plays.

Moreover, the future realization of a single global CDR will be possible only by achieving the highest level of accuracy.

Figure 2.   Call Detail Record Collection
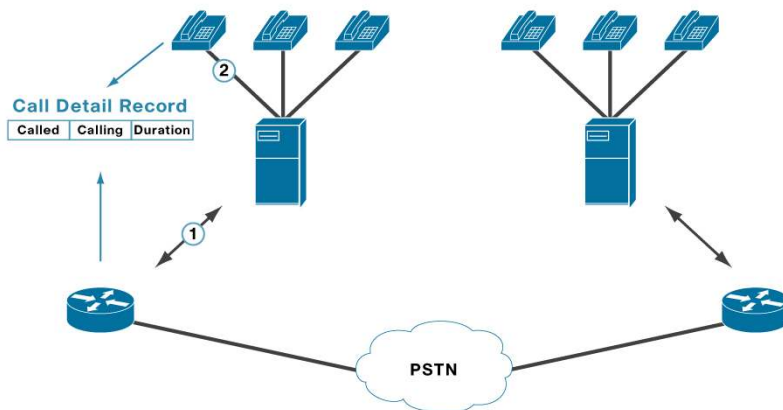
Enhanced Revenue with Accurate Timekeeping

The financial implications of accurate timing are especially significant for voice-over-IP (VoIP) services. Often the business model for IPT deployments relies on allocation of revenue, billing, forecasting, monitoring, costs, chargebacks, billing integration, invoice management, productivity, and communication management expenses to a host of other services such as unified messaging, video conferencing, bandwidth on demand, and other real-time provisioning services.

In the hospitality industry—hotels, motels, and resorts—hospital, and bill-back telephony environments, time equates to revenue and is a primary source of profit. Time is a common denominator in billing, in exception-management reports that address long call durations, in call cost and chargeback consolidations, in service comparisons, in management of network misuse and abuse, and in network management. All these tasks rely on accurate and synchronized network timekeeping.
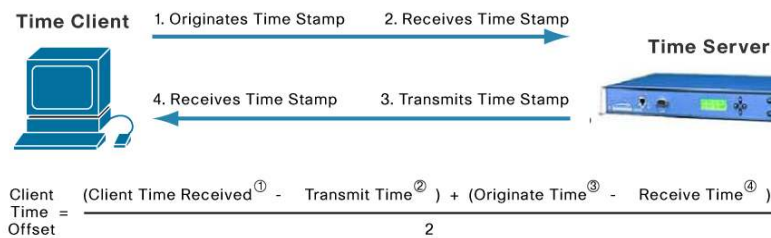
NETWORK TIME SYNCHRONIZATION OPTIONS

The goal of network time synchronization is to help ensure that all embedded system clocks in servers and networking equipment use accurate time. Time on these devices is kept by counting the cycles of the local oscillator in the equipment. Varying in quality, local oscillators usually run fast or slow (determined by measuring their drift from a reference point over time). During synchronization, the time on these clocks is adjusted back to the correct time. In Cisco Unified CallManager systems and gateways, time is synchronized by using one of two primary methods: Network Time Protocol (NTP) and Windows Time Service (W32Time).

NTP

NTP (RFC 1305) synchronizes servers and network devices using a reliable time source, such as a dedicated network time server that references the global positioning system (GPS). As shown in Figure 3, with NTP a client-initiated packet is time-stamped by the client and the time server; the result is that the client removes its time offset relative to the network time server. Many operating systems and network devices already incorporate support for NTP.

Figure 3. NTP High-Level Client-Server Communication

$$\text{Client Time Offset} = \frac{(\text{Client Time Received}[1] - \text{Transmit Time}[2]) + (\text{Originate Time}[3] - \text{Receive Time}[4])}{2}$$

(Assumes Symmetric Path Latency for Outbound and Return Paths)

One of the strengths of NTP is that it uses Coordinated Universal Time (UTC), which can be easily accessed through the GPS satellite system. Because UTC is the same worldwide, networks synchronized to UTC avoid interoperability problems with other networks. This synchronization is particularly important when administrators are troubleshooting IP telephony traffic and need to compare log files from various networks.

Reliability and accuracy are the primary advantages of the NTP approach to time distribution. NTP uses a stratum hierarchy (see Figure 4). With the time source defined as stratum 0 and the network time servers as stratum 1, servers and clients operate in strata 2, 3, and so on and link their clocks to the primary time source. Because accuracy declines a little in each successive stratum, servers and clients can access multiple sources over diverse network paths, providing redundancy and greater reliability. Complex algorithms allow each server and client to achieve greater accuracy by reducing jitter, rejecting information that varies too widely from that of other sources, and accounting for the drift rate of its own clock oscillator.

Figure 4. Hierarchical Clocking
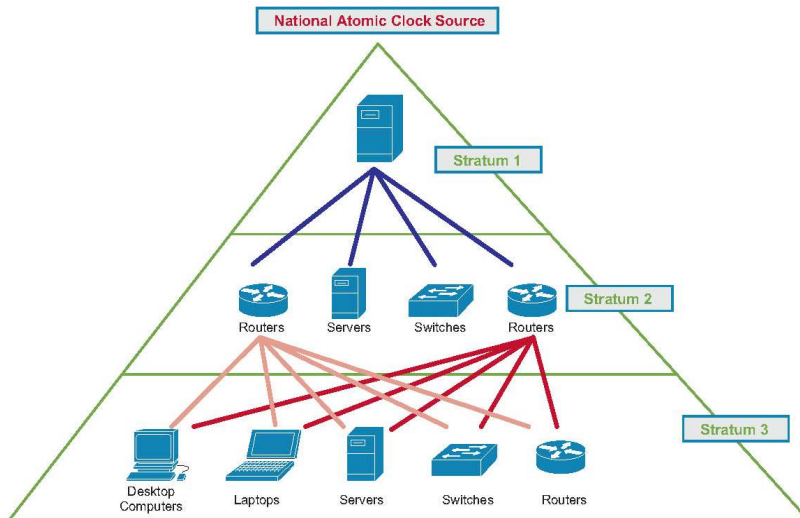
NTP provides extraordinary accuracy and ease of use in part because of its reliance on network time servers at stratum 1. Network time servers such as Symmetricom SyncServer S200 provide accurate, reliable, and secure time to the network and, with their slim rack-mount configuration, are cost effective, simple to use, and easy to install. With atomic clock accuracy from its embedded GPS receivers, a network time server can provide highly accurate time synchronization for thousands of clients on the network (refer to Symmetricom, "Synchronization Essentials of VoIP," pp. 3–4, at http://www.symmttm.com/pdf/Network_Timing/wp_voip.pdf).

W32Time

The second option for network time synchronization in Cisco Unified CallManager systems and gateways is W32Time, supplied with Microsoft Windows and based on the Simple Network Time Protocol (SNTP). The purpose of the W32Time protocol is to make sure that all computers in an organization running Windows 2000 or later use a common time. Time synchronization is important in a Windows 2000 environment because Windows 2000 implements the Kerberos Version 5 authentication protocol, a standards-based authentication protocol defined by RFC 1510.

W32Time works by periodically checking local time on a server or client with the current time on the time source, usually the authenticating domain controller. This process starts as soon as the service turns on during system startup. This protocol attempts synchronization every 45 minutes until the clocks have successfully synchronized three times. When the clocks are correctly synchronized, W32Time then synchronizes at 8-hour intervals, unless a time stamp cannot be obtained or a validation failure occurs. If a failure occurs, the process begins again.

By default, Windows-based computers use the following hierarchy, shown in Figure 5 (from Shala Brandolini and Darin Green, "Microsoft Windows 2000 Server, Windows Time Service," April 2001):

- All client desktop computers nominate the authenticating domain controller as their inbound time partner.
- All member servers follow the same process as client desktop computers.
- Domain controllers can nominate the primary domain controller (PDC) operations master as their inbound time partner, or they can use a parent domain controller based on stratum number.
- All PDC operations masters follow the hierarchy of domains when selecting their inbound time partner.

**Figure 5.**   W32Time Cascading Hierarchy
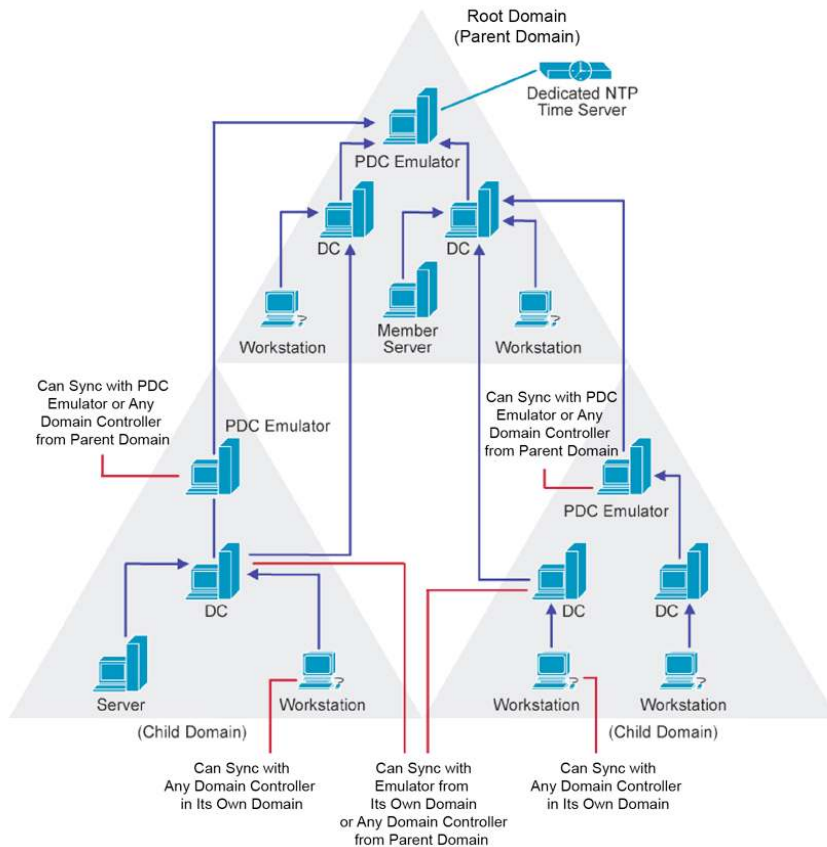
Following this hierarchy, the PDC operations master at the root becomes the authoritative time server for the organization. Microsoft highly recommends configuration of the authoritative time server to gather the time from a hardware source, such as the Symmetricom SyncServer S200. If the authoritative time server is configured to synchronize with an Internet time source, there is no authentication (refer to Microsoft, "How to Configure an Authoritative Time Server in Windows 2000," at http://support.microsoft.com/default.aspx?scid=kb;en-us;216734). This configuration is especially important in Cisco Unified CallManager systems—and arguably in all large-scale IP telephony deployments.

Running W32Time helps ensure that all computers within the enterprise reliably converge to a loosely synchronized common time. Although this loose synchronization satisfies the requirements specified by the Kerberos authentication protocol, W32Time is not designed for use by applications that require greater precision, such as today's emerging IP telephony environments (refer to Microsoft, "The Windows Time Service," at http://www.microsoft.com/windows2000/techinfo/howitworks/security/wintimeserv.asp).

NTP or W32Time: Which Is the Best Option?

The essential difference between NTP and W32Time can be seen in the results from a simple test jointly conducted by Cisco and Symmetricom. Figure 6 shows the accuracy of W32Time over an 18-hour period. W32Time updated the clock only about every 8 hours, allowing the time to accumulate almost 350 milliseconds of error.

Figure 6.　W32Time Offset Error

(~18 hours, ~4,400 samples)
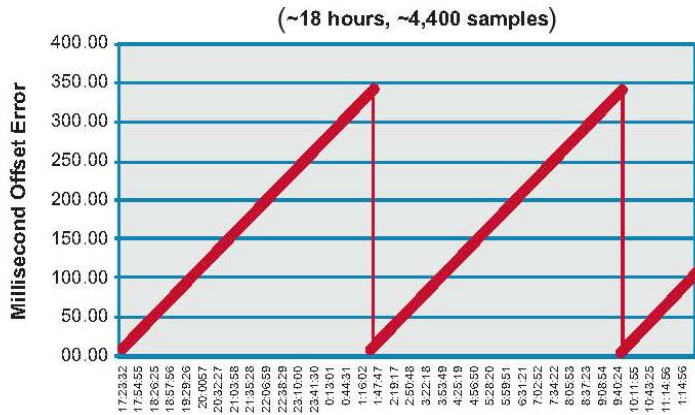
Because W32Time uses SNTP to synchronize the time only periodically, it provides a much greater chance for drift. With a fixed schedule of time updates, the accumulated offset error from server to server can be significant. If one server clock is fast and the other slow, clock differences in excess of 10 seconds over an 8-hour period will be likely. This kind of inaccuracy is the source of many log file inaccuracy problems and billing errors. Another type of troublesome clock error is the result of fast clocks that are updated infrequently. In the case illustrated in Figure 6, the clock was approximately 340 milliseconds fast prior to adjustment. Time stamps during the first 340-millisecond period will be interleaved in some way, or, worse, files may be overwritten during the repeated 340 milliseconds. This time inference can occur every 8 hours with W32Time.

In contrast to W32Time with its periodic synchronization technique, NTP adjusts time gradually and far more frequently. For example, when Cisco Unified CallManager is configured to use NTP, the NTP client tunes the local oscillator once every 8 to 16 seconds. The drift of the local clock is minimized to keep time constant, as shown in Figure 7.

Figure 7.　NTP Synchronization

(138 samples over 34 minutes)
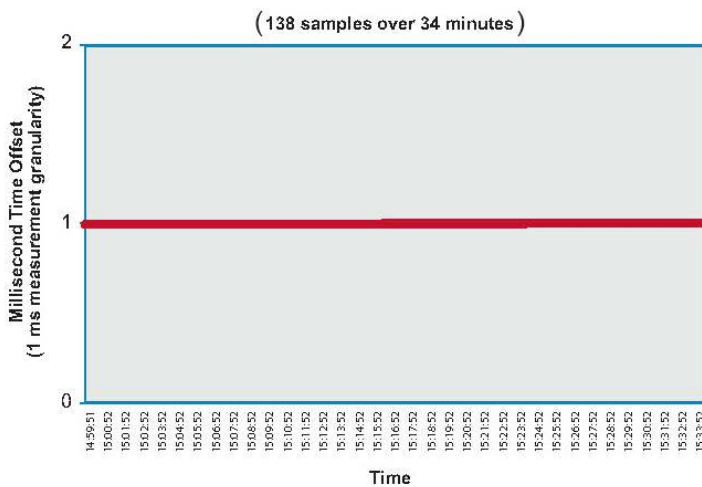
It is important to note that the results shown here do not account for network impairments of asymmetric latency (because of delays in the network from routers and path differences) or dropped packets. When asymmetric latency is added between the NTP client and the NTP server, the client clock offset corrections are not perfect, and jitter is added to the clock corrections. To mitigate network impairments, it is important to follow the best-practice guidelines that follow.

SUMMARY OF BEST PRACTICES

As Figures 6 and 7 show, NTP is superior to W32Time for network time synchronization in IP telephony networks. To build a good clocking scheme for large-scale IP telephony networks, administrators must use dedicated NTP servers and clients and minimize network impairments, thereby allowing clocking protocols to maintain a stable clocking scheme.

To provide a stable clocking environment, network administrators should do the following:

- Create a clocking architecture that defines the top-level clocking source and all the components in the downstream topology (refer to Figure 4).

- Establish the correct polling interval. The bandwidth of a typical NTP client-server exchange starts at 1 kilobit per minute, using the minimum polling interval. The minimum polling interval is 64 seconds, and the maximum is 1024 seconds. The recommendation is to follow the minimum polling interval.

- Cisco Unified Call Manager's NTP client has a hard-coded polling interval:
  - Cisco Unified CallManager Version 4.2 and earlier uses a polling interval of 64 seconds.
  - Cisco Unified CallManager Version 5.0 and later uses a polling interval of 1024 seconds (Note that even with this longer polling interval, tests have shown that Cisco Unified CallManager can maintain very accurate time.)

- Using the network clocking architecture outlined in Figure 4 as a reference, the following is recommended:
  - Use architecture in which all the distribution points use IP Unicast to talk to their upstream clock source.
  - At the distribution points, use IP Unicast with other protocols for security purposes and to minimize broadcast storms inside the LAN network.
  - Mark the NTP packets to flow in the network management quality-of-service (QoS) stream, with the recommendations listed in Table 1.
  - Identify the settings and equipment required to configure each component within the architecture.
  - Help ensure proper bandwidth and minimize network impairments by deploying appropriate QoS.

Table 1. QoS Marking Recommendations for NTP Packets

Clocking Protocol	Layer 2	Layer 3		
	LAN Class of Service (CoS)	IP Precedence	Differentiated Services Code Point (DSCP)	Per-Hop Behavior (PHB)
NTP	2	2	16	Class Selector 2 (CS2)

Configuring Cisco Unified CallManager for Accurate Timekeeping

Cisco Unified CallManager provides an NTP client on all servers. All subscribers should point to the publisher within the cluster. In turn, the publisher server should point to a dedicated NTP clock source. The following guidelines should help achieve accurate timekeeping within the Cisco Unified CallManager cluster:

- For Cisco Unified CallManager Versions 4.2 and earlier:
  - Use only the XNTP client for synchronizing the local IP telephony server oscillators; do not use W32Time for timing the cluster.
  - Manually change the settings in the ntpd.conf file to reflect the location where the NTP clients get their external timing. Within this file, change the server line of the configuration file after the server command to point to the IP address of the NTP server.
    - For the publisher, this NTP server should be a dedicated clock source.
    - For subscribers, this NTP server should be the publisher. This change must be made on all IP telephony servers.
  - The location of the ntpd.conf file is C:\WINNT\system32\drivers\etc.

- The following is a sample ntpd.conf file for a subscriber:

```
server 10.1.1.1     # IP Address or name of Primary Unified CallManager
driftfile C:\WINNT\system32\drivers\etc\ntp.drift # path for drift file
```

Note: After major operating system upgrades, verify that the ntpd.conf file is set correctly.

- For Cisco Unified CallManager Versions 5.0 and later:
  - Manual configuration of the ntpd.conf file is not allowed.
  - Cisco Unified CallManager prompts the administrator to specify the external NTP server at publisher installation and automatically populates the ntpd.conf file on the publisher server.
  - The ntpd.conf files of all other cluster servers are automatically configured to point to the publisher as the NTP source.
- Configure the publisher to point to a traceable stratum 1 clock source.
- Configure network latency between the publisher and the dedicated clock source so that it is within 40 milliseconds round-trip time (RTT).
- For additional information about NTP time synchronization for Cisco Unified CallManager, please refer to the Solution Reference Network Design (SRND) guide for your specific version at http://www.cisco.com/go/srnd.

Disabling W32Time

To disable W32Time on the local computer, you must be a local administrator on the PDC emulator. To perform this procedure on a remote computer, you must be a member of the Domain Admins group.

Follow these steps:

1. Open the Services snap-in.
2. Right-click Windows Time and choose Properties to open the Windows Time Properties dialog box.
3. In the Startup Type box, choose Disabled from the drop-down list.
4. Click OK.
5. Verify that the startup type for the time service is Disabled.

CISCO SYSTEMS