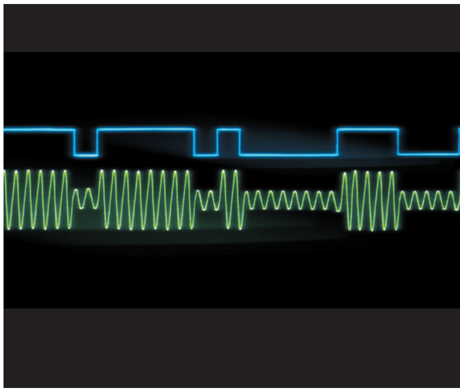
Symmetricom®

Network Time Synchronization:
5 Essential Elements



WHITE PAPER

Organizations today need network time synchronization that ensures the integrity of network operations and applications, yet needs little in the way of management overhead.

Computing today is a shared experience. Most computers are attached to the Internet and even when they are not shared, most of the things that are done on a computer — reports, spreadsheets, even calendars — are meant to be shared or are done with the understanding that they might be. Then there are all the applications that only work in a shared environment, including email, online financial apps, and transaction processing — not to mention network operations itself, which includes everything from directory services, to security, to fault diagnosis and auditing. Even before the Internet, many commercial applications depended on the smooth operation of networked computers: industrial process control, funds transfers, telecommunications, and EDI (electronic data interchange of business documents like purchase orders and invoices) to name just a few.

So what is it that people (or their computers) share? Two things: first, of course, they share whatever information needs to be sent, received, or processed. But the second thing they share is time. We may not all inhabit the same space. Indeed we may be continents apart. But what we do all share everywhere, all the time, is time itself. There is a well-known Alan Jackson song the title of which is: It's 5 o'clock somewhere. That is a non-technical way of implying that if you add the right offset, it is 5 o'clock here too. As a matter of fact, the "here" is usually much less important to users than the "now." Anyone who uses the Internet to make a credit card payment cares much less about the physical location of the bank's server than whether the transfer occurred before the bank's deadline (and has proof it did).

To create a shared experience, the network must provide shared time, just like it provides a sharing of information. "Now" must be the same everywhere on the network. Specifically, what is shared is a reference point called UTC (Coordinated Universal Time). UTC gives applications everywhere a common index with which to synchronize events and prove that events happened when timestamps say they happened.

Sharing time — as opposed to timesharing — means applications can operate completely independent of each other yet remain completely synchronized because they are all synchronized to UTC. The key is distributing UTC to those applications and proving that you did. You also want to do both (i.e., distribute UTC and prove you did) in a way that is easy to implement and manage. Timekeeping is not something that most computer users, and certainly most network administrators, want to spend time thinking about. Nor will they have to — provided that network administrators recognize the importance of network time synchronization and apply its five essential elements.

The Importance of Network Time Synchronization

To understand the ingredients of network time synchronization, it is first necessary to understand why synchronization is important. In other words, what are the needs of the organization that network time synchronization fulfills? Once you know which needs apply to your particular situation you can better assess your technical requirements.

In general, network time synchronization fulfills two missions: 1) It allows events to occur at the proper time (i.e., event synchronization); and 2) It provides proof when events occurred or did not occur (i.e., computer forensics). The first mission occurs during the fact; the second mission occurs after the fact. Table 1 summarizes some typical examples:

Mission Type	Employed	Artifact(s)	Purpose	Example Apps
Event Synchronization	During events	Application messages, flags	To ensure events occur on time, in correct sequence	Transaction processing, process control, authentication
Computer Forensics	After events	Time stamps	To determine when events occurred and in what sequence	Digital signatures, Crime investigation, Fault diagnosis

Table 1: Why Organizations Employ Network Time Synchronization

Organizations employ event synchronization to accomplish one or both of the following objectives: 1) To schedule a process — i.e., to ensure that it starts or stops on time or runs for a specified period regardless of when it starts or stops; and 2) To ensure that cooperating processes can interoperate correctly — so that if one process hands a task off to another process, the second process will in fact be ready to accept the handoff.

Examples of both scenarios occur frequently in industries like pharmaceuticals where a batch of materials, let's say, is supposed to be heated to a certain temperature for a certain period of time and then passed to another process for mixing. If the timing of these events is not exact, the batch may be ruined. Furthermore, if one process begins before another has finished, then the actual production equipment — not to mention the product itself — could be damaged. Similar scenarios happen all the time in industries like manufacturing, logistics, and electric power.

In the power industry, generators thousands of miles apart must constantly stay in phase at 60 cycles per second. If generators operate at different points in the phase cycle, they will actually work against each other — not only reducing power output but also potentially destroying the generators themselves. In addition, there are literally millions of electric clocks around the world that use the 60 Hz cycle (or 50 Hz, depending on which country you are in) to maintain the correct time. During periods of high load this cycle tends to slow down and, alternatively, to speed up during periods of light load. Utilities in turn rely on accurate network time synchronization to ensure that they do in fact provide the correct phase cycle their customers expect.

Financial services are another good example. Take the networks over which traders buy and sell securities. These networks must execute trades within a very small, and also extremely accurate, window of time (three seconds, in the case of NASD). The reason for this accuracy is obvious: prices change rapidly in financial markets and traders want to make sure that deals happen when the posted prices are still in effect. They also want to know (and regulators want to know as well) a particular trader's holdings in an investment at a particular point in time. So not only must trades actually occur on schedule, they must be accurately time stamped to reflect that they were.

Timestamps on financial documents are a time-honored way of proving that money changed hands, documents were signed, letters were posted, and other business events occurred when they were alleged. In the digital realm, this type of after-the-fact time lining is called computer forensics — the second category of network time synchronization applications.

Computer forensics has received considerable publicity lately — much of it due to high-profile Wall Street indictments. Many of these were made possible because of paper trails established in part through the use of timestamps on email and other electronic documents. Recent legislation such as Sarbanes Oxley and HIPAA (Health Insurance Portability and Accountability Act) mandate how organizations maintain records. With respect to digital documents, they expressly require the use of verifiable timestamps to prove the state and control of documents at particular points — like when they were created or subsequently altered and by whom. Officers in companies covered by these laws risk jail time and heavy fines if their organizations do not comply.

Timestamps, of course, are valuable evidence even apart from criminal investigations. Most organizations have a business need to trace the chain of events that led to key events, such as an important decision or an action by an employee. Then there are a host of technical reasons companies employ timestamps. Diagnosing a computer problem, for example, often means going back over log files to trace back the series of events that led up to a fault. Without demonstrably accurate timestamps, it is hard to see whether an issue was the cause or the consequence of some other issue.

Again, the electric power industry provides a good example. As previously mentioned, operators have to phase synchronize generators running hundreds of miles apart. In addition, they also must monitor the power grid for events such as voltage spikes or equipment outages using systems known as SCADA (supervisory control and data acquisition). SCADA systems are heavy users of time synchronization technology — specifically GPS timeservers — and timestamps are a major reason why. Timestamps from these servers provide the critical evidence needed to identify the causes of power failures, such as the one that crippled the Northeastern US and Ontario in August 2003. As it was, backtracking events to the power blackout's root causes took weeks and could have been done much faster had timestamps been more accurate and more widely distributed throughout the grid. They weren't, so investigators often had to cross check one timestamp against another in order to recreate the actual time index over which events took place.

In pharmaceuticals, the issue of time stamping is a regulatory imperative under FDA regulation 21 CFR Part11 and others. Companies have to not only synchronize production steps according to strict recipes; they of course have to also document the fact that these precisely synchronized actions did in fact happen as scheduled. Without accurate and pervasive time stamping, pharmaceutical companies risk severe penalties plus the possibility they might be able to trace back a problem if something did in fact go wrong during the manufacture of one of their products.

The Five Elements

In fact, the same is true for most organizations — accurate time needs to be pervasive on the network. It's hard to imagine a company that would knowingly tolerate only some of its processes running on time or only some of its timestamps being accurate — especially if they realize how easy and inexpensive good network timekeeping can be to accomplish. To make time both pervasive and accurate, five essential elements must be present:

- An accurate and reliable time source
- A timekeeping architecture that fits the organization
- Robust server management
- Robust network time management
- A secure, verifiable audit trail

An Accurate Time Source

Ultimately, accurate time must come from somewhere. How accurate that time needs to be depends on the applications and operations performed. Most network operations (e.g., online security, log file updates) require accuracy on the order of 1 to 10 miliseconds. Depending on the specific application, electric utilities may require time measured in microseconds. Most financial and general business applications require accuracy in the 100 millisecond to 10 second range — even if only to accurately establish the order of events.

Even if set to an absolutely accurate time reference, a PC clock may still be off by 50 milliseconds at the very the instant when it is set. Then, once set, the clocks in computers will start to drift, some by as much as several minutes a day. It is possible for a workstation to achieve consistent accuracy of half-a-millisecond, but only if its clock is reset repeatedly over the course of a day. The challenge is to reset the clock before it drifts too far, and to do so using a time source that is itself accurate.

Most organizations, unfortunately, are

unaware of the importance of maintaining accurate time on their networks. Those that are aware will typically acquire time in one of three ways: 1) over the Internet from the National Institute of Standards and Technology, NIST, or a third-party time service; 2) from GPS satellites; or 3) over a dial-up connection from NIST. If over the Internet, organizations almost always use NTP (Network Time Protocol), an internationally recognized protocol for synchronizing the clock on client machines with clocks on network timeservers. NTP is available on virtually all computing platforms — either as a built-in service of the operating system (Unix, Mac OS) or as widely available client software (Windows).

When evaluating the time source, organizations must also take into account the asymmetric path delays between the time clients and the time server as well as the security of the time source. Of the three methods listed above, GPS is the only one that offers a direct, accurate and secure connection from UTC to inside the security of the organization's network firewall. There are no intervening WAN infrastructures or routing tables to cause uneven delays between the client requests and server responses — as can happen with internet based NTP time sources.

On a WAN, NTP client time accuracy can be as good as 10 to 50 milliseconds on average, individual time corrections can frequently vary by quite a bit more. However, if UTC is provided via GPS to the LAN, NTP can usually distribute UTC locally with an accuracy of from one to 10 milliseconds on the client side. (GPS time stamp accuracy inside the timeserver is typically about 1 microsecond to UTC). That means for most organizations, and for most applications, a GPS referenced timeserver is sufficient to deliver time to the local net and distribute time to client machines once it is there.

There are also other reasons to use GPS besides just performance. GPS avoids the need for a dial-up connection, a potential security most IT managers would like to avoid. It also avoids keeping open the networking port (#123), which NTP uses, and is therefore a potential point of entry for an intruder.

A Timekeeping Architecture that Fits

Acquiring UTC from GPS requires taking the signal off the air and delivering it to the clients (PCs, workstations, servers, controllers, etc.) that rely on accurate time for event synchronization and time stamping. Since few clients come equipped with reliable network time synchronization software, this means that besides the GPS receiver itself there also has to be a way to distribute time from the GPS receiver to the clients. In other words, the organization needs a time distribution network - typically an architecture of deployed timeservers and time clients.

Why an architecture? Because timeservers acquire time from GPS receivers and distribute time in response to client requests. Depending on the size and topology of your network, you may want to install multiple timeservers and do so in a certain configuration:

- To support multiple LANs
- As a backup
- To handle peak load volumes of client requests
- To handle special time-sensitive applications

As just noted, distributing time via a WAN introduces delays that can be avoided when each LAN takes UTC from GPS. But there are also good reasons to have multiple timeservers at each LAN — as a backup, for example, in case one of the timeservers goes down or becomes overloaded if there is a spike in requests. In these scenarios, clients might have a primary and a secondary (or even tertiary) timeserver to reference in case of a problem. Also, one timeserver might itself be set up to receive time from another timeserver (say, if its connection to GPS were somehow lost). The purpose of these types of architectures is to make the network self-healing with respect to timekeeping.

The choice of GPS receiver, the choice of network timeservers, and the architecture of the network are key. Here are factors to consider in making those choices:

Server performance — While the ability to synchronize tens of thousands of clients is catchy marketing, the real test is the volume of peak load requests the server can handle simultaneously — while maintaining accuracy and availability.

100Base-T Ethernet — A timeserver that supports both 10Base-T and 100Base-T will accommodate networks today and when those networks and after those networks are upgraded in the future.

Redundant time sources — A timeserver may employ various strategies to ensure continuous UTC availability — such as multiple GPS receivers or a GPS receiver with NIST dialup as a backup.

Single satellite timing — In urban canyon environments where satellite visibility can be limited or when roof access is restricted, an automatic single satellite timing mode provides accurate time with intermittent satellite coverage. It can also track satellites using a window-mounted antenna.

Built-in time reference — When access to UTC is interrupted, a network should be able to maintain the required timing accuracy for a period sufficient to enable continuity of the supported applications and business mission. For example, Rubidium time references can be employed in those timeservers that acquire time from GPS and which redistribute the time to other timeservers or to clients.

Once the network hierarchy is in place, the next issue is how to manage it. As previously noted, time is not something most organizations want to think about — so management of network time distribution should be simple and straightforward. Management occurs on three levels: 1) monitoring and controlling the network devices (e.g., setup and configuration); 2) comprehensive management of clients and devices as a network; and 3) providing a verifiable audit trail of the time synchronization across that network.

Robust Server Management

How should an organization control a network timeserver? The answer probably is: Any way it wants. Some companies, for example, might prefer to access devices via a web interface; some through a comprehensive network management suite, such as OpenView, while others might rather just work with the device using a built-in keypad. Or the same organization may, from time to time, employ more than one of these methods, depending on the specific situation. For example, working with a keypad probably makes the most sense when initially setting up the device, while ongoing monitoring and control might best be done from a central location via the network. A web or telnet interface offers a light client "footprint" and easy accessibility from anyplace there is a networked PC. The virtues of network management solution are that it allows the operator to monitor all network devices from a single console. In that scenario, the device must offer SNMP (Simple Network Management Protocol) support so managers can monitor the device via any management software that also supports that protocol. They can then control the device remotely via the web or telnet interface, or work with the device locally using its keypad.

Robust Network Time Management

Managing a network of devices as a whole is different from managing particular devices. The availability and reliability of accurate time across the entire network — not just a part of it — must be assured. Network timeservers must work in synch, their operation should be centrally controlled, the processing load should be balanced, and enterprise policies on issues like security should be enforced. Nor should providing network-level management be so burdensome to network administrators that it interferes with the other priorities they have to attend to.

The network management would typically run on a central workstation, from which it would connect securely to a trusted network time source (such as a dedicated GPS referenced timeserver) and to clients. It then distributes that time accurately and verifiably to every time-aware machine on the network. All timeservers and clients should be individually identified using a unique serial number assigned when the management layer software is installed on the workstation. Administrators at this workstation are then able to perform the following activities:

- Perform core setup functions of all timeservers:
    - Installation
    - Monitoring
    - Configuration
    - Troubleshoot device-level problems
- Monitor time synchronization enterprise-wide (i.e., are all clients synchronized within a specified threshold)
- Implement network-wide timekeeping mitigation strategies
    - Time-source averaging
    - Clock training
    - Skewing (offsetting clocks by a constant value)
    - Target seeking (skew clocks by a value needed to achieve a certain accuracy)
- Automatic failover of servers
- Multiple levels of fallback time sources
- Safeguard against malicious or inadvertent tampering with network timekeeping

The ideal result: a reliable time synchronization system that requires little management overhead and offers tremendous value to the integrity of network operations and applications. The network management layer is also the foundation for the last piece on top of the network timekeeping infrastructure, discussed next.

A Secure, Verifiable Audit Trail

Lastly, but certainly not least, a time synchronization infrastructure requires an audit capability. The whole point of a timekeeping infrastructure is to provide assurance that events happen on time and that the actual time of events can be verified. The audit trail is that assurance.

This capability would typically take the form of a dedicated audit server — which makes sense given the fact that good auditors usually stand apart from the entities that they audit. That ensures integrity of the process — because time can be verified independently of the clocks subject to the audit — and because the process can be better isolated from security threats. That doesn't mean that an audit server should require the installation of a redundant management layer (the focus of the preceding section). On the contrary, the audit server should be able to leverage the capabilities of an existing management layer.

In general, the function of the audit server should be to prove conclusively (and on demand) whether the time on any monitored system was correctly synchronized at a particular time and date with a specified time source.

In general four requirements must be met for successful time auditing:

- Monitored machines must be able to be reliably and individually identified
- Time on individual machines must be synchronized regularly and accurately with a known time source
- Vital information must be easily retrievable, such as when the local clock was last adjusted and with what time source
- Sync information must be collected and regularly and complied into concise and complete audit records
- Immediate email alerts must be generated when any monitored machine fails to be synchronized with desired tolerances or if a machine misses more audits than a specified threshold

It is this type of audit capability that is typically required by federal regulations (e.g., FDA, Sarbanes-Oxley, HIPAA) as well as by major securities organizations like NASD to prevent fraud and establish the validity of transactions. In the case of event synchronization (as opposed to timestamp applications) an audit server enables compliance with ISO 9000 and cGMP (current Good Manufacturing Requirements).

5 Elements, 1 Objective

The value of networks is that everything, virtually speaking, can happen in the same place. There is only one incontrovertible physical reference point, which is time. Where things happen is less important than when they happen. Universal time (UTC) establishes the sequence of events in this shared space — either prospectively to enable cause and effect — or retrospectively to prove cause and effect. In either case, users and applications need to assume that their version of "right now" agrees with everyone else's. That's a simple objective — and it should be equally simple to achieve.

Timekeeping is not something people — in particular, network administrators — want to spend a lot of time thinking about. If they start with a simple timekeeping infrastructure — based on the five elements presented here — they won't have to. And neither will their users.

SYMMETRICOM, INC.
2300 Orchard Parkway
San Jose, California
95131-1017
tel: 408.433.0910
fax: 408.428.7896
info@symmetricom.com
www.symmetricom.com