Symmetricom

NGN SERIES

How NTP Impacts the New Telecom

No longer confined to the data center, Network Time Protocol is an emerging force in the delivery of new digital services and the reliable operation of telecom networks.

Table of Contents

In the last few years, telecom networks have undergone a substantial change — both in terms of technology as well as in the services they support. In particular, packet-based technologies have placed new demands on network-wide timing. This paper will discuss how these changes are creating new requirements in the accuracy, security, and availability of NTP-based time. Network operators will have to adjust.

NTP history and background

Network Time Protocol (NTP) is an Internet protocol for synchronizing clocks of computers and other equipment to a common time reference over a network. It is also a program (NTP daemon with utilities) that implements the protocol and controls the computer clocks.[1] As originally conceived, NTP was designed to provide time to computer hosts. Over time, its use grew so that today it is the universally accepted method to synchronize system clocks in computers, servers, and data communication equipment.

Historically, NTP has been deployed in the Internet, and then in telecom IT departments or datacenters for post processing functions to support operational activities such as billing, AAA (access, authentication and accounting), and event log generation. Today, many new needs for NTP have emerged, from both a services and operations perspectives. Many of these needs require more accurate and assured time than current NTP servers can provide. Moreover, some datacenter functions are moving into field offices. That means the systems supporting these functions no longer reside on the same LAN as their NTP source — so that source is further away. These trends have significant implications on both the services and operations side.

Service Drivers of NTP

Many new services — like IPTV, VoIP, wireless content downloads and multi-player gaming — involve real-time delivery of multimedia. Not only do these services require more accurate time than networks previously needed to provide, they also need time delivered in more places and more often. Real-time services with high QoS expectations require real-time monitoring and measurements at many points in the network — not just at a few as in the past — right to the customer premise and end-user device. Moreover, these services typically employ multiple systems to complete service requests, leading to an explosion in the number of systems that need NTP services. Let's consider two services in particular: IPTV and VoIP.

[1] IETF RFC 1305

IPTV

IPTV or Internet Protocol TV is the end-to-end delivery of video and related content using Internet Protocol over a managed network to consumers (i.e., not over the Internet). NTP based time is used for many critical processes in IPTV service delivery, such as:

- Conditional Access (CA)
- QoS
- Execution of channel change requests
- RTCP feedback
- Transaction logs and measurements

To support these functions, the IPTV service delivery infrastructure — and therefore the time synchronization that supports it — is distributed across a hierarchy of locations that includes the video headend office, the video service offices, the access network, and the customer premises. Consider Conditional Access (CA) — a very important function in IPTV service delivery. Conditional Access (CA) licenses provide the basis for content usage and associated revenue generation. They also prevent content misuse and piracy that can steal revenue and lead to content use agreement violations.

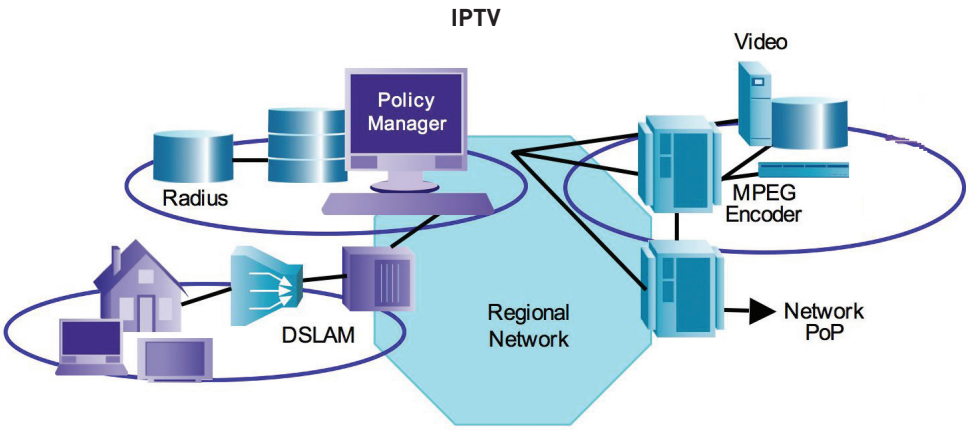

FIG 1 Evolution to a Highly Distributed Service Infrastructure

All these references must be synchronized so that QoS measurements — and the analytics those measurements support — will be relevant.

Time on the systems that enforce CA licenses needs to be accurate and secure — for example, to ensure that operators can only show content during specific periods (such as after a DVD's release date). CA license keys also need secure time to prevent theft of service and alteration of the license usage duration. The ATIS IPTV Interoperability Forum (IIF) recently published a specification for CA that calls for secure time. Many vendors have already begun implementations of secure time for CA. However, currently deployed NTP servers may not meet the new requirements or scale to handle the large transaction volumes generated by thousands, and potentially millions, of IPTV subscribers.

QoS measurement is another key function in end-to-end IPTV service delivery. Some unique aspects of QoS measurement for IPTV are the number of measurement points in the network, the need for real-time monitoring, and the collection of metrics inside the customer premise at both the residential gateway and the set top box. While the core and metro networks are tightly controlled to ensure network performance, the last mile is more problematic — and therefore poses a greater QoS challenge.

Just as to ensure CA, accurate and distributed time references are also needed to enable QoS measurements across hops in the network, or between a group of systems or customer devices. All these references must be synchronized so that QoS measurements — and the analytics those measurements support — will be relevant.

By far, most of the time references that need to be synchronized are those serving customer endpoints — the population of which is many hundreds of times larger than the number of core network elements. Providing NTP at all these points calls for consideration of NTP server transaction capacity, bandwidth overhead, security, authentication, consistency, and network delay. Operators may also wish to source NTP services closer to customer premise equipment (such as at the central office). This avoids relying on a centralized NTP that's distributed from the headend and the central datacenter over a wide area network (and avoids the greater latency and jitter that results and can lead to synchronization inaccuracies at the edge of the network). It also mitigates the impact of a loss of NTP service by reducing the number of end-points tied to a particular NTP source.

Voice over IP

Time synchronization also plays a key role in VoIP or Internet telephony services in several areas, such as:

- SLA measurements
- Fault analysis
- CDR generation and billing
- Security
- E911

The service level agreement (SLA) is a contract between a service provider and a customer that guarantees a certain quality and availability. SLAs are therefore a critical part of any VoIP service. Recently, service providers have started providing end-to-end SLAs guaranteeing VoIP QoS. These SLAs are being offered for IP PBX and Centrex services and include collection of data from the customer premise. Until now, most SLAs for enterprise-managed services have only been guaranteed between PoPs — not to the customer premises, as they are now.

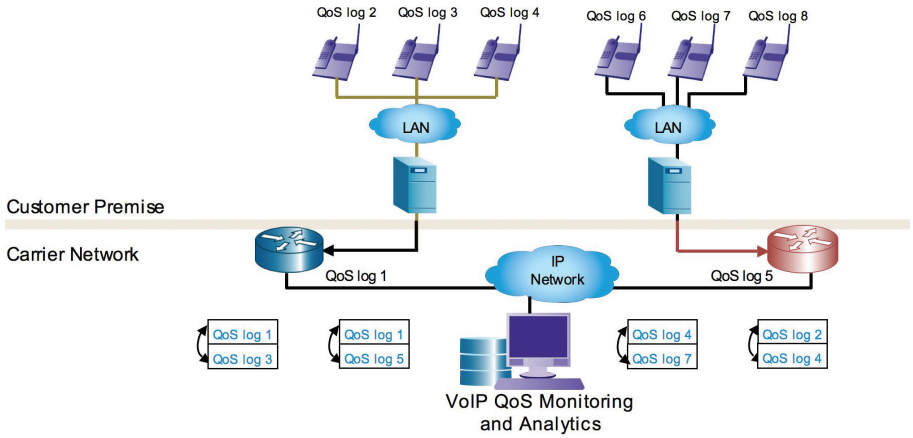

FIG 2 NTP Application Example: VoIP/IP Telephony

Both active and passive measurement techniques are used to measure VoIP QoS. Since active measurement consumes bandwidth, such measurements are taken at periodic intervals — typically every 15 minutes. Active testing offers the benefit of a more deterministic measurement but only captures a periodic snapshot of the network. Transient problems can be missed so real-time monitoring is also required. Furthermore, measurements taken at constant intervals where the call volumes vary greatly, as in most enterprises, do not provide an accurate picture of QoS or network performance.

The alternative — passive monitoring — provides real-time data and does not use any bandwidth. Passive monitoring measures the quality of actual calls. It does, however, require the network be equipped for passive measurement at the points in the network that best reveal call quality, such as the IP phone itself and the network's voice paths.

IP phones from leading IP Telephony vendors can now report IP network impairments or MOS scores back to a centralized management system. Here data gathered at customer premises is evaluated by network analysis software. In many cases, enterprise customers use management tools to internally monitor VoIP quality. When issues arise, the data collected by the network operator can be compared with the data from the customer's internal management tools. However, for any comparison to be valid, the time index on the two systems' logs (the operator's and the customer's) must agree. That requires a reliable and consistent time reference in the customer premise.

Many other packet-based services — in addition to IPTV and VoIP — also require accurate, secure, reliable, and network-wide NTP. Both gaming and video conferencing, for example, require that the end points be tightly synchronized to enable a satisfactory user experience.

VoIP QoS measurements are not simply about end-to-end performance measurement, however. They also determine the quality in various segments of the network — which is particularly useful when different providers are involved in end-to-end service delivery. Quite often, a VoIP call involves a VoIP provider's network, a backbone carrier, and a customer's local ISP network — any of which can be the source of a problem. Any fault isolation or event correlation could require comparison of logs from many different sources — again, which calls for accurate and consistent time across all of them.

Network time synchronization also plays an important role in VoIP billing and security. Accurate CDRs — based on actual call minutes — are an obvious prerequisite for correct billing and revenue realization. With respect to VoIP security, data theft and denial of service attacks are of particular concern. Many of the techniques used to launch attacks spoof the system time of the customer's VoIP gateway. This is done both to mask the timeline of such intrusions and to frustrate network forensics. It's yet another example of why secure time should be provided to customer endpoints.

Closely aligned with digital security is E911 compliance since emergency 911 services directly protect the physical security of the customer. While E911 is a legal requirement, it can also be notoriously complex to implement on IP networks. A key element is the carrier's ability to prove 911 log accuracy — which is made much easier if time is traceable back to a known secure and accurate NTP source.

Many other packet-based services — in addition to IPTV and VoIP — also require accurate, secure, reliable, and network-wide NTP. Both gaming and video conferencing, for example, require that the end points be tightly synchronized to enable a satisfactory user experience. But beyond NTP's services impact, NTP also has an impact on network operations itself.

NTP for Network Operations

Operators are currently being tasked with converging network services into an IP based network (NGN). This migration to managed packet based networks has created environments not previously encountered in the Internet. High service assurance requirements coupled with best effort protocols put additional burdens on operators to stringently manage their networks. Higher quality NTP services are required to mitigate the effects of jitter and path or load asymmetries in the network. These quality services will allow greater agility in capturing or diagnosing problems as they develop.

Performance monitoring and measurements

Network traffic is becoming asymmetric because the new services generate disproportionately larger and faster traffic flows downstream. These are services like IPTV or VoD in wireline networks or content downloads like ring tones, video clips, and songs in mobile services. Here the best performance indicator is latency from source to destination (e.g., headend to customer STB) — because that is what affects the user's experience most. Since upstream and downstream traffic flows are very different, roundtrip measurements don't capture packet delay and packet loss that only occur in one direction. That's why network measurement methods are giving way to one-way measurements. One increasingly popular method is one-way delay measurement or OWD.

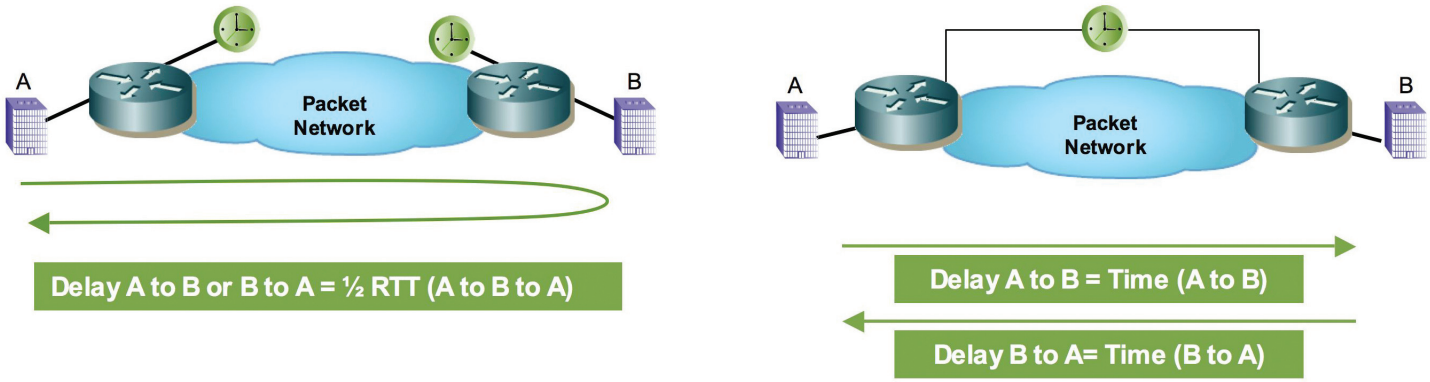

FIG 3 NTP Application Example: IP Performance Monitoring

A one-way measurement is reliable, however, only if the time indexes at both ends are highly synchronized. That's because packet networks are particularly prone to short-term transient effects. Even if a network performs well on average, it can still suffer frequent and extreme packet delays and jitter for very short periods — episodes that won't be detected unless packet transit times are precisely clocked.

OWD measurements therefore call for an accurate time base. IETF RFC 2679 specifies measurements at the 1 milisecond level today with momentum to push that lower. This will require measurement device accuracies in the 10-100 microsecond range for network QoS. Such accuracy cannot be achieved if the NTP source is located some number of hops away from its client. It needs to be nearby. That's because NTP packets are subject to the same network effects as the packets whose transit times they help measure. At the time operators most need accurate time, delivery could be compromised by the same impairments affecting other parts of the traffic flow.

Network fault diagnostics and recovery

Measuring performance is one part of QoS — diagnosing and recovering from faults when QoS degrades is another. There, too, precise time index synchronization is key. Often it's not till there's a failure that the need for synchronization becomes clear — either as part of the problem or as a tool to help fix the problem. Process monitoring systems use time indexes to track and log events. Non-synchronized indexes in different systems and servers result in inconsistent timelines.

A good example is the capturing of performance data. Without the same time index, logged events may appear to have happened sooner or later than they really did or in a different sequence — making it difficult to assess events like:

- Loss of connection
- Buffer overflow
- Missing packets
- Crashes
- Denial of service attacks

Another concern is denial of service (DOS) attacks. RMON event logs help network security experts reconstruct the timeline of a network crime — which again requires mutually consistent time stamped network packet transits.

Events are trapped, reported, and logged using the Remote Monitoring (RMON) services that reside in servers, routers, switches, and dedicated instruments. When a crash occurs, a stream of RMON events is reported. Each event is indexed with a NTP timestamp affixed by the reporting RMON agent. If these time stamps are synchronized, investigators can establish proper event sequence and more quickly identify the root cause.

Another concern is denial of service (DoS) attacks. RMON event logs help network security experts reconstruct the timeline of a network crime — which again requires mutually consistent time stamped network packet transits.

Billing and CDR generation

Billing is also a core function in telecom. Accurate billing is not only a regulatory requirement but also a major factor in customer satisfaction. As previously noted, CDRs are the primary source of billing information in VoIP — and so they are too in the rest of telephony. They provide information about call origination, destination, and duration. CDR duration includes the timestamp indicating when the call was initiated and either the call duration or the time the call was terminated. While the role of the CDR is much the same as before the packet era, the way CDRs are created has changed. It has gone from a batch process to more of a real time and distributed process — making time index agreement among cooperating systems even more important than it already was.

In the circuit switched era, central office switches and a few signaling transfer points (STPs) in the SS7 system collected CDR information. When calls went across multiple networks, VoIP gateways, and servers, batch processes combined data from the various segments the calls traversed.

Even in that non-real-time scenario, billing integrity very much relied on the accuracy of timestamps. Synchronization was and is particularly critical when billing discrepancies between carriers require time-consuming mediation. Today, many networks now include unified messaging, audio conferencing and other on-the-fly services — that require unified billing and therefore unified time measurement. Content delivery services such as ring tones, video, music etc. have a more complex delivery infrastructure utilizing many more systems distributed throughout the network — and these too must be unified with respect to time.

Take, for example, a content download request. The request must first be logged at a gateway, next validated by an AAA server, then served through a content server in one location, and the transaction recorded in yet another location. Thus, time stamped records from a number of systems, servers, and databases must be compiled to provide a single usage summary. That can be prohibitively difficult if their respective time indexes don't agree.

What's more, CDR production is migrating away from a batch process to one based more on decentralized systems interoperating in real time. This is partly in response to the real time nature of services and partly in response to an explosion in transaction volumes — which can be handled more efficiently closer to the network edge. This has led to two consequences. First, these systems need reliable NTP and second, serving them NTP from a core office creates inconsistent time indexes for all the reasons cited earlier.

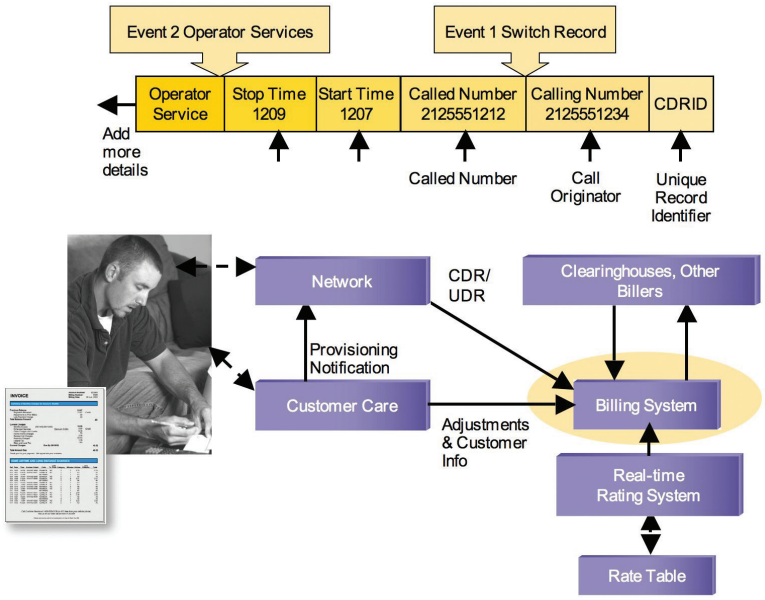

FIG 4 Telecom Billing Process

Such time discrepancies, of course, have a real impact on revenue realization and customer satisfaction. Most billing questions are resolved in favor of customers in order to keep their business. As a result, carriers lose significant revenue. Beyond billing, however, CDRs also help in fraud management, network analysis, mediation, and inter-carrier settlements.

Closer to the edge

The pressure to meet all these needs — support new services, improve network performance, accelerate fault diagnosis and recovery, and streamline billing — is pushing NTP services closer to the network edge. Correlation of events across multiple systems and devices requires a single global time scale. As the numbers have increased greatly — the number of events, the number of systems and devices, the number of hops, and the number of "9s" availability — operators must move accurate time closer to the points of need. The organic model of NTP deployment, as an outgrowth of enterprise/datacenter NTP, no longer measures up.

That's why a fourth revision of the protocol is underway. Although NTP functionality has been refined and strengthened over the years, it typically achieves accuracy on the order of milliseconds over a LAN and anywhere from hundreds of milliseconds to a few seconds over a wide area network. Network latency, jitter, and a constantly changing environment all take their toll. As the timebase of most devices is not monitored, it is possible for clients configured to receive NTP over a WAN to lose that connection and drift over a period of time by minutes, or even hours, without raising any alarms. That's unacceptable for carrier class performance.

The emerging requirements for NTP call for a reassessment of current NTP deployment practices. These practices have created NTP islands — unwieldy, unreliable, expensive and unmanageable disparate sources of time in the network. What's required is a scalable, reliable, manageable and accurate NTP infrastructure. Since use of the protocol is pervasive, the network engineering and NTP server performance aspects must also be refined.

The emerging requirements for NTP call for a reassessment of current NTP deployment practices. These practices have created NTP islands — unwieldy, unreliable, expensive and unmanageable disparate sources of time in the network. What's required is a scalable, reliable, manageable and accurate NTP infrastructure. Since use of the protocol is pervasive, the network engineering and NTP server performance aspects must also be refined.

Abbreviations and Acronyms

AAA - Access, Authentication and Accounting
ATIS - Association for Telecom Industry Standards
CDR - Call Data Record
CO - Central Office
DoS - Denial of Service
CA - Conditional Access
IETF - Internet Engineering Task Force
IIF - IPTV Interoperability Forum
IPPM - IP Performance Monitoring
IPTV - Internet Protocol TV
IT - Information Technology
LAN - Local Area Network
LAN - Local Area Network
NOC - Network Operations Center
NTP - Network Time Protocol
OWD - One-Way Delay
POP - Point of Presence
QoS - Quality of Service
RFC - Request for Comments, IETF Standards Document
RMON - Remote Monitoring
RTD or RT Delay - Round Trip Delay
RTCP - Real Time Control Protocol
SLA - Service Level Agreement
SS7 - Signaling System 7
STB - Set Top Box
STP - Service Transfer Point
TCP/IP - Transmission Control Protocol over Internet Protocol
UDP - User Datagram Protocol
VHO - Video Hub Office
VoD - Video on Demand
VoIP - Voice over IP, also know as IP Telephony
VSO - Video Serving Office
WAN - Wide Area Network

References

IETF RFC 1305 - Network Time Protocol (Version 3)
IETF RFC 2679 - A One-way Delay Metric for IPP

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