

The Future of Time and Frequency Dissemination

Judah Levine
Time and Frequency Division
National Institute of Standards and Technology
Boulder, Colorado 80303

Abstract

I will try to extrapolate the changes in the dissemination of time and frequency information that have taken place during the last 25 years to predict the future developments both in the methods of disseminating time and frequency and in the kinds of customers we will be asked to serve. Two important developments are likely to play pivotal roles in driving the evolution of dissemination. The first is the commercial availability of very high quality clocks — devices whose performance may eventually rival that of the current generation of primary frequency standards. The widespread use of these devices may blur the traditional distinction between client and server, and may replace it with a more symmetrical interchange of data among peers. The second is the increasing demand for digital time and frequency information driven by the increasing sophistication of everything from traffic lights to electric power meters. The needs of these individual users may not tax the state of the art of primary frequency standards in principle, but their large numbers and wide geographical distribution present a technological challenge that is difficult to meet at a reasonable price using existing methods.

Some of these problems may be solved (or at least addressed) using developments in communications and consumer electronics such as the increasing use of fiber-optic telephone circuits and the increasing bandwidth and sophistication of the cable network used to transmit television pictures. To be useful, these advances in hardware must stimulate parallel advances in software algorithms and methods. These advances are more difficult to predict with great confidence, but the developments of the last few years will be examined to provide some indications of the future.

Introduction

In 1964, the 1-day stability of what was then a state-of-the-art cesium clock was about 20 ns. It was very difficult for a user far from a timing laboratory to receive a signal that maintained this stability, and people who were serious about time usually used portable clocks to transfer it. It took a number of years for dissemination methods to improve to the point where portable clock trips were no longer necessary — the last trip from NIST to the BIH at the Paris Observatory was in the middle 1970s. State-of-the-art clocks today are about 100 times better than those of 30 years ago, and dissemination must improve accordingly. This is not a trivial undertaking; transit times, which may be tens of milliseconds or longer must be measurable and stable to a fraction of a nanosecond. The parameters that affect the transit time must also be understood well enough so that this level of stability can be achieved over long periods of time, which is perhaps even more difficult.

In addition, two other trends will also have great impact on the dissemination process. The first is the availability in the commercial market-place of devices which are at least as accurate as the primary standards of a generation ago. Sites equipped with this hardware cannot really be considered potential users of most of the services currently provided by a national timing laboratory — they are closer to peers, and their hardware may be more efficiently used by implementing two-way data exchanges rather than by the one-way time transmissions and the client-server relationships that have been the norm up to now. The second is the increasing demand for moderately accurate time in various digital formats. These uses do not tax the state of the art from the point of view of accuracy, but satisfying the sheer volume of the requests and their great diversity requires some careful planning. Both of these trends are likely to continue for the foreseeable future, and meeting them is likely to be a significant challenge during the next 25 years.

The High-End Dissemination Problem

The transit time of a signal between a timing laboratory and a moderately distant user is larger than the fluctuations of the signal produced by the frequency standard itself, so that the full accuracy of the clock can only be realized at a distant site if the transit time of the signal to the user is measured and the fluctuations about its mean value are well-characterized. Three techniques can be used to estimate the magnitude of this delay and to correct for it; the principles underlying all of them have been around for many years, although the details of the implementations have evolved with time.

1. Two-Way Methods

These methods use signals that travel in both directions along the path between the clock and the user and they estimate the one-way delay as one-half of the round-trip travel time. They obviously depend on the reciprocity of the path; while reciprocity is a simple concept, it is difficult to realize in practice. The basic problem is simple: if the hardware is inherently bi-directional, the message protocol is often half-duplex — the messages in the two directions traverse the path at different times and the fluctuations of the path delay in time may limit the reciprocity. If, on the other hand, the channel is full-duplex so that messages can travel in both directions simultaneously, then at least some of the hardware is inherently uni-directional so that the reciprocity will be limited by the imbalances between the send and receive channel end-point hardware even if the intervening path (through the atmosphere, for example) is highly reciprocal. Examples of both types of system are easy to find. Computer networks, for example use bi-directional hardware, but the channel is used in a half-duplex manner — the messages in the two directions are sent at different times, and the reciprocity will be limited by short-term fluctuations in the load on the network and especially at the switching points. Two-way time transfer using communications satellites, on the other hand, is a full-duplex method which exacts the expected price: different hardware is used for the two directions and the reciprocity is likely to be compromised unless the two channels are very carefully balanced.

2. Common-View Methods

These methods use one-way signals that travel along two different paths at almost the same time to two separated receivers. Unlike a two-way method which depends on the equality of the transit times in both directions on the same path, a common view method depends on the equality (or at least the stability) of the transit times in one direction along two different paths. Near-equality of the physical lengths of the two paths is obviously desirable since paths of different lengths can only be imperfectly compensated for by extrapolation — if the signals arrive at the end-points simultaneously they must have left the source at different times, whereas if they left the source at the same time they will not arrive at the end-points simultaneously. One of the advantages of this method is that the active cooperation of the transmitter is not required — it may not even “know” that it is part of a measurement. Signals from television stations, from LORAN stations and from GPS satellites have been used for common-view dissemination, and there have even been proposals to use the 60 Hz power system in this way.

3. Dispersion-Based Corrections

These methods estimate the correction to the transit time due to the refractivity of the transmission medium by measuring its dispersion — the dependence of the transit time on signal frequency. The system sends the data using several different frequencies simultaneously. The power of the technique depends on the fact that the time delay introduced by the medium can often be expressed as the product of a known function of the signal frequency and a known function of some path parameter (such as column density). If this is the case, the medium-dependent portion of the distance and dispersion equations are the same, so that a measurement of the dispersion can be used to correct the transit time itself. This method is not universally applicable since not all paths are dispersive, and some dispersive media cannot be characterized by a separable dispersion equation. Given the dependence on path parameters that we have assumed above, the fractional error in the determination of the transit time will be equal to the fractional error in the dispersion. The dispersion is usually much smaller than the transit time itself so that its measurement accuracy must be correspondingly higher. The error budget of the dispersion estimate often dominates the error budget of the entire measurement as a result.

All of these methods are currently used in disseminating time and frequency, and some systems use more than one method simultaneously. The GPS system, for example, is often used in common view mode, and its dual-frequency signal structure is designed to facilitate an estimate of the ionospheric refractivity using the observed dispersion between the L1 and L2 signals. The tropospheric refractivity, on the other hand, usually cannot be estimated from the dispersion. The effect is very small to begin with and is not completely separable because of the presence of a term due to tropospheric water vapor. Various methods have been used to estimate the tropospheric refractivity including balloon measurements and water-vapor radiometers, but none has proven completely successful. Corrections for the tropospheric index of refraction are already widely used in geodetic measurements using GPS, and they are likely to be necessary in time-transfer in the future.

Many of the current high-end users are more interested in frequency than in time, and the

ideal dissemination system for them would be one whose fluctuations could be characterized as white phase noise at all averaging times. The uncertainty in the frequency transmitted by such a system can be made as small as desired by increasing the averaging time until some other process (such as the performance of the clock itself) becomes important. No currently operating system satisfies this condition, and developing such a system will be one of the goals of the near-term future. Although it may be possible to improve the characteristics of the transmission media, further decreases in the uncertainty of the transit time may only be realized with more sophisticated averaging schemes, perhaps including extensive post-processing.

One of the difficulties in improving time dissemination is that many of the systems that are currently used for this purpose were in fact originally designed for some other function which was not very sensitive to the propagation delay itself or to fluctuations in the delay about some average value. A system that can meet the increasingly stringent requirements of the next 25 years may have to be designed as a time-transfer system from the start rather than as a piggy-back on a system designed for some other purpose.

At least in the short term, dissemination will probably depend on the two-way and common-view methods. Both of them are inherently symmetrical and do not single out either participant as client or server. Any system based on these methods is able to support the peer concept that is likely to be the optimum way of interacting with customers with very high-quality clock hardware. Common-view GPS would seem to have an initial advantage in terms of the number of sites that can participate in a simultaneous measurement campaign, but this advantage may not be fully realized in practice because of the intentional degradations of the GPS signal. A two-way system using communications satellites is less likely than a common-view GPS system to be limited by fluctuations in the atmospheric component of the transit delay, but this advantage may be lost due to time-varying asymmetries in the hardware.

The Low-End Dissemination Problem

The fundamental problem of high-end dissemination is finding a way of preserving the inherent accuracy of the clock hardware during the transmission of the information to the user. This issue is usually not so important to the low-end user. The more significant issues are often cost, reliability, ease of use and possibly legal traceability. The accuracy of the signal transmitted by WWV, for example, is adequate for many users, but receiving the signal reliably often requires an outside antenna, and extracting the time in a digital format requires something more than a bottom-of-the-line receiver. A telephone time service like the NIST Automated Computer Time Service may be easier than WWV to install and more reliable to operate, but each use requires a telephone call to the server. A user with many systems to synchronize may be faced with appreciable telephone costs, and a significant investment in server hardware will be required to provide an adequate level of service. A single ACTS server can handle perhaps 1500 calls/day; more than 500 servers would be needed if only 1% of the estimated 80 million domestic PC users used ACTS only once per day. Similar scaling arguments apply to almost any publicly-available service which attempts to estimate the transit time of the signal along the path to the users. (The cost of running a radio service like WWV obviously does not depend on the number of users, but the path delay can only be estimated using tables of the average

propagation characteristics at any site.)

The load on the servers could be reduced with better client oscillators that required less-frequent synchronization and more sophisticated use of the calibration data when they are acquired. The NIST Internet Time Service is a first attempt in this direction — it has demonstrated performance accuracies substantially better than one second using less than one calibration message per day. A single stratum-1 server can handle at least 20 000 requests per day so that the required server hardware increases much more slowly with heavy use than would be true for an ACTS-type system. Although ACTS can deliver time signals more accurately than the internet, the difference is not important in many applications.

Both ACTS and the internet time service are based on the two-way principles outlined above; the ACTS system is more accurate because its signals travel over a dedicated telephone circuit that is more easily characterized. Several internet methods are based on a combination of two-way and common view in an attempt to compensate for the relatively poor characteristics of the internet. Improvements here are quite likely, both in the speed of the network itself and in the sophistication of the software at each client node.

None of the currently available dissemination systems addresses the issue of authentication — the need for some users to be able to obtain a time-stamp that can provide legal proof of its authenticity. A simple example is the need to prove that a document describing the disclosure of an invention actually existed on a certain date. This requirement is currently satisfied for conventional documents by obtaining the stamp or the signature of a disinterested third party like a notary or a post-office, but there is no correspondingly simple method for authenticating the time-stamps on documents that are in digital format such as computer files.

A related issue to authenticity is the need for anti-spoofing — a method of allowing a client to verify that a time message originated from a genuine time server. It is unfortunate that the probability of fraud will increase as time-stamping becomes more important and as the certified data become more precious, but this is likely to be true, and some means of authenticating time transmissions may be required in the future. Several proposals have been advanced for adding authenticating digital signatures to messages transmitted over wide-area computer networks, and various public-key encryption systems such as the method currently used for GPS transmissions have been proposed for authenticating messages sent by other means, but none of these methods is generally available or particularly easy to use. It is quite possible that simplicity and robustness will continue to be conflicting goals, and there may be no alternative to systems that are complex and somewhat awkward to use as a result.

Conclusions

The continuing improvement in the stability and accuracy of frequency standards must be accompanied by corresponding improvements in the means used to disseminate time and frequency data to users if these standards are to be widely usable. Several dissemination methods currently exist, but none is likely to be completely adequate either from the point of view of adequate stability and accuracy to satisfy high-end users or from its ability to be scaled up to serve an increasing number of users with moderate requirements who are more

interested in ease of use and low cost. Several different techniques are likely to be required to serve these very disparate demands, although they are all likely to be based on the general principles we have discussed. The increasing demand for automated time-stamp systems and the increasing value of time-critical data make various forms of cheating or even fraud more lucrative, and some form of authentication of time transmissions may eventually be required. Existing authentication techniques tend to be cumbersome and inefficient, and finding a simpler method that maintains adequate security may become a high priority task in the next few years.

QUESTIONS AND ANSWERS

Dr. Cutler, Hewlett-Packard Labs: I would just like to comment on the very last thing that you talked about, anti-spoofing. There are some people at AT&T labs that have already developed a system of authentication that is very neat and very clever for time stamping using hashing techniques.

J. Levine: I have a similar proposal along the same line which is similar to the proposal. NTP, the Network Time Protocol that I spoke about this morning, has a DES incription mode connected to it. So people are beginning to think about it. But it is just a beginning. It is going to become a real problem in a few years.

Dr. Winkler, USNO: I have two comments. First, your interesting story about Ramsey, I was not aware of that. What that means really is the decision, if it would have been a decision of the BIPM or the CIPM to define the second; if it would have been done in respect to universal time seconds during these years when Ramsey estimated the correct frequency of cesium, we would be all right today, because that has not changed very much. We would still be offset 300 parts in ten to the tenth from ephemeris time. The problem is they selected ephemeris time in order to provide the continuity with the decision of the general conference of 1956 to adopt ephemeris time. And I think the reason why we have these leap seconds really goes back to the fact that we have pushed for a change too rapidly. We have done that before we really understood the implications of what was going to be done.

J. Levine, NIST: I would say that you are right, but it is also true — I mean, I was only a graduate student in 1965, and I spoke only when I was spoken to; and I wasn't really in the time and frequency community. But it was also true that there were two different communities because we haven't spoken at all about the spectroscopy and frequency community. And they were pulling in a very different direction.

Dr. Winkler: You are right. There was this difference of languages and everything, philosophy and everything, after that. But let me go to the second point, which I think is more of immediately practical importance and that is the point which you made about the problems of access and costs. I think the worldwide availability of GPS will eliminate them. There is simply no problem. You have a passive reception of — 10 billion people can do that if they want to. And it doesn't cost them anything because to receive all you have to provide the little receiver which will in five years be within a wristwatch; and everybody will have time within 10 ns at minimal cost. So I really think the problem which you have construed here as being a danger is going to go away.

And the last point which I would have to make is the problem of the legal time reference in the

United States. And we have to state that by law; the responsibility of NIST is to provide the standard of frequency, the standard of time to seconds. The responsibility of the Observatory is to provide the time reference.

J. Levine: Sorry, you misunderstood the essence of my point. The essence of my point was not that NIST had a unique responsibility; it's that some downstream user had to be able to have —

Dr. Winkler: I agree with that, and that is a problem. But the problem is very similar to what you have in navigation. If an accident happens in the high seas and the question comes up, "Has there been sufficient care in navigation or has there been a lack of care? Has the access to the navigation system been so that prudence has been exercised or not?" And that also immediately applies to any other technical information. And if anyone, just without making any checks, accepts the next time reference from the next church tower and makes critical decisions, this is not sufficient prudence and not sufficient care; and it will not stand up in court. So I refer to the legal practices which govern questions of accidents at sea and the use of navigational information; because exactly the same situation exists here in such expressions as when you have to make decisions on time or have to be certified on certain moments and so on.

Ralph Partridge, Los Alamos National Lab: You were commenting about the escalating costs of having to supply a million telephone calls today and saying that it has to be done in a reasonable way. And of course, Gernot's point is well made with the wristwatch GPS receiver and so forth; except those things need antennas and the antennas have to be outdoors. So it may not be as universally usable as he was hoping perhaps. But if you can have a system as it expands does so exponentially rather than linearly, you come out ahead. And so if you simply have the time disseminated locally at each central office, each one can serve 10,000 telephones, that can get it from the next higher level; and now your system increases exponentially and the cost goes as the logarithm in the number of servers.

J. Levine: I agree with you. That brings up the issue of traceability. You then have to be sure that all of these folks are doing the right thing. I agree with you 100 percent.

Ron Beard, Naval Research Lab: I just wanted to make the point that time is already money, because one of the bigger users of the computer networks in data exchange that is currently ongoing is commodities and monetary exchange. I think it is highly important that they know exactly when these transactions take place. And as they are going through these networks, the timing of them, when they are actually associated with the various buyers is already a critical point. So I think we are already in that state. We just may not appreciate many of the implications yet.

Dr. Winkler: I think this is an extremely important point. You are absolutely right. And the most critical one is the confusion which is possible during a leap second, when it is introduced. And it may be the most powerful argument to revise the rules which govern the insertion of leap seconds and the coordination of universal time. It is really something which we will have to consider.

But let me say another thing here. And that is there is always going to be a minimum interval of time within which actions must be considered as simultaneous. And they will have to deal with the ability of the human mind to do different things, or to distinguish two different things. Where time can be measured electronically, of course there is no question. But when you talk about actions and certification and decisions I think we are not talking about microseconds.