Survivable, Real Time Network Services

Defense Advanced Research Projects Agency
Contract F30602-98-1-0225, DARPA Order G409

Quarterly Progress Report
1 April 1999 - 30 June 1999

David L. Mills
Electrical Engineering Department
University of Delaware

1. Introduction

This report covers the work done in support of the DARPA Information Technology Office program in computer networking. Contributors to this effort include Prof. David L. Mills and graduate students Qoing Li and Robert Redwinski. The project continues previous research in network time synchronization technology jointly funded by DARPA and US Navy. The technology makes use of the Network Time Protocol (NTP), widely used in the Internet, together with engineered modifications designed to improve accuracy in high speed networks. Specific applications benefiting from this research include multicast topologies, multimedia, real-time conferencing, cryptographic systems, and management of distributed, real-time systems.

This quarterly report is submitted in traditional report form on paper. As the transition to web-based information dissemination of research results continues, almost all status information and progress reporting is now on the web, either on pages belonging to the principal investigator or to his students. Accordingly, this and future progress reports will contain primarily schedule and milestone data; current status and research results are reported on web pages at www.eecis.udel.edu/~mills in the form of papers, technical reports and specific briefings.

1.1 NTP Version 4

Work continued on the NTP Version 4 reference implementation and distribution for Unix, VMS and Windows. We have incorporated numerous patches required by the over two dozen ports of the code to various Unix architectures and operating systems. There are now some 40 volunteers who maintain various ports, test interoperability, measure performance and provide special purpose drivers and subsystems.

We have automated the software update process and converted to a CVS repository. This has considerably reduced the risk of multiple simultaneous updates, as well as the manual labor to upload the software and documentation, including the rapidly changing lists of public NTP servers.

2. Nanokernel Project

The motivation and status of the nanokernel implementation was described in the most recent quarterly report. Since then, a number of improvements have been made, including a reduction in residual error by a factor of ten and improved resilience to network transients and malfunctions in hardware and software. Following is a brief history of the project as it evolved over the last several years. Following this is a brief summary of the scope and rationale and some lessons learned.

As workstations and networks have evolved to higher and higher speeds, so have the expectation that computer network timekeeping will evolve to greater and greater accuracies. The benchmark accuracy regime during the Fuzzball era in the late 1980s was of the order of one millisecond. By 1994 expected accuracies with then-modern workstations like the SPARC IPC were in the order of 100 microseconds. In the last few years computer and network speeds have increased by a factor of at least ten to 500-MHz computer clock rates and 100-Mbps LAN speeds. For at least some time, computer and network speeds are expected to accelerate.

In order to keep pace with these developments, the detail engineering of the NTP algorithms has experienced many enhancements, some incremental, some profound. However, the degree to which performance can be enhanced using only the NTP foundation code has essentially reached its limit. The most recent refinements have improved the performance in very congested networks and increased the interval between updates to well over a day with only minor reduction in performance.

By 1994 it became clear that the principal barrier to improved performance was the stability of the oscillator used to maintain system time. We developed analytical models for room-temperature quartz crystal oscillators used in typical workstations and concluded that the interval between oscillator corrections must be reduced well below the typical minimum of 64 s necessary to avoid interfering with revenue traffic on the local network. In the NTP Version 4 implementation, it is possible to reduce this interval to 16 s, but this is useful only when justified by allowable overhead.

Another problem arises from the mechanism used to discipline the Unix clock, where new corrections are computed one per second and implemented using the Unix adjtime() system call. However, errors up to several hundred microseconds can accumulate between adjustments. The only solution to this problem is to discipline the clock oscillator at every clock interrupt, which in modern workstations ranges from 10 ms to less than 1 ms.

In 1994 a package of kernel modifications, called the microsecond kernel, was implemented for several experimental Unix kernels, including Hewlett Packard, Sun Microsystems and Digital RISC. Subsequently, this package was incorporated in the stock kernel for Digital Alpha and Solaris SPARC. Later the package was implemented for the Linux and FreeBSD kernels and possibly others. While the package included an enhanced version of the clock discipline loop refinements for the NTP daemon, it also contained a special hybrid discipline loop for use with precision pulse-per-second (PPS) signals generated by some radio clocks and laboratory instruments. This improved the overall accuracy generally to the order of a few microseconds, but the resolution was limited to no better than 1 μs.

As the latency for a system call has approached 1 μs in modern workstations, it has become clear that the inherent resolution no better than this may become the latest performance limiting factor. Accordingly, the microsecond kernel was redesigned with result the nanosecond kernel now deployed in expreimental kernels for SunOS SPARC and Digital Alpha. A number of experiments have been carried out that verify that the nanokernel in a fast workstation can indeed discipline the clock to the order of 50 ns RMS relative to a cesium oscillator. The theory, design and operation of the nanokernel algorithms is described in the web documentation at www.udel.edu/~mills/resource.htm and in the distribution nanokernel.tar.gz in pub/ntp at ftp.udel.edu.

A key issue to identify in the development of the nanokernel is the role of simulators. Based on past experience with the NTP simulator, the design exercise used a special purpose simulator in which the various algorithms were embedded. Based on our earlier work in modelling computer oscillators, we were able to replicate quite closely the behavior of typical computer oscillators and study the response of the nanokernel algorithms. Notwithstanding the positive experience with this approach, the most important factor may be the fact that the algorithms themselves are lifted from the simulator and patched in the existing kernel code with few if any changes. In this way we can reduce the likelihood of bugs and also make strong assertions about the behavior of the real kernel in response to extremes in the performance envelope.

Another important issue is the completeness and accuracy of the documentation. Prosecution of a research agenda which convince a number of other researchers to adopt and evaluate invasive schemes like the nanokernel need to be justified, motivated and calibrated in persuasive ways. A good deal of effort went into the nanokernel documentation. Like all recent projects, the documentation was done for web browsing and put up not only in the actual software distribution, but also put up in more public places, like the NTP web site and this investigator's home page. The documentation includes the rationale, design approach, theoretical and analytical models and measured proof of performance. A status report and briefing slides on this work are at www.eecis.udel.edu/~mills/status.htm. Not the least useful result of this work is the ease of turning the pages into a scientific paper.

3. Miscellany

The routing plan for the DCnet (128.4) network was reviewed and many changes made in local topology. All machines connected to more than one subnet run the gated routing daemon. A Torrent router on indefinite loan handles routes in and out of the DCnet subnets, as well as a multicast tunnel to the 128.4.2 Backroom facility subnet. Routes via CAIRN/DARTnet are imported from the DARTnet router and provided to all subnets. One interesting fact is that the number of such routes has exploded recently, with result that many clients of our busy time servers use the CAIRN routes both in and out of the local wires.

A PC used for program development has been upgraded to FreeBSD 3.2 in anticipation of the expected upgrade of the CAIRN routers. This version has the nanokernel and provisions for precision PPS signals. We intend to operate this machine as a dedicated CAIRN time server, as well as for general program development.

4. Personnel

Robert Redwisnki completed his thesis and graduated with the degree of Master of Science in June, 1999. His thesis, “Simulated Routing Protocol Survivability in Networked Environments,” is the result of implementation and experiment with very large networks and the simulator described previously. A status report and briefing slides on this work are at www.eecis.udel.edu/~mills/status.htm.

Ajit Thyagarajan continues work to complete his dissertation. His topic is the analytical and experimental study of autoconfigure algorithms suitable for deployment in a survivable internet. He has completed the design and analysis phase and is working on simulation and experiments. He expects to complete all requirements by the end of the Fall semester 1999.

David Mills was elected ACM Fellow in May 1999.

5. Meetings

Following is a list of presentations and meetings made during the current period of performance. Note that a full complement of briefing slides is archived at www.eecis.udel.edu/~mills/colloq.htm. In most cases the slides are organized as an extended briefing designed to be read via the web, but not all the slides were used at the presentation. Thus, a considerable amount of extra information beyond the oral presentation is available.

A briefing on the issues of securely authenticating and automatically configuring assets on a large, survivable network was presented at the DARPA Next Gerneration Internet Meeting on 26-29 October 1998. Emphasis at this presentation was on the autokey and autoconfigure algorithms described in previous reports.

A briefing on the issues involved with extremely precise time synchronization was presented at the DARPA High Confidence Networking Meeting on 15-16 April 1999. The principal issue of interest to the community is the robustness of engineered kernel enhancements in the face of timing errors introduced by hardware or software disruptions, including component failure and hostile attack.

A background briefing was presented for DARPA Program Manager Mari Maeda on 28 April 1999. The presentation was conducted by telephone using previously sent briefing slides in PowerPoint format. This investigator concluded this mode of presentation is an efficient means of progress reporting, since it avoids travel and minimizes the reporting effort.

6. Plans for Next Quarter

We plan to continue refinement of the NTP algorithms and data structures as the result of a program of analysis and simulation developed previously for a paper published in IEEE Trans. on Networking. We also plan to continue implementation of necessary support software for the autokey and autoconfigure functions developed for demonstration with NTP.

7. Publications

All publications, including journal articles, symposium papers, technical reports and memoranda are now on the web at www.eecis.udel.edu/~mills. Links to the several publication lists are available on that page, as well as links to all project descriptions, status reports and briefings. All publications are available in PostScript and PDF formats. Briefings are available in HTML, PostScript, PDF and PowerPoint. The project descriptions are cross-indexed so that the various interrelationships are clearly evident. Also included are the documentation pages for various public software distributions. Links to other related projects at Delaware and elsewhere are also included on the various pages. Hopefully, the organization of these pages, which amount to a total of about 300 megabytes of information pages and reference documents, will allow quick access to the latest results and project status in a timely way.

Following is a retrospective list of papers, reports and memoranda supported wholly or in part by this project and the immediately preceeding project “Scalable, High Speed, Internet Time Synchronization," DARPA Order D012. The complete text of all papers and reports, as well as project briefings, status reports and supporting materials is at www.eecis.udel.edu/~mills.

7.1 Papers

1. Mills, D.L. Adaptive hybrid clock discipline algorithm for the Network Time Protocol. *IEEE/ACM Trans. Networking 6, 5* (October 1998), 505-514.

2. Li, Qiong, and D.L. Mills. On the long-range dependence of packet round-trip delays in Internet. *Proc. IEEE International Conference on Communications* (Atlanta GA, June 1998), 1185-1191.

3. Mills, D.L., A. Thyagarajan and B.C. Huffman. Internet timekeeping around the globe. *Proc. Precision Time and Time Interval (PTTI) Applications and Planning Meeting* (Long Beach CA, December 1997), 365-371.

4. Mills, D.L. Authentication scheme for distributed, ubiquitous, real-time protocols. *Proc. Advanced Telecommunications/Information Distribution Research Program (ATIRP) Conference* (College Park MD, January 1997), 293-298.

5. Mills, D.L. The network computer as precision timekeeper. *Proc. Precision Time and Time Interval (PTTI) Applications and Planning Meeting* (Reston VA, December 1996), 96-108.

6. Mills, D.L. Improved algorithms for synchronizing computer network clocks. *IEEE/ACM Trans. Networks 3, 3* (June 1995), 245-254.

7.2 Technical Reports

7. Sethi, A.S., H. Gao, and D.L. Mills. Management of the Network Time Protocol (NTP) with SNMP. Computer and Information Sciences Report 98-09, University of Delaware, November 1997, 32 pp.

8. Mills, D.L. A precision radio clock for WWV transmissions. Electrical Engineering Report 97-8-1, University of Delaware, August 1997, 25 pp.

9. Mills, D.L. Clock discipline algorithms for the Network Time Protocol Version 4. Electrical Engineering Report 97-3-3, University of Delaware, March 1997, 35 pp.

10. Mills, D.L. Proposed authentication enhancements for the Network Time Protocol version 4. Electrical Engineering Report 96-10-3, University of Delaware, October 1996, 36 pp.

11. Mills, D.L. Simple network time protocol (SNTP) version 4 for IPv4, IPv6 and OSI. Network Working Group Report RFC-2030, University of Delaware, October 1996, 18 pp.

12. Mills, D.L. Simple Network Time Protocol (SNTP). Network Working Group Report RFC-1769, University of Delaware, March 1995, 14 pp.

13. Mills, D.L. Simple Network Time Protocol (SNTP). Network Working Group Report RFC-1769, University of Delaware, March 1995, 14 pp.

7.3 Internet Drafts

14. Mills, D.L., T.S. Glassey, and M.E. McNeil. Authentication Scheme Extensions to NTP. Internet Draft draft-mills-ntp-auth-coexist-01.txt, University of Delaware and Glassey-McNeil Technologies, September 1998, 9 pp.

15. Mogul, J., D.L. Mills, J. Brittenson, J. Stone, P.-H. Kamp and U. Windl. Pulse-per-Second API for UNIX-like Operating Systems, Version 1.0. Internet Draft draft-mogul-pps-api-02.txt, IETF, June 1998, 25 pp.