Internet Throughput and Delay Measurements Between Sequoia 2000 Sites

Joseph C. Pasquale, George C. Polyzos,
Kevin R. Fall, Vachaspathi P. Kompella

Computer Systems Laboratory
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093-0114

December 1991

Technical Report CS91-219

Abstract

We report performance measurements of Internet connections between five Sequoia 2000 sites. Throughput and delay statistics are presented for various message sizes and for both daytime and nighttime. The highest throughput observed was 85 KB/s between UCSD and UCLA at night and the lowest was 1 KB/s between UCSD and DWR during the day.

For more information contact: Joseph Pasquale at pasquale@cs.ucsd.edu, 619-534-2673, 619-534-7029 (fax), or George Polyzos at polyzos@cs.ucsd.edu, 619-534-3508, 619-534-7029 (fax).

1. Introduction

The Sequoia 2000 project [1] is developing computing and communications technology to support Global Change research. The goal is to build a Terabyte database of important, very large datasets to be accessible by Earth Scientists in real-time from their workstations at the Sequoia 2000 sites distributed across California.

To provide acceptable performance and a testbed for communications research a new computer network interconnecting the Sequoia 2000 sites is planned to be deployed. This network will be dedicated to serve Sequoia 2000 applications and research on these issues.

The Sequoia 2000 sites are already interconnected through the Internet. The current Internet uses the NSFNET as a backbone network and relies on many regional and other networks for further distribution. The NSFNET backbone is in the process of being converted to a T3 network, i.e., using links with a 45 Mb/s data rate.[1] The T3 links are already in place, but some traffic is still routed over T1 links (1.5 Mb/s).

2. Goals

The goal of this study was to provide estimates of the bandwidth available to and the delay experienced by applications distributed across Sequoia sites using Internet connections for communication. No special optimizations were attempted. Therefore, the results that are presented here show what typical applications should expect to see as network performance. In particular, the performance metrics do not necessarily reflect what is possible to achieve through the Internet, even though we have no reason to believe that it is different from what is presented, particularly since one has no control over intermediate systems.

The TCP protocol was used at the transport layer. This protocol provides reliable data delivery which is what almost all current applications desire or assume. The implementations of TCP used were the ones provided by the hosts participating in the measurements.

3. Topology

The five Sequoia 2000 sites we have considered are the campuses of the University of California at Berkeley (UCB), Los Angeles (UCLA), Santa Barbara (UCSB), and San Diego (UCSD), and the Department of Water Resources of the State of California (DWR).

We have not addressed the problem of local connectivity. This problem is more interesting and severe on the San Diego campus, where three Sequoia 2000 groups are located: (i) the Computer Systems Laboratory (CSL) in the Department of Computer Science and Engineering, (ii) the Scripps Institution of Oceanography (SIO), and (iii) the San Diego Supercomputer Center (SDSC). All measurements at San Diego were made from a machine at SDSC. SDSC is a node on the NSFNET backbone, manages CERFnet, and is the main network gateway for the whole UCSD campus. Therefore, it is a natural ‘central’ point for UCSD networking. CSL is connected to SDSC by the campus FDDI ring with a data rate of 100 Mb/s. SIO is currently connected to the UCSD campus by broadband Ethernet with a 5 Mb/s data rate.

Two regional networks, BARRNet and CERFnet, provide network access for all five sites. Maps of BARRNet and CERFnet are provided in Appendix A. Figure 1 provides a high level description of the topology of the relevant parts of the networks interconnecting Sequoia 2000 sites. Note that many details are hidden in this high level map.[2] Link speeds range from 512 Kb/s to 45 Mb/s.

---

[1] We use B as an abbreviation for byte and b for bit.

[2] An idea of the level of abstraction can be obtained by comparing this map with those of Appendix A.

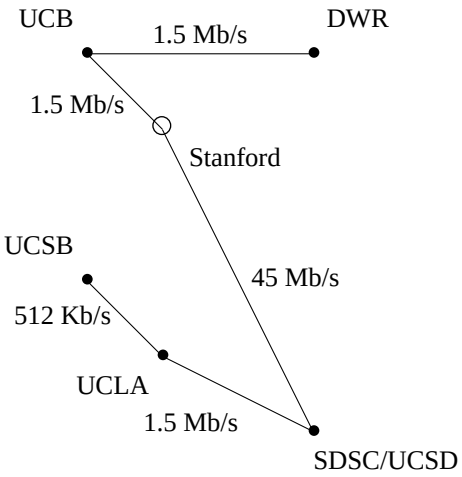

Fig. 1: High-level connectivity map of Sequoia 2000 sites.

The machines and their operating systems used for the measurements are shown in the following table:

Site	Host	Type	Operating System
UCB	springer.berkeley.edu	Sun 4c	SunOS 4.1.1B
UCLA	yosemite.atmos.ucla.edu	Sun 4	SunOS 4.1.1
UCSB	ucmbo.ucsb.edu	Sun 4c	SunOS 4.1.1B
UCSD	hydra.sdsc.edu	Sun 4	SunOS 4.1.1
DWR	quimby.water.ca.gov	Sun 4c	SunOS 4.1.1

4. The Measurement Tool

The Swedish Institute of Computer Science (SICS) developed a general purpose protocol performance measurement system called the SICS Protocol Implementation Measurement System (SPIMS) [2]. Through a measurement specification language, it is possible to describe a set of experiments directed toward obtaining a desired performance metric. Two machines, connected by a computer network, using one of the protocols that SPIMS understands are required for the measurements. A daemon runs at each host, waiting to start a set of experiments. When the initiator begins an experiment, the responder replies by establishing a connection which is used to control the experiment (and which is different from and independent of the data path being measured). The actual module that performs the experiment can be a standard SPIMS benchmark, or one specially designed by the user.

Using the measurement specification language, the user can specify the particular measurement module to be run at both ends. As an example, the user can specify: `10 rr 32*1000.` This means that the `request-response` benchmark (abbreviated `rr`) should be run 10 times, with each run consisting of 1000 transmissions of 32-byte data packets. Furthermore, it is also possible to specify which network protocol is to be used for the measurements.

Various benchmarks are available with SPIMS. The ones we used to measure the network performance were `bulk_get` and `request-response.` The `bulk_get` benchmark is initiated by a message from the initiator to the responder. The responder starts sending data units of the size set in the specification. When the required number of trials have been completed, the initiator sends a terminate transfer message to end this run. The process is repeated until the required number of runs is completed. Various statistics are then computed.

With the `request-response` benchmark, the responder sends a control message to the initiator requesting the transmission of a message of the specified size. When the message is received, the responder sends it back immediately. The initiator can then easily measure the round-trip time. This process is repeated as many times as

specified.

For both benchmarks SPIMS reports the minimum, maximum, mean, and standard deviation of the target metrics over the specified runs. The default values for the number of runs provided very low standard deviations, compared to mean values, leading to very tight confidence intervals.

5. Experiments

We have measured effective throughput and message round-trip delay of TCP connections for all 16 combinations of source-destination pairs of Sequoia 2000 sites for various message sizes ranging from 1 to 16 KB and for both daytime and nighttime. Measurements were made between all pairs of nodes, in both directions. Individual benchmarks were run for throughput and delay. The benchmarks, measuring one performance metric for various message sizes, were staged two hours apart in order to avoid interference among them (since many of the paths used to connect the sites overlap). Daytime experiments were run between 10 am and 6 pm and nighttime measurements were performed between 1 and 9 am. The schedule of measurements is provided in Appendix B. The measurements were performed between September 9 and September 25, 1991.

6. Results

We provide graphs for mean measured throughput as function of the message size from each source to each of the four destinations in Appendix C. Similar graphs for mean round-trip delay are provided in Appendix D. As can be seen from the shape of the graphs, the results do not reveal unexpected behavior, except that the best performance results are far below what is possible with quiet laboratory networks (see below).

Barring a few outlying points, throughput is relatively constant during the daytime sets and during the nighttime sets, with the daytime sets showing almost uniformly lower throughputs. The round-trip delays show a nearly linear increase with message size. The slope of the delay curves, as computed by a curve-fitting routine, closely matches the throughput figures obtained through separate measurements. Finally, the 95% confidence intervals obtained were very tight, suggesting that the figures for the means presented below are descriptive of the actual performance during the measurement period.

Mean throughputs for 8 KB messages for each source-destination pair are presented in Table 1 for daytime hours and in Table 2 for nighttime hours. The maximum and minimum throughputs observed over all experiments are shown in Table 3 and Table 4, respectively. Mean round-trip delays for 1 B messages for day and night are shown in Table 5 and Table 6, respectively. Similar results for 8 KB messages are provided in Tables 7 and 8.

Table 1: Day Throughput (KB/s) for 8 KB Messages[4]

from \ to	UCB	UCLA	UCSB	UCSD	DWR
UCB	—	14.58	12.33	29.63	38.81
UCLA	16.88	—	33.47	42.68	12.05
UCSB	14.44	41.45	—	35.35	10.77
UCSD	27.35	50.93	14.52	—	20.97
DWR	29.99	12.63	8.06	19.26	—

[4] These results were obtained with an 8 KB message size. However, the measurements show insensitivity to message size, as can be seen from the complete measurement set shown in Appendix C.

Table 2: Night Throughput (KB/s) for 8 KB Messages

from \ to	UCB	UCLA	UCSB	UCSD	DWR
UCB	—	21.23	17.29	39.56	55.48
UCLA	19.72	—	39.95	68.39	11.34
UCSB	18.47	50.80	—	46.36	17.58
UCSD	53.98	69.38	38.95	—	32.61
DWR	59.31	20.10	16.60	18.22	—

Table 3: Maximum Observed Throughput (KB/s)

from \ to	UCB	UCLA	UCSB	UCSD	DWR
UCB	—	32.42	23.75	55.08	77.05
UCLA	28.98	—	55.55	80.63	22.41
UCSB	26.21	56.72	—	52.59	23.35
UCSD	69.19	85.57	53.44	—	50.49
DWR	78.43	26.81	24.35	38.99	—

Table 4: Minimum Observed Throughput (KB/s)

from \ to	UCB	UCLA	UCSB	UCSD	DWR
UCB	—	3.05	6.15	9.01	12.83†
UCLA	4.48	—	12.46	10.95	1.11
UCSB	8.92	23.54	—	15.64	4.87
UCSD	3.04	32.30	9.05	—	5.89
DWR	8.51	5.49†	2.03	1.02	—

† Night measurement.

Table 5: Day Round-Trip Delay (ms) for 1 B Messages

from \ to	UCB	UCLA	UCSB	UCSD	DWR
UCB	—	117.80	164.50	70.00	28.70
UCLA	144.90	—	47.70	51.30	N/A
UCSB	149.10	37.30	—	58.70	149.30
UCSD	62.80	86.90	71.00	—	80.50
DWR	34.70	140.80	208.10	94.50	—

Table 6: Night Round-Trip Delay (ms) for 1 B Messages					
from \ to	UCB	UCLA	UCSB	UCSD	DWR
UCB	—	N/A	124.80	42.80	112.70
UCLA	117.80	—	37.00	37.10	N/A
UCSB	98.50	37.20	—	58.90	124.70
UCSD	46.00	49.00	50.40	—	62.70
DWR	25.40	140.70	114.20	88.60	—

Table 7: Day Round-Trip Delay (ms) for 8 KB Messages					
from \ to	UCB	UCLA	UCSB	UCSD	DWR
UCB	—	N/A	1095.00	747.50	495.40
UCLA	1081.00	—	562.10	357.20	N/A
UCSB	1322.40	559.80	—	539.20	2124.50
UCSD	653.00	638.90	848.30	—	689.90
DWR	537.30	1428.30	1674.80	987.90	—

Table 8: Night Round-Trip Delay (ms) for 8 KB Messages					
from \ to	UCB	UCLA	UCSB	UCSD	DWR
UCB	—	N/A	1140.70	424.40	324.70
UCLA	1003.90	—	434.80	268.70	N/A
UCSB	858.90	464.00	—	499.20	1098.30
UCSD	380.20	503.50	475.30	—	684.10
DWR	322.20	1127.60	911.00	1708.30	—

Figure 2 shows the average throughput and delay for 8 KB messages transmitted during the day as a function of distance in hops. Hops between sites are determined according to the topology in Figure 1. For example, UCSD is one hop away from UCB, while UCSB is four hops away from DWR. We fit two exponential curves, which are $\text{throughput}(d) = 54.011 \times 2^{-0.65615 \times d}$ and $\text{delay}(d) = 375.88 \times 2^{0.40527 \times d}$, where d is the number of hops. On average, throughputs at night are 50% greater than during the day. (Delays at night are approximately 25% less then during the day, however this statistic is not very reliable).

7. Comments

The obtained measurements show end-to-end, application-to-application performance. They are affected by network traffic sharing the same links, gateway load and software performance, and load on the measurement hosts. We believe the variations in performance, more easily seen in the throughput measurements, reflect differences in some of these conditions since there is no general trend for throughput as a function of message size, and delay shows a linear trend, as expected. Little can be done to remove this variation.

We have not attempted to attribute performance loss to specific components. However, we have investigated one specific element: protocol overhead. The results, shown in Table 9 for 8 KB messages, suggest that the

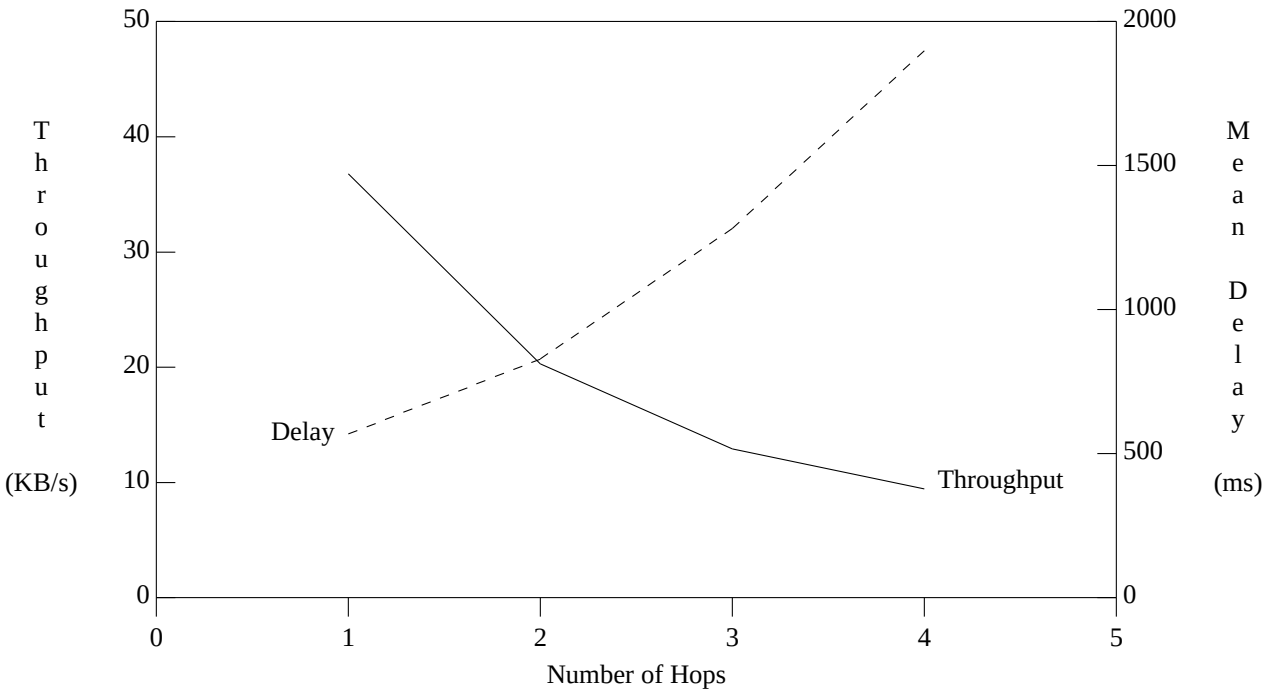

Fig. 2: Average Throughput and Delay during the Day by Distance.

implementations of TCP/IP used at the hosts do not limit performance severely, at least up to Ethernet speeds. The measurements were performed over relatively quiet LANs and links entirely within our lab. The T1 measurements used the TurboChannel T1 adapters [3]. Throughput was relatively insensitive to message size.

Table 9: Throughput Performance for Various Host Architectures and Physical Links

from	to	Link Type	Link Data Rate (Mb/s)	Throughput (Mb/s)	Throughput (KB/s)	% of Link Data Rate
DS 5000/200	DS 5000/200	T1	1.544	1.467	183.36	95
DS 5000/200	DS 5000/200	Ethernet	10	7.517	939.67	75
DS 5000/200	DS 5000/200	FDDI	100	11.53	1441.24	11.5
SparcStation 2	SparcStation 2	Ethernet	10	6.094	761.76	61
SparcStation 2	Sun 3	Ethernet	10	3.765	470.59	38
Sun 3	Sun 3	Ethernet	10	3.187	398.40	32

8. Comparison to a Hypothetical T1 Network

Consider comparing the throughput performance of the Internet with a hypothetical private T1 network connecting the Sequoia sites. The maximum bandwidth we were able to achieve with an unloaded T1 line was approximately 183 KB/s, as shown in Table 9. The average one-hop throughput we measured over the Internet was 37 KB/s during the day, and 55 KB/s during the night. The average one-hop maximum throughput (which is the average of the maximum throughputs between sites separated by one hop), was 70 KB/s. Thus, in very rough terms, a private T1 network will increase the throughput performance over the existing Internet by 400% during the day, 250% during the night, and at least 150% under the best conditions we observed with the Internet.

9. Conclusions

We have performed throughput and delay measurements for Internet connections between Sequoia 2000 sites for various message sizes and during peak (day) and non-peak (night) hours. Our best throughput result between UCLA and UCSD achieves approximately half the nominal link speed. Daytime throughputs are of the order of a few tens of KB/s between any two sites. It is interesting to note that the topology of the sub-network connecting Sequoia 2000 sites through the Internet is linear, with DWR at one end and UCSB at the other. Thus, as expected, the worse performance was typically encountered between sites that are farther apart.

Based on the Internet measurements and lab measurements for a T1 link and FDDI, we conservatively estimate that a dedicated, relatively quiet, T1-based network would provide a 150-400% improvement over the currently available bandwidth. A T3-based network with FDDI for local distribution seems to be limited to approximately 11 Mb/s (1.4 MB/s) between remote applications. However, if achieved, this is a factor of 100 improvement over the average measured day throughput between UCB and UCSB.[5]

References

1. M. Stonebraker and J. Dozier, ‘‘Sequoia 2000: Large Capacity Object Servers to Support Global Change Research,’’ Sequoia 2000 Technical Report #1, UC Berkeley, July 1991.
2. *SICS Protocol Implementation Measurement System: User Manual*, Swedish Institute of Computer Science, Kista, Sweden, February 1991.
3. D. Boggs, ‘‘TurboChannel T1 Adapter,’’ WRL Research Report 91/4, DEC, Palo Alto, CA, April 1991.

Appendix A: Maps of BARRNet and CERFnet

[5] These remarks assume that there will be no considerable contention among applications for the dedicated network, or that contention problems will be addressed, e.g., by the mechanisms described in [1].

Appendix B: Schedule of Measurements

Table B.1: Day Throughput Measurements Schedule					
from \ to	UCSD	UCLA	UCB	UCSB	DWR
UCSD	—	2.00 pm	12.00 pm	10.00 am	4.00 pm
UCLA	12.00 pm	—	4.00 pm	2.00 pm	10.00 am
UCB	2.00 pm	10.00 am	—	4.00 pm	12.00 pm
UCSB	4.00 pm	12.00 pm	10.00 am	—	2.00 pm
DWR	10.00 am	4.00 pm	2.00 pm	12.00 pm	—

Table B.2: Night Throughput Measurements Schedule					
from \ to	UCSD	UCLA	UCB	UCSB	DWR
UCSD	—	5.00 am	3.00 am	1.00 am	7.00 am
UCLA	3.00 am	—	7.00 am	5.00 am	1.00 am
UCB	5.00 am	1.00 am	—	7.00 am	3.00 am
UCSB	7.00 am	3.00 am	1.00 am	—	5.00 am
DWR	1.00 am	7.00 am	5.00 am	3.00 am	—

Table B.3: Day Delay Measurements Schedule					
from \ to	UCSD	UCLA	UCB	UCSB	DWR
UCSD	—	3.30 pm	1.30 pm	11.30 am	5.30 pm
UCLA	1.30 pm	—	5.30 pm	3.30 pm	11.30 am
UCB	3.30 pm	11.30 am	—	5.30 pm	1.30 pm
UCSB	5.30 pm	1.30 pm	11.30 am	—	3.30 pm
DWR	11.30 am	5.30 pm	3.30 pm	1.30 pm	—

Table B.4: Night Delay Measurements Schedule					
from \ to	UCSD	UCLA	UCB	UCSB	DWR
UCSD	—	6.30 am	4.30 am	2.30 am	8.30 am
UCLA	4.30 am	—	8.30 am	6.30 am	2.30 am
UCB	6.30 am	2.30 am	—	8.30 am	4.30 am
UCSB	8.30 am	4.30 am	2.30 am	—	6.30 am
DWR	2.30 am	8.30 am	6.30 am	4.30 am	—