Timing Experiments in the Internet

Benessa Defend, Final Report 2004 DMP

August 5, 2004

I. Description and Importance of the Research

The overall objective of the research was to conduct experiments on the Internet to understand message latency. These experiments will lead to the validation of the partial synchrony model on overlay networks on the global Internet. Cristian and Fetzer have previously validated their Timed Asynchronous Model [1] within a LAN environment, and this research sought to extend the application of their model.

The partial synchrony model imposes fewer restrictions than the fully synchronous model and more restrictions than the fully asynchronous model. Under the purely synchronous model, bounded and known processing speed, delivery delays and difference among local clocks can be assumed. Under the purely asynchronous model, these times are either unbounded or their bounds are unknown. Therefore, it is impossible to distinguish between processes which are slow and those that have failed. Strong assumptions of the synchronous model do not hold for many practical distributed systems, such as the Internet, and the assumptions of the asynchronous model are too weak for practical implementation. The assumptions of the Timed Asynchronous Model may be practically realizable. This model requires that local services are timed, that processes have access to local hardware clocks having bounded drift rates from real-time, that processes can have crash or performance failures, and that the communications system can have omission or performance failures.

One benefit of partial synchrony is that protocols will be more scalable. Also, it will be possible to evaluate predicates which cannot be evaluated in the fully asynchronous model. For example, since the partial synchrony model allows for the existence of synchronized clocks, predicates which refer to time or to intervals of time can be evaluated.

The development of a valid partial synchrony model will permit the development of new algorithms and internet-scale services which are not realizable under purely synchronous systems and not currently possible under purely asynchronous systems. It will also enable the development of scalable monitoring and predicate-evaluation algorithms. A scalable on-line monitoring and predicate-evaluation service will enable future distributed systems to assess and respond to changes in distributed state. Overall, this will permit better realization of the full potential of the Internet by providing an increased level of assurance and allowing Internet-scale applications to reliably and efficiently assess global state in a timely manner.

II. Experiments

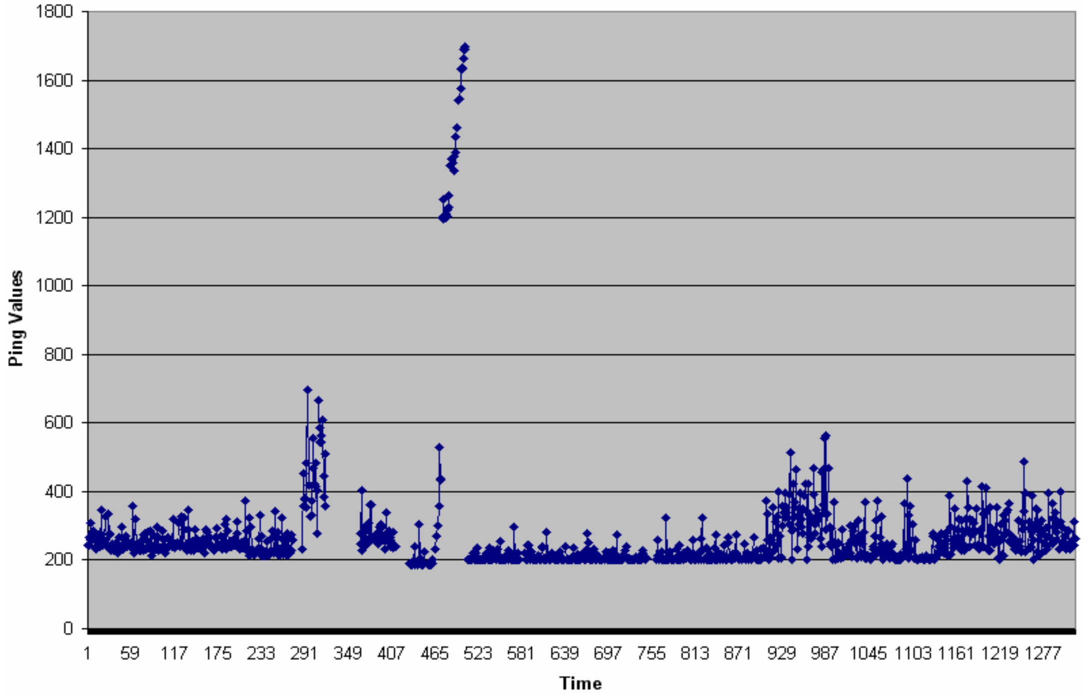
The first objective was to develop the necessary technical skills to conduct and execute network experiments. These skills included the use of the PlanetLab and Scriptroute infrastructures and network programming. Background reading was also necessary to understand the context of the research. The next objective was to develop an experiment plan, conduct experiments, and summarize the results in order to provide an initial understanding of real internet timing behavior.

I learned network programming by running and modifying programs from a textbook and then writing my own programs. For example, I wrote my own TCP ping server and client. My ping server listened on a well-known port for client requests. It forked a child to respond to the request while the original server process continued listening. The child called a function that printed the host name and the output of the uptime command. The client pinged the server, read what came in on the connection and echoed it to standard output. If it couldn't connect, it retried a set number of times and slept a set number of seconds in between tries. I also wrote UDP client and server programs. Through writing these programs, I developed an understanding of TCP, UDP, and networking that I needed to conduct the experiments.

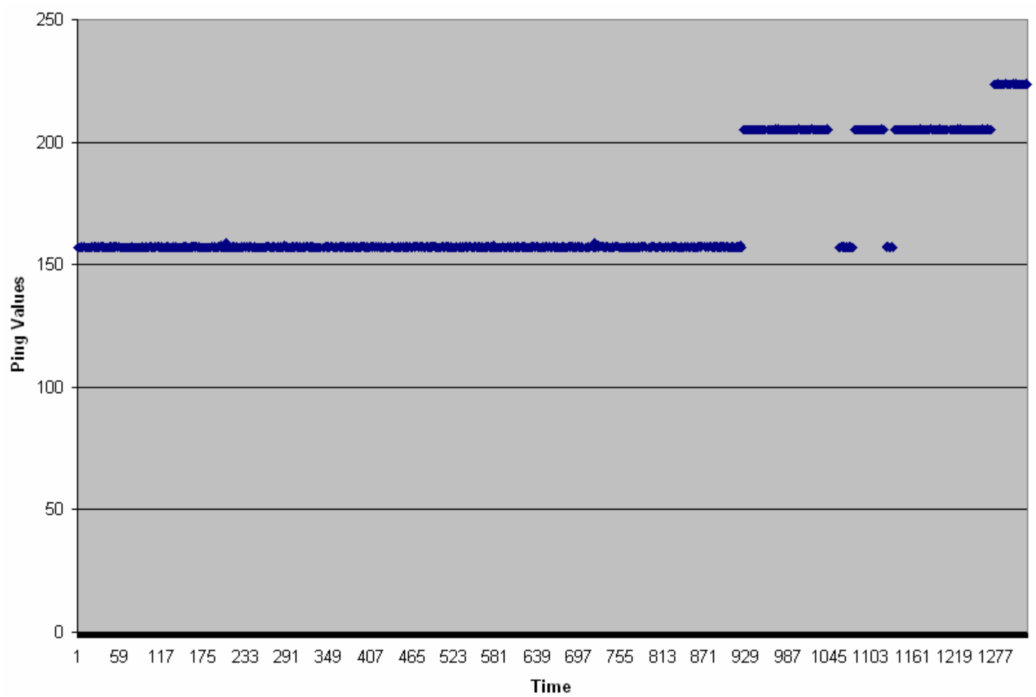
My experiment involved using the fping command to gather information about message latency between nodes. I sent a burst of ten pings to a node, waited one minute, sent another ten bursts, waited one minute, and continued this process for two hours. In order to gather diverse sets of data, I used a total of ten different nodes. Four were in the United States, and one in each of Taiwan, Australia, Tele-Avi, Denmark, China, and Greece. By choosing nodes that are geographically separated, I gathered data that was more representative of the Internet than if I had used nodes that were in the same geographic location.

I first attempted my experiments in the PlanetLab environment, but the fping command did not work correctly. Thus, I used Scriptroute to conduct the experiments. I chose a Scriptroute server that was in the United States. I wrote my fping scripts in the Ruby language and submitted them to the server. As a security precaution, the Scriptroute servers have a time limit on script execution. Since my scripts involved timed loops, they exceeded the time limit. To solve this problem, I wrote a Perl script that sent a Ruby script containing ten bursts to the Scriptroute server. Then the Perl script waited in a timed loop before submitting the script again. This process continued for two hours, thereby producing the same output had the Scriptroute server never imposed a time limit.

I ran my programs at different times of the day, including in the afternoon and early morning hours. I wrote programs to parse the data and delimit the fields so that the results could be imported into MS Excel. I calculated the minimum, maximum, mean, and standard deviation for each IP address during each time of the day (early morning, morning, afternoon). I also created histograms and line graphs of the data. I discovered that some of the connections were very stable, such as connections in the US. The node in China was very unstable, as was the node in Australia. The node in Taiwan exhibited periods of stability interrupted by timeouts.



The above line graph shows the values obtained by pinging a PlanetLab node in China from a node in the United States over the course of two hours. Spaces in the graph represent timeout values. The ping values are in milliseconds. Since the ping values are not close together, it shows that the node is not stable. The period around time 465-523 is especially unstable given the high ping values compared with the other values during different times.



The above line graph shows the values obtained by pinging a PlanetLab node in Taiwan from a node in the United States over the course of two hours. Spaces in the graph represent timeout values. The node is very stable from time 1-929 because the ping values are very close together. Then, a timeout occurs, and the ping values increase significantly. Timeouts occur again, followed by a return to the previous low ping values. Then, the ping values increase again, a timeout occurs, and they increase even more.

The results of my experiments established a framework for preliminary analysis of the partial synchrony model. My data also suggest that parts of the Internet display partially synchronous characteristics. For example, the line graph of China shows asynchronous behavior and the graph of Taiwan shows partially synchronous behavior.

III. Future Work

Plans for future work include continuation of experiments to gather information about message latency. This will lead to the evaluation and validation of the partial synchrony model for the Internet. These experiments will be continued by an undergraduate and a Ph.D. student at Wayne State University. The results will lead to publication and a final thesis dissertation.

IV. References

[1] F. Cristian and C. Fetzer. The Timed Asynchronous Distributed System Model, *Transactions on Parallel and Distributed Systems*, vol. 10, no. 6, pp. 642-657, June, 1999.