Characterizing Internet Performance to Support Wide-area Application Development

Gerco Ballintijn Maarten van Steen Andrew S. Tanenbaum

Vrije Universiteit Amsterdam, Faculty of Sciences,
De Boelelaan 1081a, NL-1081 HV Amsterdam, The Netherlands
{gerco,steen}@cs.vu.nl

Abstract

To help in our wide-area application development, we have done an informal study of the relation between wide-area latency, the number of routers, and geographical distance between Internet sites. We did this by performing ping and traceroute measurements between 19 sites distributed across the globe. Contrary to our expectation there is almost no correlation between distance, latency, and number of routers in the current Internet. To help further studies of these characteristics the problems of this pilot project are also described.

1 Introduction

When developing an application, it is important to have accurate knowledge of the characteristics of the environment in which the application is to run. This is especially true for wide-area applications since their performance (or lack thereof) will depend heavily on these characteristics. Obvious characteristics for wide-area networks are latency and bandwidth. Knowledge of these characteristics will be used not only when designing a wide-area application, but especially when testing the viability of the application.

For instance, as part of the Globe project [5], we are currently designing and implementing a wide-area location service [4]. The goal of this service is to track the current location of mobile objects in our wide-area distributed system. Our objects usually represent computer resources, but can also be used to model users. Every object in our system is identified by an object identifier. This identifier can be used to query the location service for the current location of an object. To validate the design of our location service, we want to measure its performance under various conditions. Since performing these kinds of measurements in the real Internet would be difficult at best, we have decided to use a simulated Internet.

We need, however, a good characterization of the performance of the current Internet if we want to test our location service on a simulated Internet. To make matters more difficult, we need a characterization of the complete Internet since our location service is expected to run everywhere in the Internet. We need to know the Internet characteristics of out-of-the-way places, like Islamabad (Pakistan) or Paramaribo (Suriname), not just the Internet covering the USA and Europe.

Unfortunately, network studies performed thus far have had a strong focus on bandwidth issues in the USA and Europe. For instance, the Network Weather Service [2], provides continues information on bandwidth and latency between sites in the USA. A similar commercial service is provided by Matrix Information and Directory Services [1]. Vahdat describes in [3] an experiment to measure latency and reliability between 43 sites in the USA.

We have therefore started to build a database of latency values covering the complete Internet. This database is essentially a matrix containing the latency between pairs of Internet sites. Since these values are of interest to anyone developing wide-area applications, we made them available to the public. These values are particularly useful for those implementing interactive wide-area applications. One aspect we are particular interested in is the correlation between geographical distance and latency. Our measurements therefore also include the geographical distance between sites.

The goal of this paper is three fold. First, we present our preliminary findings. Since this only a pilot project, conclusions should be drawn only tentatively. Second, we discuss the problems faced when performing a more fundamental study into latency in the Internet. Third, we place a call for further cooperation in extending our latency database.

This rest of this paper is structured as follows. Section 2 describes how we performed our latency measurements. In Section 3 we present some figures based on our preliminary findings. In Section 4 we describe the problems encounter during the project, and in Section 5 we draw our conclusion and present future work.

2 Experimental Setup

We used our personal connections to find Internet sites willing to cooperate. Previous (unrelated) attempts at finding cooperating sites by asking help on Usenet newsgroups and mailing lists turned out to be useless. People do not seem to respond unless asked directly. Luckily, we managed to find sites willing to cooperate on every continent.

We chose the Internet sites we asked in such a way as to cover the earth as evenly as possible. To keep the project manageable we decided to use 3 sites per continent, if possible. The cooperating Internet sites are listed in Table 1. We found in total 19 cooperating sites. One significant problem is our coverage of the African continent. Despite several attempts, we managed to obtain only the site in South Africa.

To ease the burden on the cooperating sites, we made the latency measurement as simple as possible. To measure the latency between sites, we used the standard ping utility as found on UNIX and Windows systems. We provided every site with a shell script. This shell script would measure the latency between the site and every other site, and would put the results in a file. All files would then collected and analyzed.

The ping utility uses the ICMP protocol to measure the the round-trip time of Internet packets between two sites. An originating site sends a ping request to the destination site, the destination responds by sending a ping reply. Every site sends 25 ping packets, and averages the round-trip times. If we assume a symmetrical Internet connection between two sites, the latency is half the round-trip time. Since ping values were not always available we also used the Internet traceroute tool to measure latency. We also used this tool to determine the number of routers between two sites.

3 Discussion of Results

Since the matrices containing the raw data (latencies, distances, and hop counts) are too large to present here, we will show a number of figures instead. The figures capture the most interesting

Continent	Country	City
North America	USA	San Francisco
	USA	Boulder
	USA	New York City
	Iceland	Reykjavik
Central America	Nicaragua	Managua
South America	Suriname	Paramaribo
	Rio de Janeiro	Brazil
	Argentina	Buenos Aires
Europe	Amsterdam	The Netherlands
	Portugal	Lisbon
	Moscow	Russia
Middle East	Israel	Jerusalem
	Pakistan	Islamabad
Far East	Shanghai	China
	Japan	Tokyo
	Indonesia	Bandung
Australia	Australia	Adelaide
	Christchurch	New Zealand
Africa	South Africa	Cape Town

Table 1: Cities of the cooperating sites

aspects found. The raw data itself can be found via the home page of the Globe project [1].

Figure 1 shows a scatter plot of the geographical distances versus the measured round-trip times. We expected to find some correlation between distance and latency. This figure, however, clearly shows there is no correlation.

Figure 2 shows a scatter plot of the geographical distances versus the number of routers between two sites. We expected to find also a strong correlation between the geographical distances versus the number of routers between two sites. The figure, however, shows there is at best only a weak correlation.

Figure 3 shows a scatter plot of the number of routers versus the latency between two sites. Given the relatively large delay packets incur at routers, we expected also find a strong correlation between the number of routers versus the latency between two sites. The figure, however, shows again there is at best only a weak correlation.

Figure 4 shows a scatter plot of the round-trip time from site A to site B and vice versa. If Internet connections were completely symmetrical (and the measured values contained no errors), the figure would show only marks on the line $X = Y$.

4 Discussion of Experiment

Our project started as a way to get some insight into wide-area latency and its relation to geographical distance. However, in the course of time, it became clear that a proper study of wide-area latency

[1] `http://www.cs.vu.nl/globe/`

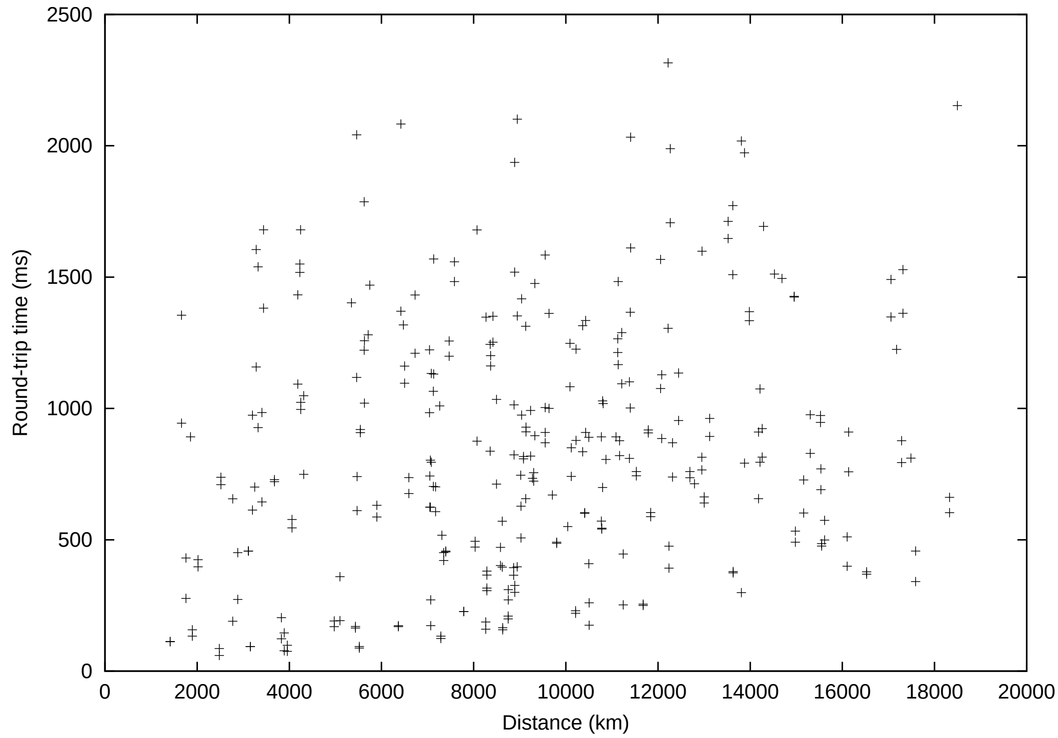

Figure 1: Plot of the distance and the round-trip time between two sites.

would be worthwhile. Such a new study would be needed to address the problems our project presented here.

The first problem concerns the Internet sites used. By using our personal connections, we were able to find sites, but most of these sites are educational. The Internet sites were not chosen at random, and the values thus clearly do not represent the whole Internet. However, the only sites likely to cooperate are educational sites.

The second problem concerns the way the measurement was performed. Since we did not want to burden (and scare away) our cooperating sites, we chose to let them perform a minimal amount work. Since measurement was performed only once, the results are strongly influenced by the time of measurement. A more thorough investigation would require regular measurement (e.g., every couple of hours) over a longer period of time (e.g., a week). Unfortunately, our experience shows us that automating these things, even over "compatible" UNIX systems, is far from easy.

A third problem concerns the conversion and analysis of raw measurement data. Even though the ping and traceroute utilities perform the same functionality across the various UNIX and Windows systems, their output formats varied greatly. This resulted in a lot of manual labour. If a new study is to use even more sites, it would be useful to spend time automating the conversion.

5 Conclusions and Future Work

Since the figures presented here are based on a small data set, strong conclusions can not be drawn. However, the results do suggest that geographical distance and number of routers play an insignificant part in wide-area latency. Latency is apparently caused by other factors. A more comprehensive study

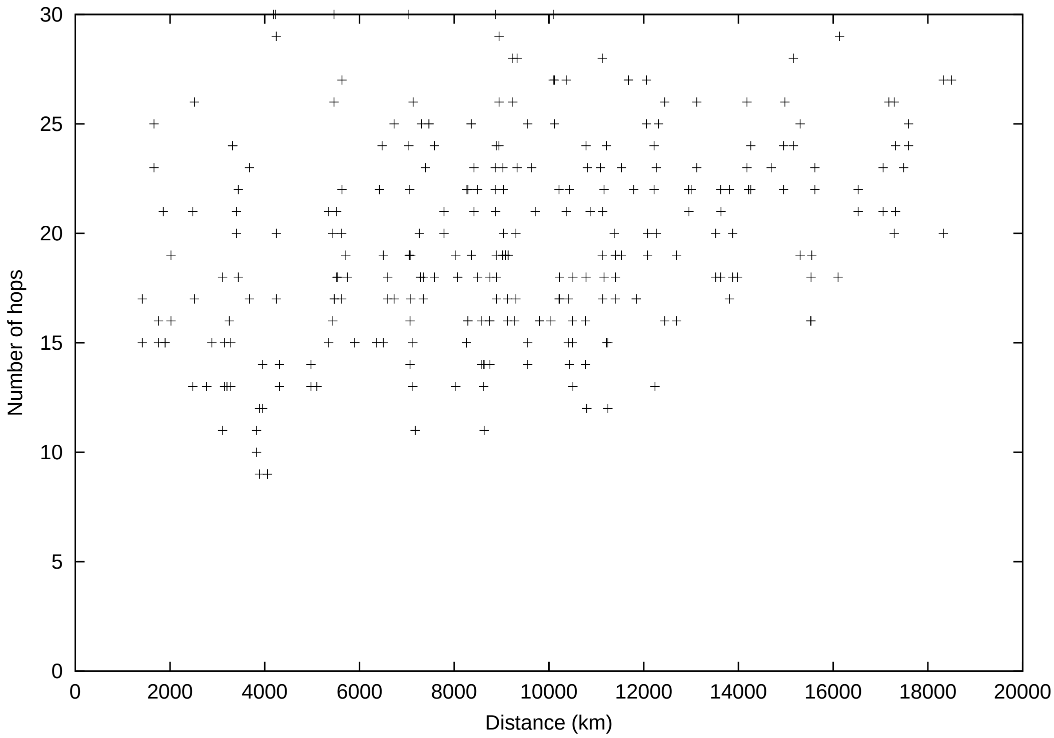

Figure 2: Plot of the distance and the number of routers (hops) between two sites.

needs to be performed to confirm the lack of correlation, and show it is not caused by for instance measurements errors.

It is our intention to continue the research of Internet characteristics as part of the Globe project. Specifically we want to perform the more thorough study suggested in this paper. We like to ask those people who are willing to collaborate with us on this topic to contact us.

Acknowledgements

We like to thank the following people for helping us gather and analyze the data, Jasmin Amidzic, Stefan Baltus, Dan Crawl, Leendert van Doorn, Arnar Fririksson, Shen Fuke, Cornelio Hopmann, Imam Kistijantoro, Amar Lior, Bruce McKenzie, Aart Middeldorp, Hugo Miranda, Andrew Ross, Ferooz Sekandarpoor, Claudio Tantignone, Marc Welz, Al Woodhull, Mikhail Zouskov, and Frank Niessink.

References

[1] Matrix Information and Directory Services, Inc. http://www.mids.org/.

[2] The Network Weather Service. http://nws.npaci.edu/NWS/.

[3] M. A. Vahdat. *Operating System Services for Wide-Area Applications*. PhD thesis, University of California, Berkeley, 1992.

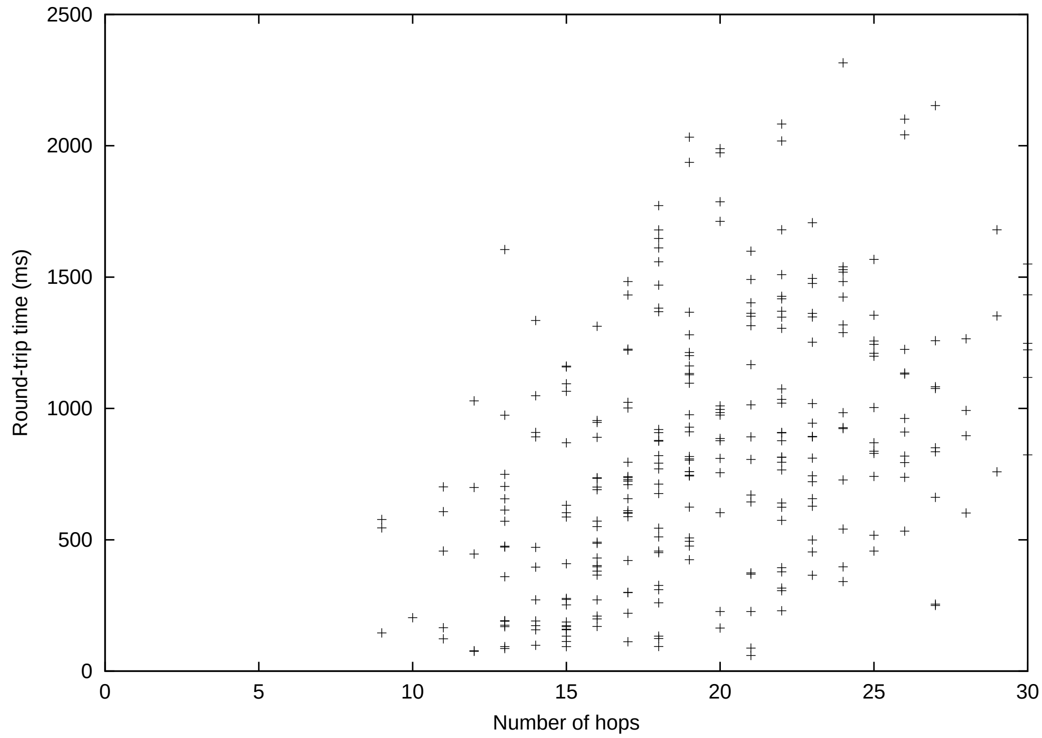

Figure 3: Plot of the number of routers (hops) and the round-trip time between two sites.

[4] M. van Steen, F. J. Hauck, P. Homburg, and A. S. Tanenbaum. "Locating Objects in Wide-Area Systems." *IEEE Communications Magazine*, pp. 104–109, Jan. 1998.

[5] M. van Steen, P. Homburg, and A. S. Tanenbaum. "Globe: A Wide-Area Distributed System." *IEEE Concurrency*, pp. 70–78, Jan. 1999.

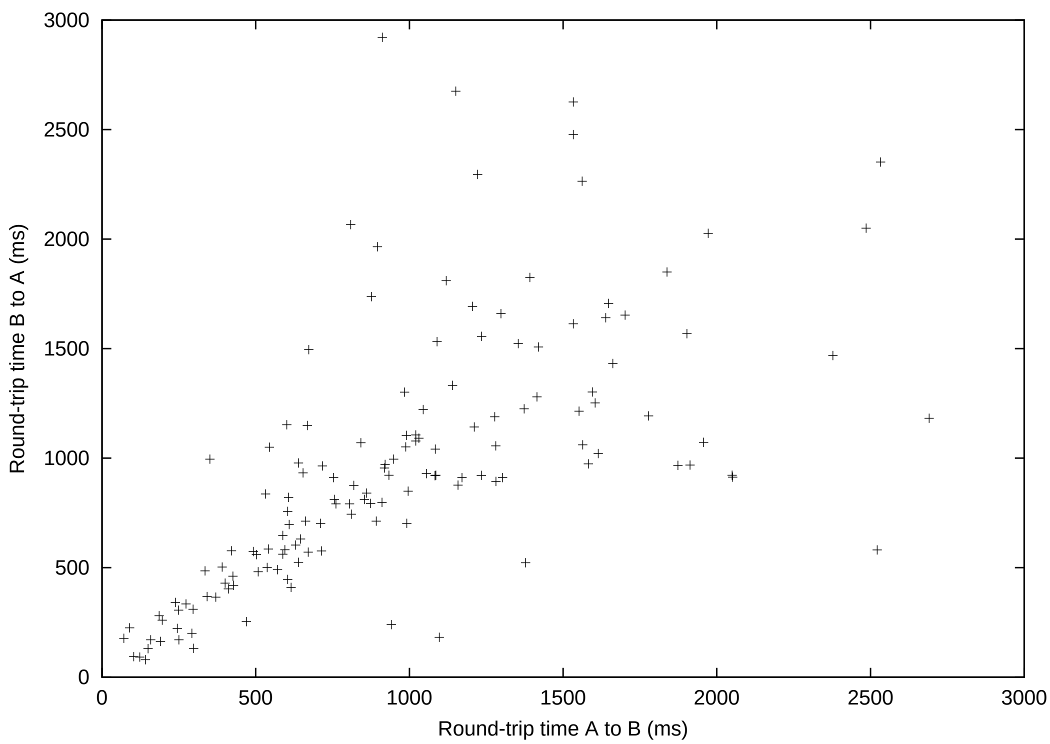

Figure 4: Plot of the round-trip time from site A to site B and vice versa.