A Study of Internet Round-trip Delay*

Anurag Acharya   Joel Saltz

UMIACS and Department of Computer Science

University of Maryland, College Park 20742

{acha,saltz}@cs.umd.edu

Abstract

We present the results of a study of Internet round-trip delay. The links chosen include links to frequently accessed commercial hosts as well as well-known academic and foreign hosts. Each link was studied for a 48-hour period. We attempt to answer the following questions: (1) how rapidly and in what manner does the delay change – in this study, we focus on medium-grain (seconds/minutes) and coarse-grain time-scales (tens of minutes/hours); (2) what does the frequency distribution of delay look like and how rapidly does it change; (3) what is a good metric to characterize the delay for the purpose of adaptation. Our conclusions are: (a) there is large temporal and spatial variation in round-trip time (RTT); (b) RTT distribution is usually unimodal and asymmetric and has a long tail on the right hand side; (c) RTT observations in most time periods are tightly clustered around the mode; (d) the mode is a good characteristic value for RTT distributions; (e) RTT distributions change slowly; (f) persistent changes in RTT occur slowly, sharp changes are undone very shortly; (g) jitter in RTT observations is small and (h) inherent RTT occurs frequently.

1   Introduction

Several recent research efforts have focussed on adapting to variation in Internet round-trip delay [1, 4, 5, 7, 16]. Amsaleg et al [1] propose an adaptive strategy for executing relational queries over the Internet; Carter&Crovella [4] propose an adaptive scheme for selecting between multiple servers offering the same data object; Chankhuthod et al [5] propose an adaptive scheme selecting between multiple caches for a data object; Etzioni et al [7] propose and analyze algorithms to optimize multi-site information gathering based on information about round-trip delay and dollar costs; Ranganathan et al [16] propose program mobility as a way to adapt to changes in Internet round-trip delay.

Performance of such schemes depends to a great extent on the rate and manner in which Internet round-trip delay varies. There have been several previous studies of Internet round-trip delay [12, 13, 15, 14, 17]. The original study by Mills [12] was done back in 1983; Internet has changed considerably since then in terms of both number of hosts and traffic volume. Sanghi et al [17] did a fine-grain study (one probe every 39.6 ms) of three links over an hour in 1992; Pointek et al [14] repeated this study for seven links in 1996. These studies focused on packet loss, duplicates, reorders and possible patterns in the occurrence of large delays. They provide excellent information on these issues but provide only summary information (min, max, mean, standard deviation) about round-trip delay. Two other limitations of these studies relate to site selection: (1) the number of sites is small and (2) commercial sites (esp popular sites) are not represented and foreign sites are under-represented (only one international link is studied). Mukherjee [13] presents an analysis of dynamics of round-trip delay over one day for three links – one regional, and two transcontinental. Quarterman et al [15] present results from a long-term study of a large number of links. Their pinging period is once every six hours and provides very coarse-grain information.

In this paper, we present the results of a study of Internet round-trip delays over ninety links. The links chosen include links to frequently accessed commercial hosts as well as well-known academic and foreign

---

hosts (see section 2.2 for details). Each link was studied for a 48-hour period. We attempt to answer the following questions: (1) how rapidly and in what manner does the delay change – in this study, we focus on medium-grain (seconds/minutes) and coarse-grain time-scales (tens of minutes/hours); (2) what does the frequency distribution of delay look like and how rapidly does it change; (3) what is a good metric to characterize the delay for the purpose of adaptation (as in [1, 4, 5, 7, 16]).

Our conclusions are: (a) there is large temporal and spatial variation in round-trip time (RTT); (b) RTT distribution is usually unimodal and asymmetric and has a long tail on the right hand side; (c) RTT observations in most time periods are tightly clustered around the mode; (d) the mode is a good characteristic value for RTT distributions; (e) RTT distributions change slowly; (f) persistent changes in RTT occur slowly, sharp changes are undone very shortly; (g) jitter in RTT observations is small and (h) inherent RTT occurs frequently.

We describe our experiments in section 2. We mention the design goals, the site selection criteria, the sites selected, the trace-collection mechanism and the periods over which the study was conducted for the different links. Using the data collected during these experiments, we computed a set of metrics to help answer the questions mentioned above. We describe the metrics and how they were computed in section 3. Following this, we present the results of our analyses. We summarize our results in section 12.

2 Experiments

The basic design of our experiments was simple. One host in each link was selected to be the *pinger* and the other was the *pingee*. The pinger periodically sends a probe packet to the pingee which echoes the packet as soon as it can. Each packet contains a timestamp and sequence number. The pinger uses the timestamp to compute the round-trip delay and the sequence number to detect packet-loss. The rest of this section fills in the details. The first subsection discusses the choice of the protocol and the pinging period and the second subsection describes how the links included in this experiment were selected.

2.1 Pinging procedure

Our design of the pinging procedure was motivated by five considerations. First, we were interested in medium-term (seconds/minutes) to long-term (tens of minutes/hours). Second, we wanted to keep the network overhead due to probe packets small – this was important as we planned to run these experiments over multiple days. Third, we wanted to avoid the possibility of flooding the network with probe packets as the behavior under such a condition is unlikely to be representative of the unprobed state of the network. Fourth, we wanted to include a wide variety of hosts in our study – including popular commercial web servers. Fifth, we wanted to place the smallest possible computation load on the pingee – this is particularly important as many of the hosts we wanted to include in our study were extremely busy machines.

On the basis of these considerations, we chose ICMP as the network protocol, 64 bytes as the packet size and one second as the pinging period. Choosing ICMP had two advantages. First, there is no need to obtain an account on the hosts being pinged, only those that are doing the pinging. This allowed us to include a wide variety of hosts in our study. Second, of the commonly available protocols, it places the least computational load on the hosts being pinged. Choosing a 64-byte packet and a one-second pinging period allowed us to limit the network overhead and to avoid the possibility of flooding the network.

ICMP has been previously used by several researchers and system administrators to measure Internet round-trip time (RTT) [4, 10, 11, 12, 15, 18]. It was also used by Merit Network Inc to measure internodal latency in the NSFNET T1 backbone [6].

The tracing was done between 8pm on the fourth of June 1996 and 11pm on the ninth of June 1996. This includes three weekdays, four weeknights, and all of one weekend. Each pingee host was pinged from two different sites - once during the week and once over the weekend. Each trace was collected over a 48-hour period.

2.2 Host selection

Our host selection criteria were derived from four considerations. First, we wanted to include a wide variety of hosts including popular commercial web servers, academic and government web servers and well-known foreign hosts. Second, since a majority of the Internet hosts are in the US and a large fraction of Internet traffic is between US hosts, we wanted to bias the selection criteria to include a significantly larger number of US hosts. Third, we wanted to measure RTT to individual hosts and not host groups. This is relevant since several busy web servers (for example, NCSA) use a single host name for a group of servers and use *round-robin DNS* [2] in an attempt to balance load between them (see [9] for application of this technique for the NCSA web server). Fourth, we wanted the sites to be spread out in a geographical sense.

We made the selections in the following way. We selected 44 pingee hosts – 15 popular commercial web servers in the US, 14 popular academic/government web servers in the US and 15 well-known international web servers. The commercial web servers were selected from the list of popular web servers made available by *Web21* at http://www.100hot.com. This included web search engines, sports information servers, news servers and so on. Among the hosts in this list, we used geographical location as a secondary selection criterion. Given the popularity of the NCSA web server, we have included it in the group of commercial sites. We selected the academic and government hosts using one of two measures: (1) popularity as shown by appearance in the *100hot* list or (2) academic fame. We selected the foreign sites primarily by name recognition. We selected at most one site from every country. For the sites that use round-robin DNS for load-balancing, we used numeric address of one of the servers to avoid the server-switch problem. The list of pingee hosts is provided in Table 1.

For the pinger hosts, we selected four locations – University of Maryland, Carnegie Mellon University, Argonne National Laboratory and the Goddard Space Flight Center. From each of these locations, we used one or more hosts as pingers in the experiments.

3 Metrics computed

We computed two kinds of metrics: aggregate metrics that summarized different properties over different time periods and temporal metrics that characterized the rate and the manner in which these properties change with time. We computed aggregate metrics over six time periods: one minute, five minutes, ten minutes, fifteen minutes, half an hour and one hour.

We computed eleven aggregate metrics: minimum, maximum, mean, standard deviation, mode, mode-fraction, 12.5-87.5-percentile, runlength and spike-isolation. We computed the minimum to help estimate the inherent RTT for the link and to determine how frequently this occurs. We computed the maxmimum to get an idea of the range of variation. The distribution of RTT for most time periods was unimodal and asymmetric with a long tail and a well-defined mode (see Figure 1 for an example and sections 5,6,7 for details). Therefore, we used the mode as the measure of central tendency. To account for errors in the measurement process, we computed *modefraction*, that is the fraction of observations that lie within an error window of the mode. The error window used was 10 ms or 10% of mode whichever is higher. As a measure of dispersion, we used 12.5-87.5-percentiles. This computes the range of RTT which covers 75% of the observations in the period of interest. We computed mean and standard deviation for comparison purposes. We computed average *runlength* as a measure of the jitter in the observations. By runlength, we mean the continuous period for which the RTT remains within a certain window. We computed the *spike-isolation* metric to determine if sharp changes in RTT observations were transient or persistent (see section 9 for details).

We computed two sets of temporal metrics. The first set measured the rate of change of aggregate metrics across successive time periods as well as at different levels of resolution. The second set measured the rate of change of distribution. Our analysis indicated that a large fraction of RTT observations are usually in a tight cluster around the mode and the mode is a good characteristic value for an RTT distribution (see section 8 for details). Therefore, we used the mode to characterize the distribution of RTT in any given period and the average runlength of the mode to determine its variability.

| Host number	Host name/number
1	`home.netscape.com (www14.netscape.com)`
2	`notme.ncsa.uiuc.edu (141.142.3.76)`
3	`a2z.lycos.com`
4	`www.altavista.digital.com`
5	`www.yahoo.com`
6	`www.microsoft.com`
7	`java.sun.com`
8	`www.sgi.com`
9	`www.best.com`
10	`www.opentext.com`
11	`www.sportsline.com`
12	`espnet.sportszone.com`
13	`pathfinder.com`
14	`www.well.com`
15	`us.imdb.com`
16	`www.cs.cmu.edu`
17	`sunsite.unc.edu`
18	`lcs.mit.edu`
19	`www.cs.umd.edu`
20	`cesdis.gsfc.nasa.gov`
21	`centro.soar.cs.cmu.edu`
22	`ink4.cs.berkeley.edu (204.161.74.8)`
23	`www-flash.stanford.edu`
24	`softlib.rice.edu`
25	`www.cs.wisc.edu`
26	`www.cs.utexas.edu`
27	`www.cs.washington.edu`
28	`www.cs.uiuc.edu`
29	`lanl.gov`
30	`www.inria.fr`
31	`www.centro.com.hk`
32	`www.monash.edu.au`
33	`sunsite.wits.ac.za`
34	`www.ac.il`
35	`www.nec.co.jp`
36	`konark.ncst.ernet.in`
37	`www.diku.dk`
38	`www.di.unito.it`
39	`dcc.unicamp.br`
40	`www.docs.uu.se`
41	`www.hensa.ac.uk`
42	`koi.www.online.ru`
43	`anon.penet.fi`
44	`www.metu.edu.tr`

Table 1: List of pingee hosts
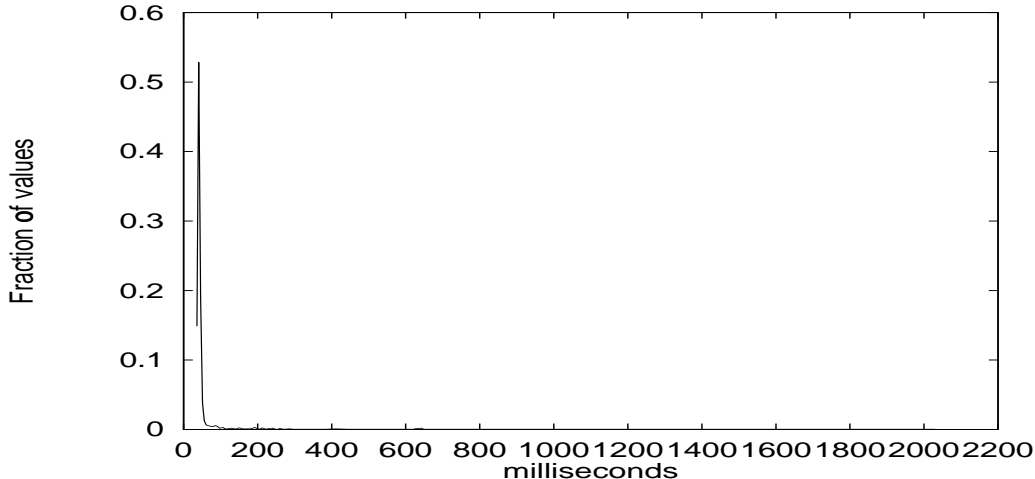
Figure 1: Sample RTT distribution. This distribution is for RTT between `baekdoo.cs.umd.edu` and `notme.ncsa.uiuc.edu` between noon and 1pm on the 5th June, 1996. The RTT values have been binned using 10 ms bins.

4 There is large temporal and spatial variation in RTT

Figure 2 presents the estimated inherent RTT for all the links traced. We assume that the minimum RTT over a two day period is a reasonable estimate for the inherent RTT of the link. In Figure 2, the links are sorted in the order of estimated inherent RTT. This order will be used throughout the rest of this paper. The (unsurprising) conclusion is that there is large spatial variation in RTT across different links.

Figure 3 presents the same data in three graphs – one for every host category (commercial, academic and foreign). The links within each category are sorted in increasing order of estimated inherent RTT. These graphs show step increases at various points. By inspection of the list of hosts, we found that these steps in RTT can be attributed to steps in the geographical distances corresponding to the links. To illustrate this, we examine Figure 3 (a). The commercial hosts were pinged from two sites – the University of Maryland and Goddard Space Flight Center, both of which are on the Eastern seaboard. Based on geographic distance, we can partition the group of commercial hosts into three subgroups – hosts in the Northeast, hosts in the Midwest and the South and hosts on the West coast and in Canada. Inspection of the graph in Figure 3 (a) shows that the points corresponding to the hosts in the Northeast occur on the leftmost part of the graph; the points corresponding to hosts in the Midwest and the South occur in middle region (in between the two steps) and the points corresponding to the hosts on the West coast and Canada occur in the rightmost part of graph. The step increases in the other two graphs can be similarly correlated with steps in geographical distance.

In the graphs corresponding to all three categories, we note that ratio of the smallest RTT to the largest RTT is similar (about 7-9) [1]. The magnitude of the variation, however, is smallest for the commercial hosts and the largest for foreign hosts.

Figure 4 (a) presents variation in RTT over individual links. The variation is measured as the range of values, that is $(max\_rtt - min\_rtt)$. Figure 4 (b) presents the same data after normalizing it using the minimum value, that is $(max\_rtt - min\_rtt)/min\_rtt$.

Common experience and previous studies indicate that temporal variation over a link varies with the time of the day. To quantify this effect for the links under study, we computed the normalized temporal variation (using $(max\_rtt - min\_rtt)/min\_rtt$ as the measure) for different four-hour periods during the day. Results are presented in Figure 5 and show that while there is a significant difference between the amount of variation during different periods, large temporal variation occurs throughout the day. Even in the most quiescent periods (4am-8am and 8pm-midnight), the range/minimum ratio is as large as 28.

---

[1] Ignoring the extremely small values at the left end of Figure 3(b) which correspond to links within the same site or links
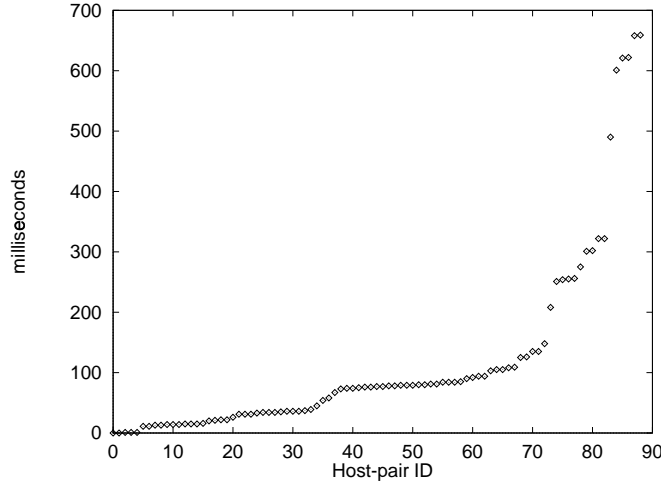
Figure 2: Estimated inherent RTT for all traced links. Links have been sorted in increasing order of estimated inherent RTT.
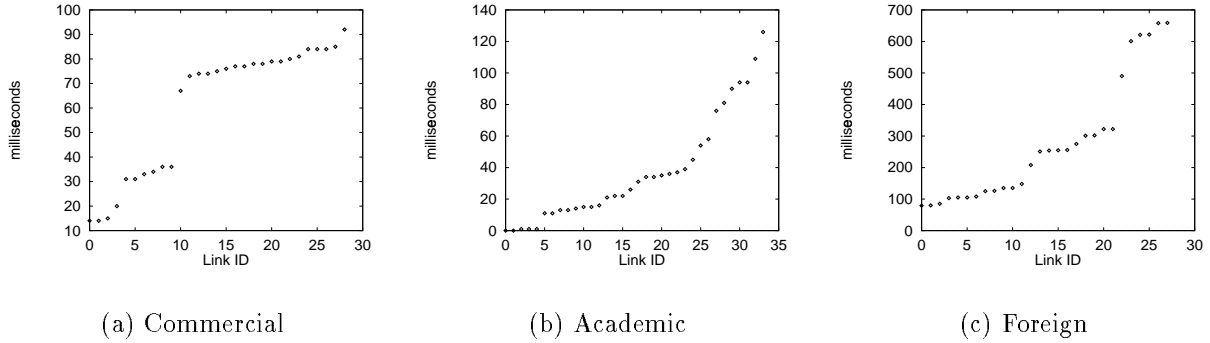

(a) Commercial   (b) Academic   (c) Foreign

Figure 3: Estimated inherent RTT for each host category. Links within each category are sorted in increasing order of estimated inherent RTT.
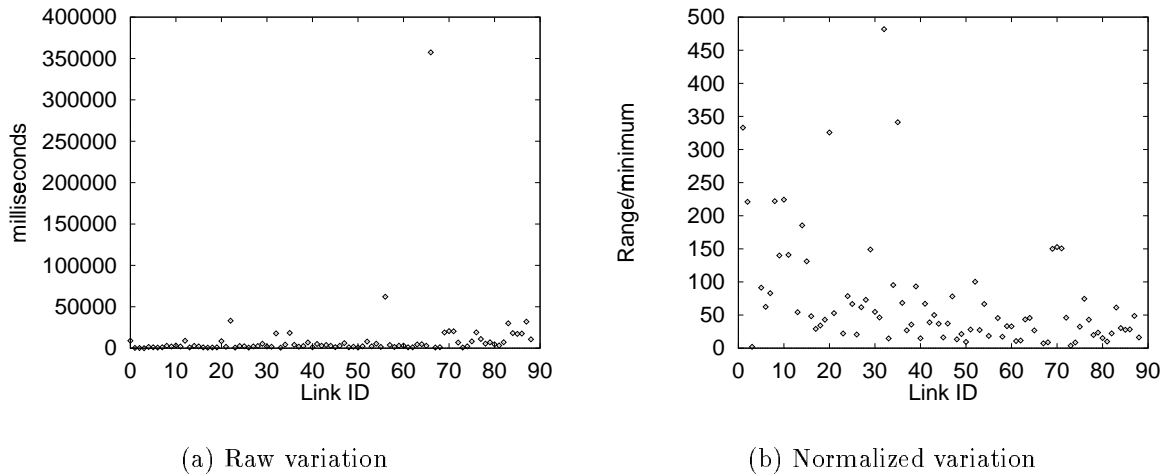

(a) Raw variation   (b) Normalized variation

Figure 4: Variation in RTT over individual links. Graph (a) plots $(max\_rtt - min\_rtt)$ as the measure of variation, graph (b) uses the $(max\_rtt - min\_rtt)/min\_rtt$ as the measure. Six outliers have been removed from graph (b) to allow a better scale along the y-axis.

6

(a) midnight-4am
(b) 4am-8am
(c) 8am-noon
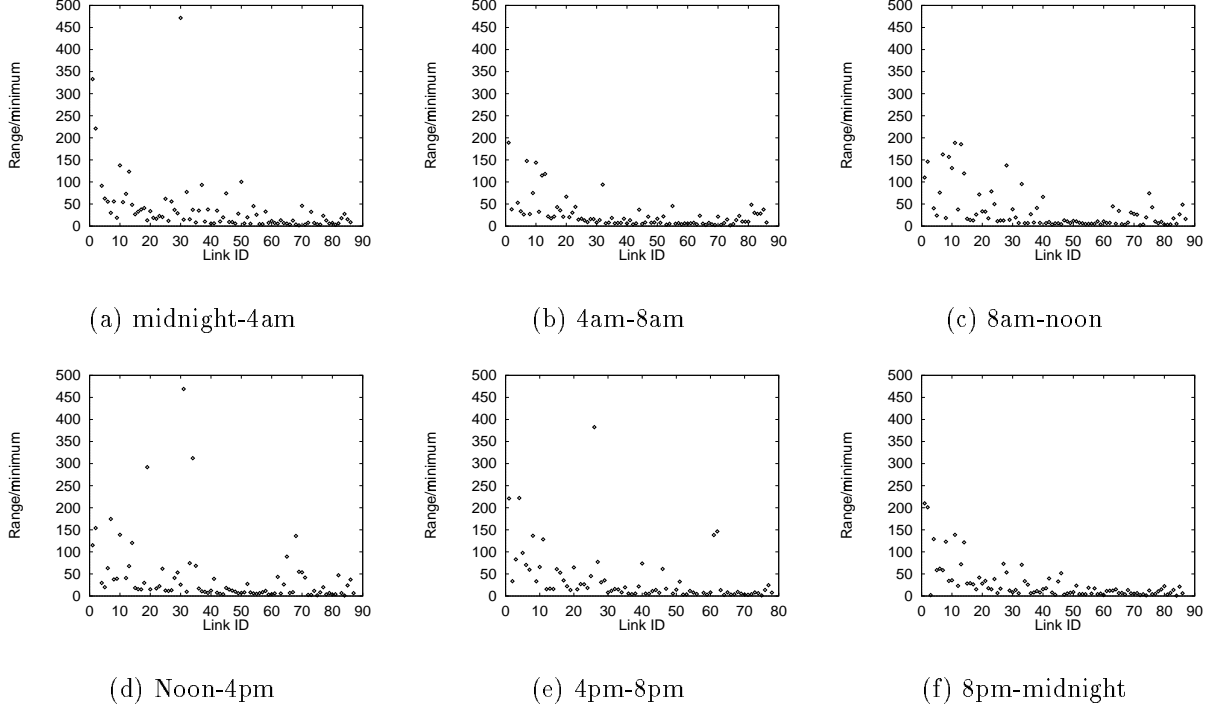

(d) Noon-4pm
(e) 4pm-8pm
(f) 8pm-midnight

Figure 5: Normalized per-link variation in RTT over different times in a day. Outliers have been removed from all these graphs: two from (a), two from (b), two from (c), four from (d), three from (e), and one from (f). The mean values for the different periods are: (a) 39, (b) 29, (c) 34, (d) 42, (e) 38 and (f) 28.

5   RTT distribution has a long tail

Given the large values of normalized range (i.e. *range/min*) shown in Figures 4 and 5, we expected the distribution of RTT to be long-tailed. Visual inspection of several histograms indicated as much (for example, Figure 1). To see if this property holds over different time-scales and different links, we computed two measures. The first measure was the ratio of the total range and the shortest range that contains 75% of the values. The second measure was the ratio of the range and tail (if any) on the left hand side (i.e. $range/(mode - min)$). The first measure indicates the sparseness of the tail. Combined with the normalized range data presented in Figure 5, the second measure indicates whether the tail is present on both sides of the mode or only on one side of the mode. We computed both measures over six time-scales between a minute and an hour. For each time-scale and for each link, we computed the average value of both measures.

Figure 6 presents the first measure for six time-scales between one minute and one hour. It shows that most RTT observations are localized in very small portions of the range of observations and that a large part of the range contains few values. As can be seen from the graphs in the figure, this behavior occurs at many time-scales. Even at the finest resolution of one minute, the average value of the measure is 18, that is 1/18th of the range of observations contain 75% of the observations and the tail covers 17/18th of the range. For coarser resolutions, the tail is even longer.

Figure 7 presents the second measure for six time-scales between one minute and one hour. It shows that the the mode usually lies at the extreme left hand side of the distribution and that tail is almost entirely on the right-hand side. At the finest resolution (one minute), the distance of the mode to minimum is, on the average, 1/18th of the range; 17/18th of the range lies on the right hand side of the mode. The one-sidedness is greater for coarser resolutions. These two measures, together, show that the distribution of RTT is usually skewed.

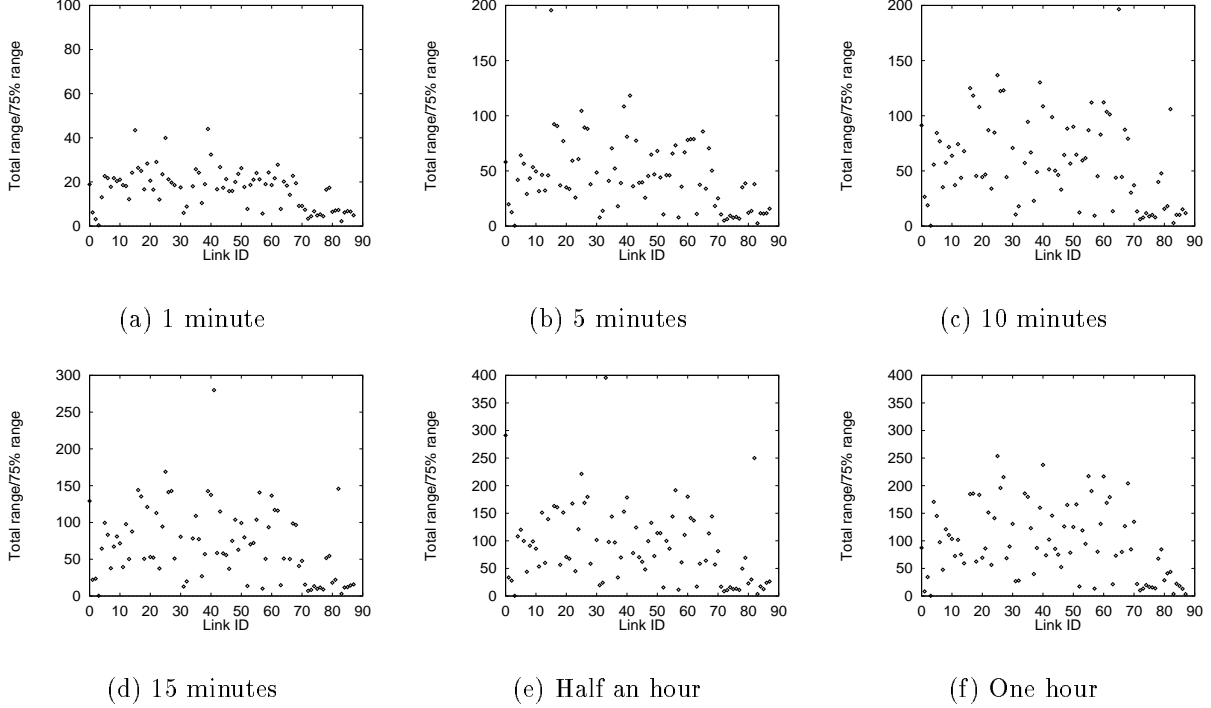---

between extremely close sites.

7

Figure 6: Measure of sparseness of the tail in the RTT distribution. For each link and at each time-scale, this measure is computed as the average ratio of the total range and the shortest range of observations that contains 75% of the observations. Note that the scale along the y-axis is not the same for all graphs. Three outliers have been eliminated from each graph. The median values for the plots are: (a) 18, (b) 41, (c) 56, (d) 64, (e) 86 and (f) 87.

6 Mode often dominates RTT distribution

As shown in the previous sections, RTT values tend to be localized. Visual inspection of many histograms indicates that a relatively short window around the mode contains a large fraction of all observations (for an example, see Figure 1). To measure the degree of *mode-dominance* (we refer to this localization as mode-dominance) across several time-scales (and all links), we computed the following measure:

1. For each time-scale, we collected all available traces at that time-scale. For example, for the one minute time-scale, we extracted all the per-minute traces for all the links.

2. For each extracted trace, we computed its histogram.

3. For each histogram, we computed the fraction of values that lie within the *mode-window*. Mode-window was selected to be either *10 ms or 10% of the mode*, whichever is higher. [2]

4. For each time-scale, we computed the fraction of traces for which the mode-window contained at least *75%* of the observations in that period. This fraction is used as the measure of mode-dominance at that time-scale.

Figure 8 presents the mode-dominance metric for all links at six time-scales. The high density of points in all graphs between 0.9 and 1.0 indicates that for a large number of links, a small window around the mode contains most of the values and that the mode (or the window around it) can characterize (or represent)

---

[2] Note that the timer resolution on most Unix systems is 10 ms.

(a) 1 minute   (b) 5 minutes   (c) 10 minutes

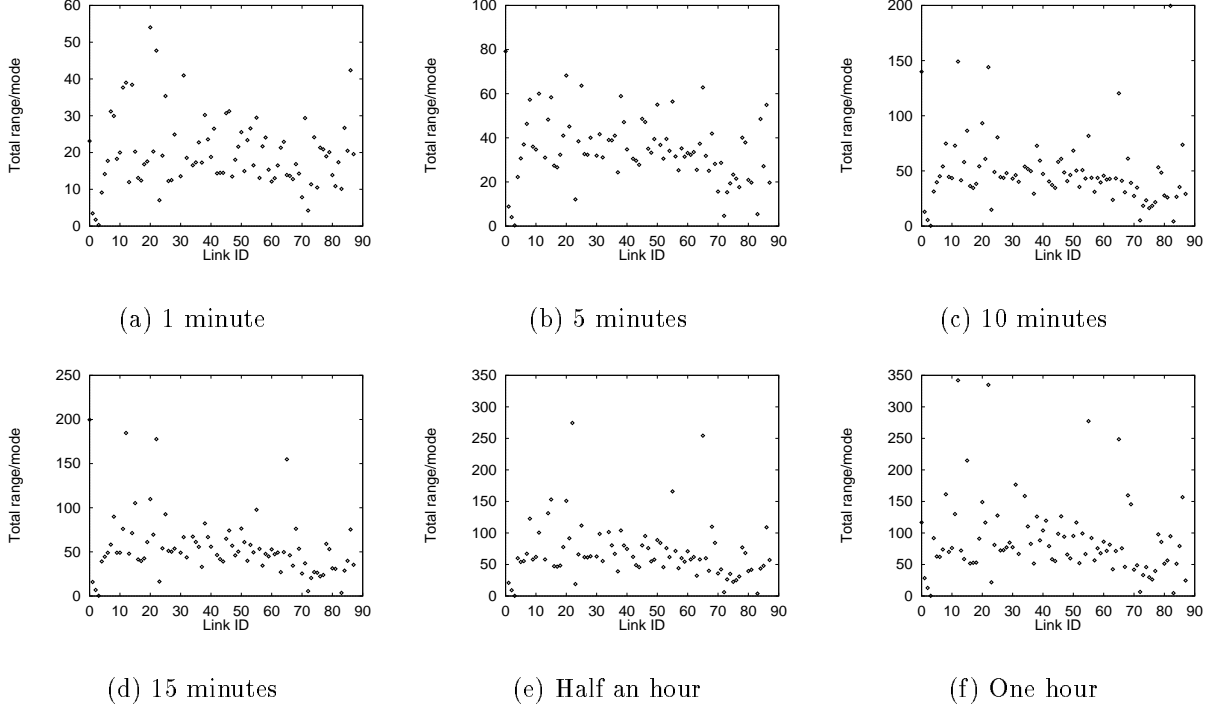(d) 15 minutes   (e) Half an hour   (f) One hour

Figure 7: Measure of one-sidedness of the tail. For each link and for each time-scale, this measure is computed as the average ratio of the total range and the distance between the mode and the minimum. Note that the scale along the y-axis is not the same for all graphs. Three outliers have been eliminated from each graph. The median values for the plots are: (a) 18, (b) 34, (c) 44, (d) 50, (e) 62 and (f) 76.

the distribution fairly well. As mentioned in the caption, about 70% of the links have a mode-dominance value greater than 0.5.

Table 2 lists the links whose mode-dominance is low at all time-scales. The goal of this list is to show that the mode-dominance property is, in some sense, invariant across several time-scales and is a property of the link. Since mode-dominance is a measure of how dispersed RTT observations are, this indicates that dispersion of RTT values is, to some extent, scale-invariant.

From the graphs in Figure 8, we note that a large fraction of the links with low values of mode-dominance occur on the right hand side. Recall that the links are sorted by estimated inherent RTT and that the links on the right hand side of the graphs correspond mostly to foreign hosts. To determine the degree of localization for different categories of hosts, we summarized the data by averaging the mode-dominance measure for each category. Figure 9 shows the average mode-dominance for each category across the time-scales used in Figure 8. It shows that for links within the US, the mode-dominance measure is usually at least 0.75 and that this is true across all the six time-scales studied. This implies that for 75% of all time periods (with length between one minute and one hour), RTT observations are clustered closely around the mode. For links to foreign hosts, the number is significantly smaller. Nevertheless, even for these links RTT observations form a tight cluster around the mode for over 40% of all time periods with length between one minute and one hour.

7   Distribution of RTT skewed in many cases

The presence of a one-sided long-tail usually indicates that the mean is larger than the mode. The magnitude of this skew depends on the values of the outliers and their frequency. As we have seen in the previous section, a relatively short window around the mode usually contains a large fraction of the RTT observations and the

| (a) 1 minute | (b) 5 minutes | (c) 10 minutes |

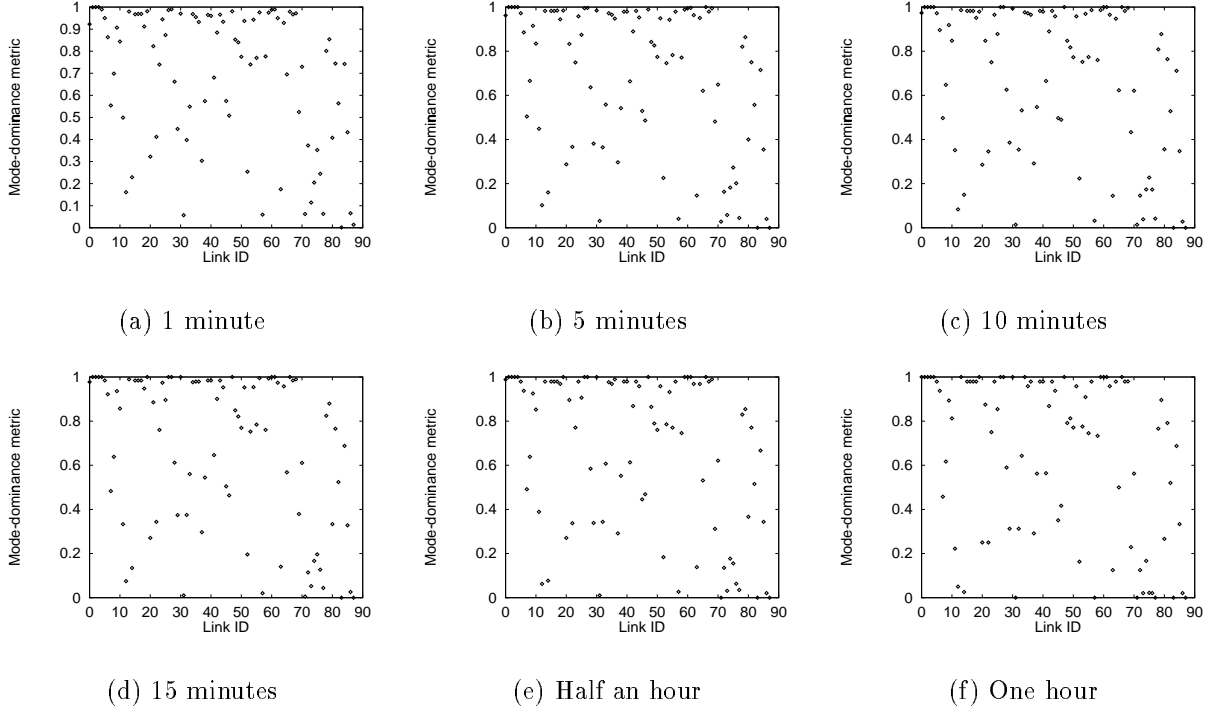| (d) 15 minutes | (e) Half an hour | (f) One hour |

Figure 8: Mode dominance at different time-scales. The process of computing the mode-dominance metric is described in text. Percentage of links with mode-dominance metric < 0.5 is: (a) 27%, (b) 28%, (c) 31%, (d) 30%, (e) 31% and (f) 31%.

Link number	Link
1	UMD-lanl.gov
2	UMD-www.di.unito.it
3	UMD-www.best.com
4	UMD-www.metu.edu.tr
5	UMD-www.nec.co.jp
6	UMD-sunsite.wits.ac.za
7	UMD-www.monash.edu.au
8	UMD-clone.mcs.anl.gov
9	UMD-www.ac.il
10	CESDIS-anon.penet.fi
11	CESDIS-konark.ncst.ernet.in
12	CESDIS-www.inria.fr
13	CESDIS-us.imdb.com
14	CESDIS-www.di.unito.it
15	CESDIS-www.hensa.ac.uk
16	CESDIS-centro.soar.cs.cmu.edu
17	CESDIS-clone.mcs.anl.gov
18	CESDIS-www.best.com
19	CMU-lcs.mit.edu
20	CMU-sunsite.unc.edu
21	CMU-www.cs.umd.edu
22	ANL-anon.penet.fi

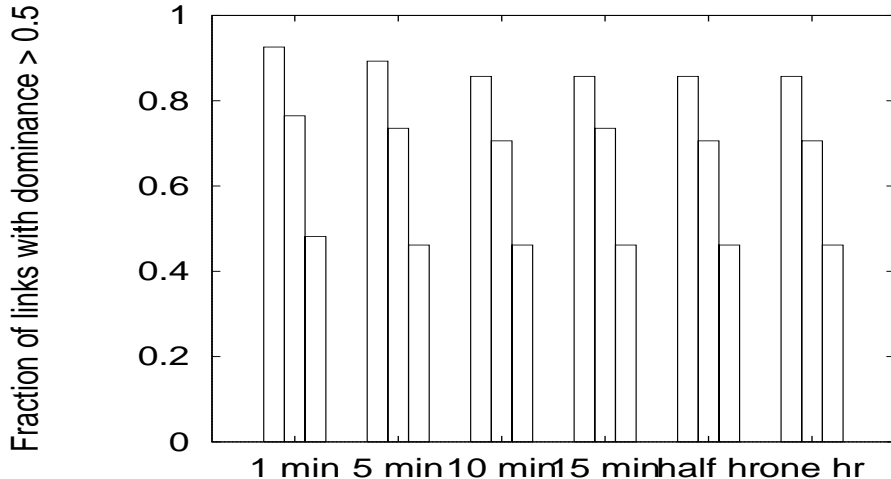Table 2: List of links whose mode-dominance metric was below 0.5 at all time-scales.

Figure 9: Fraction of links in each category (commercial, academic, foreign) with mode dominance value > 0.5. In the set of bars for each time scale, the left-most bar corresponds to academic sites, the middle bar corresponds to commercial sites and the right-most bar corresponds to foreign sites.

frequency of the outliers is small. On the other hand, we have also seen that the range of RTT observations, at several time-scales up to an hour, is one to two orders of magnitude larger than both the minimum RTT and the mode RTT. The first property tends to reduce the skew and the second tends to increase it. From Figure 5, we note that for individual links, normalized temporal variation is highest in the noon-4pm period and relatively high for the 8am-noon and 4pm-8pm periods. From Figure 8, we note that for most links and most time-periods, a large fraction of RTT observations are clustered around the mode.

To better understand the tension between these tendencies, we computed the skew for all hour-long periods studied. Figure 10 summarizes the results for the six four-hour periods used in Figure 5. For each period and for each link, it computes the fraction of hour-long traces whose mean was at least 1.2 times the mode. While the number of links that have significant skew varies with the time of day as does the frequency of skew within each time period. On the whole, the variation, however, is small: (1) the number of links with a non-zero skew measure varied around the 40% mark and (2) the frequency of skew within each time period varied around the 0.2 mark.

8 RTT distribution changes slowly

As we have seen in previous sections, RTT observations tend to cluster tightly around the mode. Therefore, in our study of the rate of change of the distribution, we used the mode RTT to characterize the RTT distribution over a given period. We consider two distributions to be different if their mode values are greater than 10 ms apart. (the time resolution in most Unix systems is 10 ms). To quantify the frequency of change of distribution, we compute the *runlength* for corresponding mode values. We summarize the information at six time-scales between one minute and one hour and present the results in Figure 11. We would like to point out the following facts:

1. At the finest time resolution, the median value of the runlength is 52 minutes. Therefore, we expect sampling the RTT values once every 45 minutes to an hour is likely to be adequate for latency-sensitive applications that wish to keep track of round-trip delay.

2. At all time-scales, there exist links whose RTT *distribution* does not change (in the sense that we have defined) for over 40 hours. For the coarsest resolution studied, that is an hour, this is true for one-third of the links.

By this, we do not intend to imply that there is no variation in RTT. As we have seen in previous sections, there is significant temporal variation across all links. Instead, it is the distribution that changes

(a) midnight-4am

(b) 4am-8am

(c) 8am-noon
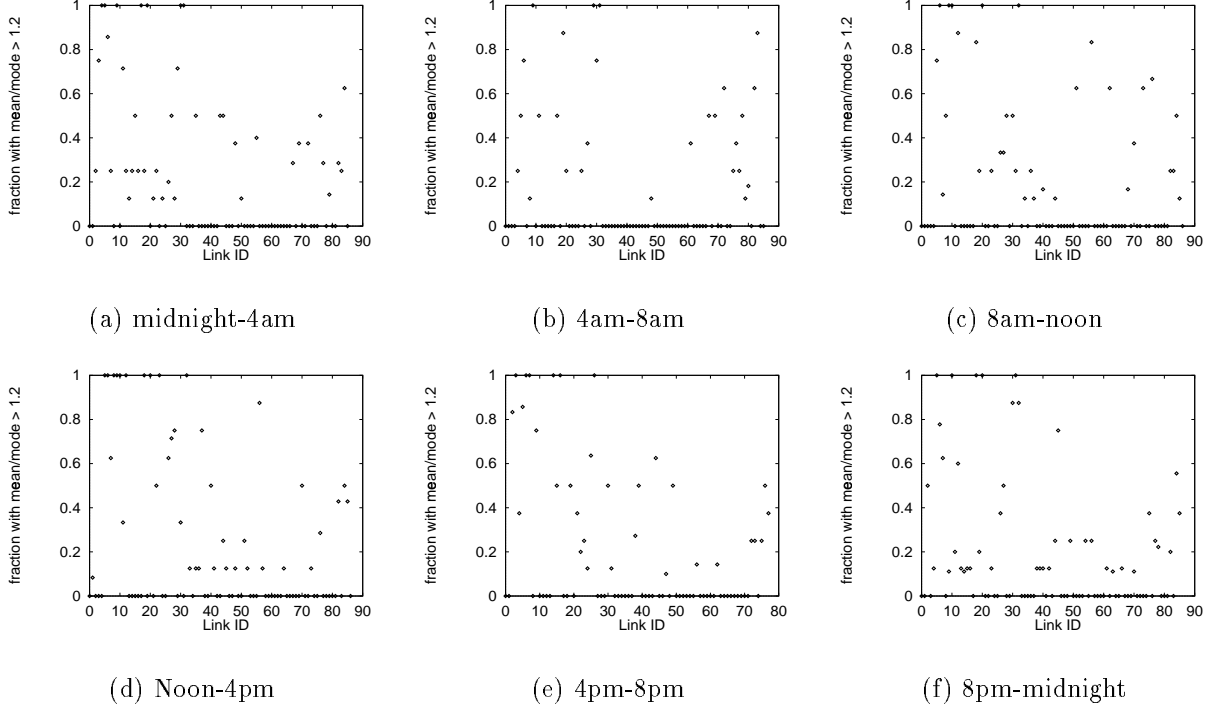
(d) Noon-4pm

(e) 4pm-8pm

(f) 8pm-midnight

Figure 10: Skew in RTT distribution over different periods in the day. The fraction of links with non-zero skew measures for the different time periods are: (a) 46%, (b) 31%, (c) 38%, (d) 44%, (e) 40% and (f) 48%. The average skew measures across all links are: (a) 0.22, (b) 0.16, (c) 0.19, (d) 0.23, (e) 0.20 and (f) 0.19.

slowly. This suggests that while efforts to predict future RTT based on previous observations may not have produced the desired results [8], efforts to predict the future distribution of RTT based on the distribution of previous observations may have a greater likelihood of success.

9   Spikes in RTT observations are isolated

As noted in previous sections, RTT distributions at all time-scales are skewed and possess a long tail on the right hand side. If RTT observations are presented as a time-series, the outliers would appear as spikes in the graph (for example, see Figure 12). Since these values occur infrequently at all time-scales, we suspected that these spikes are isolated - that is, they appear in the time-series for extremely short intervals. Additional evidence for this was provided by visual inspection of the observations for several links (for example, see Figure 12 (b)).

To provide quantitative evidence for this hypothesis, we computed an aggregate metric that we refer to as the *spike-isolation* metric. We computed this metric as follows:

1. We defined a spike as an observation that was at least two times the previous observation.

2. We considered a spike to isolated if it occurs for at most two observations. In other words, an observation is an isolated spike if it is at least twice both the previous observation and either of the next observation or the one after it.

3. For each trace, we computed the spike-isolation metric as the fraction of spikes that were isolated.

Figure 13 plots the spike-isolation metric for all the traces. It shows that most spikes for most links are isolated - as indicated by large number of points above the 0.8 mark. The summary measures presented in the caption make the same point.

(a) 1 minute   (b) 5 minutes   (c) 10 minutes

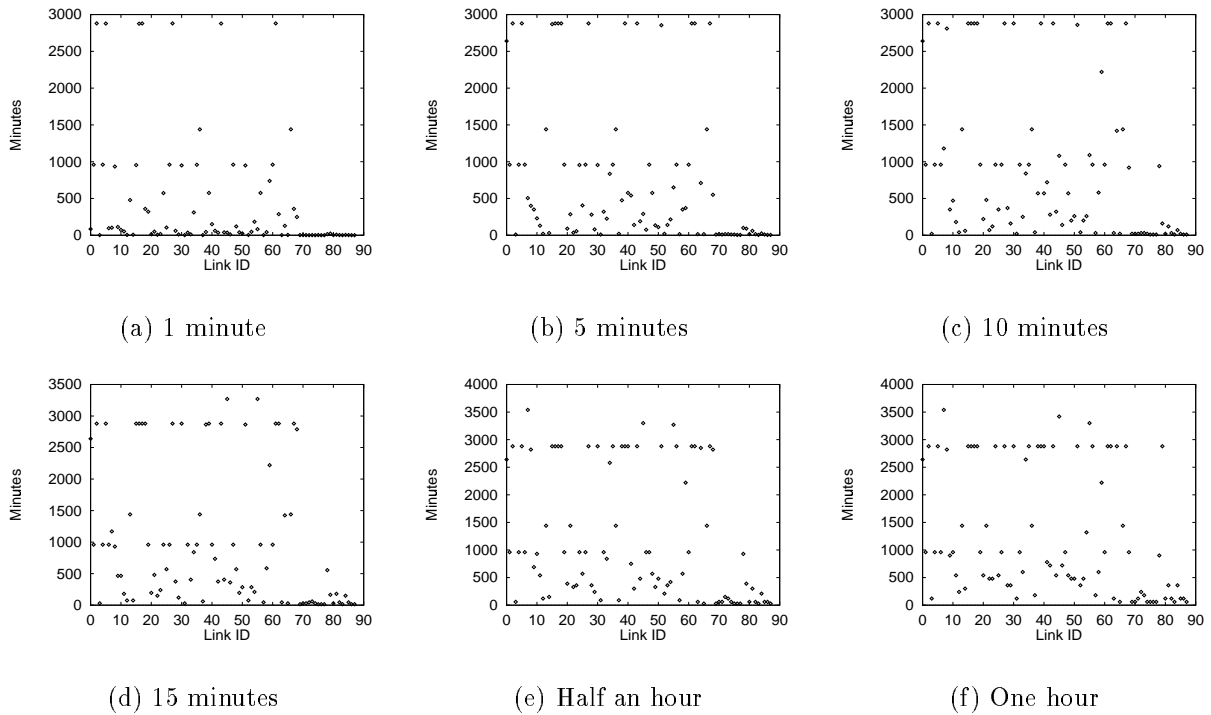(d) 15 minutes   (e) Half an hour   (f) One hour

Figure 11: Average runlength for mode at different time-scales. This measure indicates how long will a value remain within a small window. The window used is 10 ms. The median values for the plots are: (a) 52 minutes, (b) 290 minutes, (c) 470 minutes, (d) 480 minutes, (e) 840 minutes and (f) 900 minutes. The percentage of links which had runlengths greater than 40 hours is: (a) 8%, (b) 16%, (c) 18%, (d) 22%, (e) 30% and (f) 31%.
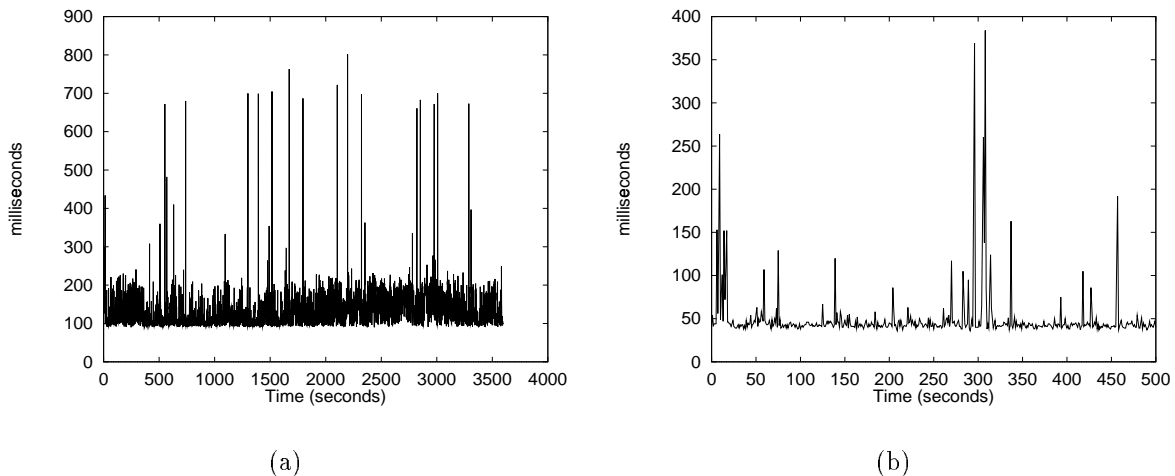

(a)   (b)

Figure 12: Sample RTT observations over two scales. Graph (a) is over an hour and graph (b) is for the first 500 seconds in that hour. These observations are for the link `baekdoo.cs.umd.edu-lanl.gov` between noon and 1pm on the 5th June, 1996.
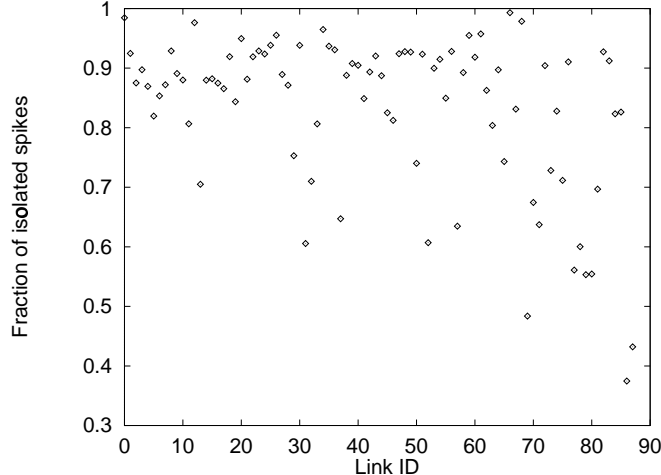
Figure 13: Fraction of spikes that are isolated for each link. The median value is 0.88, the average is 0.82 and 75% of the links have a spike-isolation metric greater than 0.8.

10 Jitter in RTT observations is small

From previous sections, we note that a large fraction of RTT observations are usually clustered in a small window around the mode and that spikes are infrequent and isolated. This indicates if the RTT observations are represented as a time-series, the observations would be relatively steady (within a jitter window). To quantify this, we computed the average runlength of RTT observations assuming a jitter window of 10 ms. Since spikes are usually isolated, we decided to compute a similar measure after eliminating all *isolated* spikes. [3] This allowed us to separate the effects of two kinds of variations: (1) sustained and (2) impulse. We speculate that the difference between these two kinds of variations reflects the difference between different kinds of congestion in the Internet. Sustained variation is likely to reflect changes in the traffic volume; impulse variation is likely to reflect short-lived events like a large transfer or transient backup at a router.

Figure 14 presents the graphs for both kinds of runlengths. We find that except for links at the right end of the graphs, the runlength is significant even without eliminating isolated spikes. For the links towards the right end of the graph, 10 ms is a small jitter window – the minimum RTT for these links is of the order of a few hundred milliseconds. Note that if the isolated spikes are removed, average runlength with a jitter window of 10 ms is about 104 seconds.

11 Estimated inherent RTT occurs frequently

An interesting fact that we noted in our study was that the estimated inherent RTT (that is, the minimum RTT over the entire trace) occurs frequently. To quantify this, we computed the fraction of 1-minute slots in all traces for all links in which the minimum RTT value occurs. Figure 15 presents the results. It shows that for most links (70%), at least half the one-minute intervals contain at least one occurrence of the estimated inherent RTT.

We note that the frequency of the occurrence of the inherent RTT falls off as we move from the left end of the graph to the right end. That is, links with a lower inherent RTT are more likely to achieve it. To summarize the information along a different dimension, we computed the average measure for the three host categories. We found that the academic hosts had the highest measure (0.83) followed by the commercial hosts (0.73) and foreign hosts (0.32).

There are two possible factors that might cause a low frequency of occurrence of the inherent RTT: (1) insufficient capacity at the network bottleneck link and (2) frequent route changes. Given the end-to-end nature of our study, it is hard for us to determine the causative factor in our experiments. We have, however,

---

[3] Note that sharp changes that survived for two seconds or more were not eliminated.

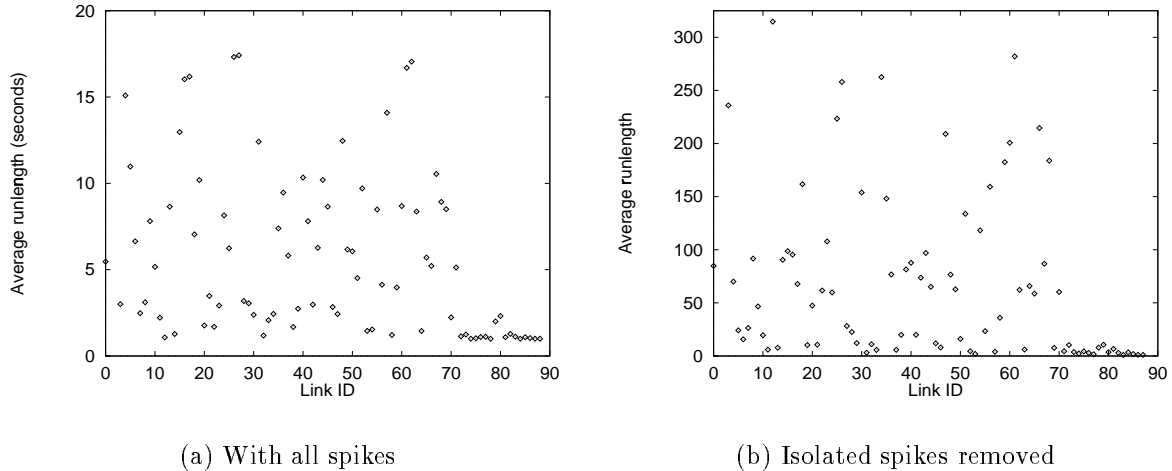(a) With all spikes

(b) Isolated spikes removed

Figure 14: Average runlength for each link. The jitter window used is 10 ms. Graph (a) shows the runlength without removing the isolated spikes; graph (b) shows the runlength once isolated spikes have been removed. Two outliers at the left end of the plot have been removed from both graphs. Note that the scales on the two graphs are about an order of magnitude different. The average runlength across all links was 7.3 sec before removing the spikes and 103.9 sec after removing the spikes. Note that sharp changes that survived for two seconds or more were not eliminated.
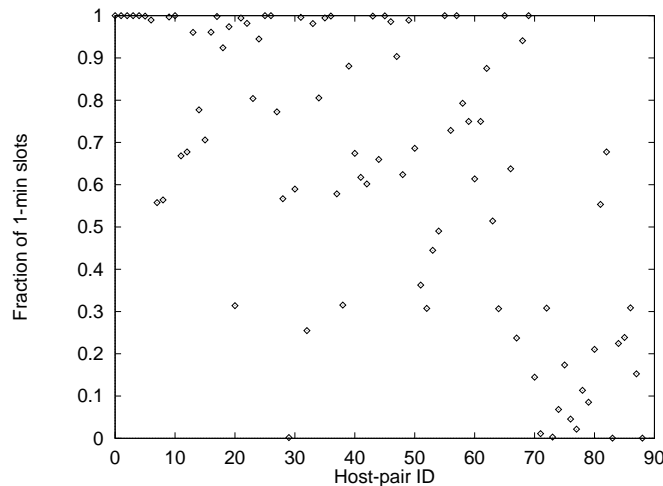

Figure 15: Fraction of 1-minute slots that the estimated inherent RTT occurs in. To summarize, the median value is 0.69, the mean value is 0.65 and 70% of the links have value greater than 0.5.

found that links from the CESDIS pinging site have had a consistently lower frequency of occurrence than other sites. This is independent of the distance of the pingee host. We illustrate this in Figure 16 which presents per-minute variation in the minimum RTT for three links originating at CESDIS. The pingee hosts for these three links are at different geographical distances. We note that the minimum RTT rises rapidly during the mid-day (noon-6pm EDT). We speculate that the bottleneck link for network connections from CESDIS lies within network segment between CESDIS and Goddard links to the external world. Using `traceroute`, we determined that this network segment consists of three hosts `zypher.gsfc.nasa.gov`, `rtr-wan2.gsfc.nasa.gov` and `rtr-internet-ef.gsfc.nasa.gov`.
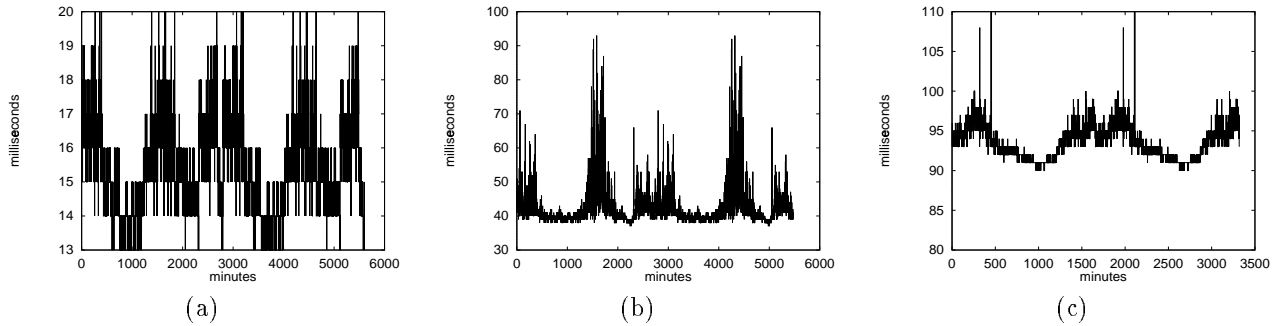
Figure 16: Time-of-day variation in minimum RTT for three links out of CESDIS. This indicates that the capacity out of CESDIS (or out of Goddard) is saturates every afternoon. Note that the scale on the y-axis is different for each of the three graphs. This indicates that this effect is independent of the distance of remote host. The three remote hosts here are (a) `zagnut.cs.umd.edu`, (b) `clone.mcs.anl.gov` and (c) `www.cs.washington.edu`.

12 Summary

To summarize, the conclusions of our study are:

1. There is large temporal and spatial variation in RTT. The extent of temporal variation depends on the time of the day.

2. RTT distribution has a long tail. The total range of observations is often one to two orders of magnitude larger than either the estimated inherent RTT or the shortest range of values that contains 75% of the observations. This conclusion holds across several time-scales from a minute to an hour.

3. Mode often dominates RTT distribution. In most cases, a short window of values around the mode contains a large fraction of the observations. This conclusion holds over several time-scales from a minute to an hour. This indicates that the mode would be a good characteristic value for RTT distributions.

4. RTT distributions change slowly. Using the mode as the characteristic value of the distribution, we found that the average period before there is a substantial change in the distribution (at a one-minute granularity) is about 50 minutes.

5. Spikes in RTT observations are isolated. That is, persistent changes in RTT occur slowly; sharp changes are undone very shortly (usually within a couple of seconds).

6. Distribution of RTT skewed is in many cases. In our study, about 40% of the links had significantly skewed RTT distributions. This indicates that there is often a substantial difference between the mean and the mode. If the mean and the mode were always close, either mean or the mode could have been used to characterize the distribution. Given the presence of skew, mode is likely to be a better characteristic value for RTT distributions.

7. Jitter in RTT observations is small. If sharp changes that last for less than two seconds are ignored, the average runlength with a jitter window of 10 ms was 103 seconds. Even without eliminating the isolated spikes, the runlength was about 8 seconds.

8. Estimated inherent RTT occurs frequently. We found that links within the US, the estimated inherent RTT occurs in about 78% of 1-minute slots. This includes extremely busy web sites such Altavista, Netscape, Lycos and Inktomi.

We would like to mention two caveats. First, even though we have tried to take geographical location and the character of hosts into consideration and have been able to achieve a good distribution along both dimensions, a sample of 44 hosts is small. We do not argue that this study captures all of Internet RTT behaviors. While we believe that this study provides useful information about Internet RTT, we would like study more hosts. Second, these experiments did not keep track of the path taken by the packets. Since route changes are a part of the end-to-end behavior seen by Internet hosts, we believe that this makes no difference to the conclusions of this study. But the lack of this information makes it impossible the differentiate between the effects of congestion and route-change. Note that given the scale and the administrative structure of the Internet, it is hard to get continuous route *and* RTT information.

13   Future work

We would like to extend this work in four directions. First, we would like to build a parameterized model for the RTT distribution of individual links. A point we note in this regard is that the nature of the distribution for individual links (degree of localization, length of the tail, skew if any) is similar across several time-scales. We would like to caution the reader that as yet we have analyzed the data only at time-scales between a minute and an hour – which leads us into the second direction that we would like to extend this work. We would like analyze the data for coarser resolutions. Third, we would like repeat at least a subset of this study at a finer time resolution. The goal of this would be to determine whether the once-per-second sampling of RTT is in some way biasing the results. Finally, as mentioned in the previous section, we would like to study more hosts.

Acknowledgments

References

[1] L. Amsaleg, M. Franklin, A. Tomasic, and T. Urhan. Scrambling query plans to cope with unexpected delays. In *Proceedings of the Fourth International Conference on Parallel and Distributed Information Systems*, December 1996. To appear.

[2] T. Brisco. DNS support for load balancing. RFC 1794, Network Working Group, April 1995.

[3] L. Cabrera, E. Hunter, M. Karels, and D. Mosho. User-process communication performance in networks of computers. *IEEE/ACM Transactions on Software Engineering*, 14(1):38–53, 1988.

[4] R. Carter and M. Crovella. Dynamic server selection using bandwidth probing in wide-area networks. Technical Report BU-CS-96-007, Computer Science Department, Boston University, March 1996.

[5] A. Chankuthod, P. Danzig, C. Neerdaels, M. Schwartz, and K. Worrell. A hierarchical internet object cache. In *Proceedings of the 1996 USENIX Annual Technical Conference*, 1996.

[6] K. Claffy, G. Polyzos, and H.-W. Braun. Long-term traffic aspects of the NSFNET. In *Proceedings of INET'95*, 1995.

[7] O. Etzioni, S. Hanks, T. Jiang, R. Karp, O. Madani, and O. Waarts. Efficient information gathering on the internet. In *Proceedings of the 1996 FOCS*, 1996.

[8] R. Golding. End-to-end performance prediction for the internet (work in progress). Technical Report UCSC-CRL-92-26, University of California at Santa Cruz, June 1992.

[9] E. Katz, M. Butler, and R. McGrath. A scalable HTTP server: The NCSA prototype. *Computer Networks and ISDN Systems*, 27(2):155–64, Nov 1994.

[10] A. Leinwand and J. Okamoto. Two network management tools (how many packets could a packet router route if a packet router could route packets). In *Proceedings of the Winter 1990 USENIX Conference*, pages 195–205, Jan 1990.

[11] M. Mathis. Windowed `ping`: an IP layer performance diagnostic. In *Proceedings of INET'94*, Jun 1994.

[12] D. Mills. Internet delay experiments. RFC-889 Network Information Center, SRI International, 1983.

[13] A. Mukherjee. On the dynamics and significance of low frequency components of Internet load. *Internetworking: Research and Experience*, 5(4):163–205, Dec 1994.

[14] J. Pointek, F. Shull, R. Tesoriero, and A. Agrawala. NetDyn revisited: A replicated study of network dynamics. Technical Report CS-TR-3696, Department of Computer Science, University of Maryland, Oct 1996.

[15] J. Quarterman, S. Carl-Mitchell, and G. Phillips. Internet interaction pinged and mapped. In *Proceedings of INET'94*, Jun 1994.

[16] M. Ranganathan, A. Acharya, S. Sharma, and J. Saltz. Network-aware mobile programs. In *Proceedings of the 1997 USENIX Annual Technical Conference*, Jan 1997. To appear.

[17] D. Sanghi, A. Agrawala, O. Gudmundsson, and B. Jain. Experimental Assessment of End-to End Behavior on Internet. In *Proceedings of IEEE Infocom*, 1993.

[18] J. Sedayao and K. Akita. LACHESIS: a tool for benchmarking Internet Service Providers. In *Proceedings of the Ninth USENIX Systems Administration Conference*, pages 111–5, Sep 1995.