# Homework 7: Round-trip time measurements with PlanetLab

Ritesh Agarwal

December 4, 2012

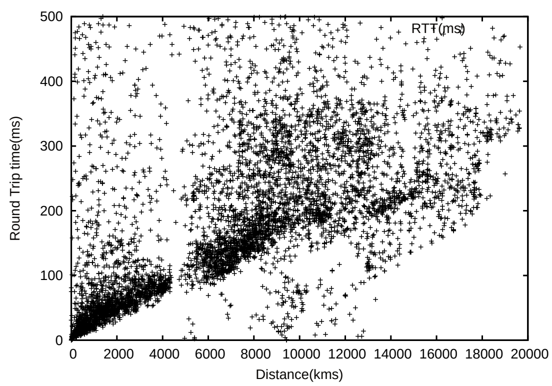


Figure 1: Scatter plot of RTTs collected at night (10:30 PM)

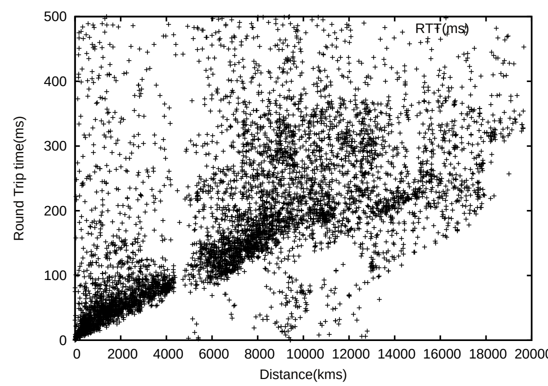


Figure 2: Scatter plot of RTTs collected at early morning (3:00 AM)

## Abstract

In this homework, we collect round trip times between 200 hosts spread across the globe and try to summarize this information by using the location information provided by PlanetLab to calculate the distances between two hosts.

## 1 Scatter Plot

In Fig. 1 and Fig. 2, we see two Scatterplots of pairwise round-trip time measurements taken over data sets collected at two different times of day. As we can see that the densely plotted areas show linear increase with distance, however there are some points that indicate either a very long RTT or very small RTT, some of them which seem to be even impossible. At distances 9500km, we can see RTT as low as 1ms. This makes the speed of bits faster than speed of light. This is possible either because some of the location details we have are incorrect.

## 2 Cumulative Distribution Function

Here we show the cumulative distribution function (CDF) plot for round-trip time, and for distance. The y-axis of a CDF goes from 0 to 1. It is zero for the minimum X value, and 1 for the maximum X value, and grows with X.

In Fig. 3 and Fig. 4, we see two Cumulative distribution function (CDF) plots for round-trip time, and for distance taken over two different sets of data collected at two different times of day. We can see that the median of the sets of the data is 7812 and 7820 respectively. Note: Values calculated from actual plots, not by looking at graph.

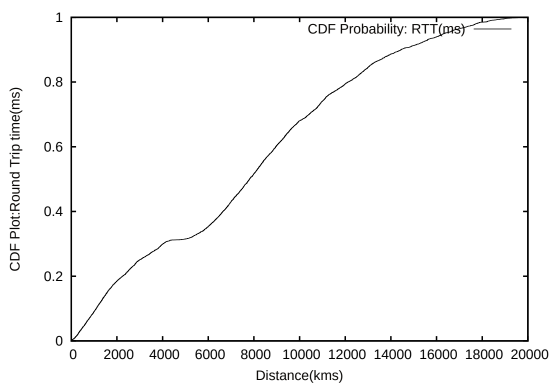


Figure 3: CDF of RTTs collected at night (10:30 PM)

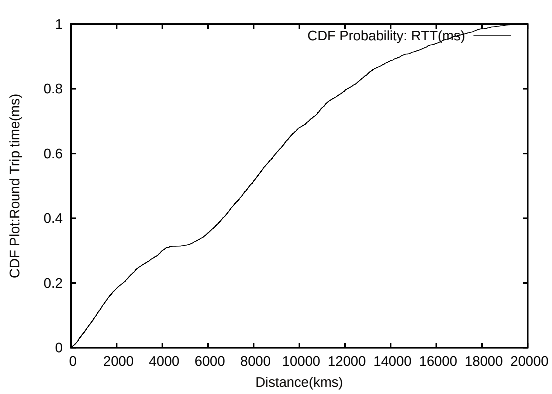


Figure 4: CDF of RTTs collected at early morning (3:00 AM)

# 3 Collection, Pruning of Data

## 3.1 Collection of Data

From the given list of 200 IPs, we try to ping every other ip from every IP for 10 samples. Every set of pings are parallely done, where each ping to every other machine is again parallely done. An aggregated result of ping per host is collected, which contains ping results to all the other hosts. The output of ping command provides an average rtt for every ping command fired, if we get a response for at least 1 packet.

## 3.2 Pruning of Data

Once we collect the ping results, we calculate the rtts and distance for every pair of hosts and combine them. We then round the numbers (distances and rtts) to integers. The results are then pruned to remove duplicate distance,rtt pairs. Again different rtts for same distance are averaged. While averaging reduces precision of dataset, it also makes it easy to visualize data.

# 4 Analysis of Data

## 4.1 Speed of bits

Speed of bits was calcuated as follows for each unique sample as below

$$speed_i = \frac{distance_i *}{\frac{rtt_i}{1000*2}}$$

## 4.2 Mean Speed of bits

Mean speed of bits was calculated as below

$$mean = \frac{\sum_{i=0}^{n} speed_i}{n}$$

$$mean_1 = 87431.8 km/s$$

$$mean_2 = 88849.6 km/s$$

## 4.3 Standard Deviation

Standard Deviation is calculated to indicate how much variation exists in the data set from the mean calculated as above. The Standard Deviation is calculated as below

$$sd = \sqrt{\frac{\sum_{i=0}^{n} (speed_i - mean)^2}{n}}$$

$$sd_1 = 272362$$

$$sd_2 = 278258$$

## 4.4 Differences of Mean and Standard Deviation calculated over two different sets

There could be several reasons for having different mean and Standard Deviation over two different sets. A very generic reason to describe the reason would be the state of the nework at different point of times. Traffic load on the internet and machine load on planet machines could be another reason that could contribute or reduce the mean speed.

## 4.5 Maximum(uncongested) speed of bits on the internet (km/s)

It is difficult to estimate the maximum speed of bits on the internet with the collected data unless, we have correct location information of every Planet Lab node, or at least unless we know which of them are incorrect. No doubt, it will be less than the speed of light. In order to be able to achieve a speed of bits same as speed of light ,we need to have a fiber optic link from point A to point B, with no repeaters or any other network devices between A and B. q

## 4.6 Misreported locations by PlanetLab

A table consisting of source IP, destination IP, speed of bits, Long Lat of Source and Destination was generated from the ping times collected. The table was then sorted in ascending order based on the speed of bits. It was suprising to know that 316 speed samples out of 29702 samples were faster than speed of light. This included ping times both from A to B and B to A. From the collected results, it is safe to assume that at least

$$\frac{316}{2}$$

i.e 158 IP addresses are misreported by PlanetLab unless we were able to collect all the information. To find out this IPs would involve considerable amount of manual work until we find a way to script the City information provided by google maps for the provided lon, lat pairs and compare it with what the whois command result returns.

On doing the above task manually, following details of some misreproted IPs whose locations are misreported by PlanetLab were found.

204.8.155.227 is reported to be in Argentina but is actually in Boston, MA.

140.112.107.82 is reported to be in Australia, but is actually in Taipei, Taiwan.

200.0.206.12 is reported to be in the North Atlantic Ocean, but is actually in Rio de Janeiro, Brazil.

## 4.7 Sites/Countries with poor connectivity

A table consisting of source IP, destination IP, speed of bits(taking care of distance and rtt) was generated from the ping times collected. Only those entries where we were able to ssh to the machine but not able to ping some or all of the other hosts were extracted. A count of unique ip on source and destination was derived as below.

The below table shows the IPs that were not pingable and their counts.

count IP

—– ————-

168 132.170.3.32

168 132.170.3.33
169 129.93.229.138
169 136.159.220.40
169 138.15.10.56
169 138.26.66.5
169 150.140.184.252
169 194.42.17.121
169 202.125.215.12
170 130.92.70.254
170 131.114.53.188
170 132.72.23.11
170 142.104.21.241
170 193.145.46.129
170 193.55.112.41
170 206.117.37.4
170 202.112.129.16

The below table shows the IPs that could not ping any other host

count IP
—– ————-
200 129.93.229.138
200 136.159.220.40
200 150.140.184.252
200 194.42.17.121

Another approach was taken to find IPs that had low speed of bits towards them but were at least reachable. A table consisting of source IP, destination IP, speed of bits(taking care of distance and rtt) was generated from the ping times collected. All the entries that did not contain rtts were removed. The table was sorted first on ip, the table was again numerically sorted on the speed of bits(such that ips who a. The top 4000 entries in the sorted table consiting of about 160,000 samples were taken. Now the no. of unique ips and their counts in the top 4000 entries were derived. The below table shows the IPs that had low speed of bits towards it.

count IP
—– ————-
88 202.141.161.43
89 137.165.1.113
89 199.26.254.67
98 206.23.240.29
110 131.193.34.38
137 200.129.132.18
160 128.138.207.54
161 128.143.6.134

# 5 Future Work

## 5.1 Estimation of the location of IP address 68.86.95.9

An approach that we can use for the estimation of location of IP address(long, lat) is to ping this IP from every planet lab host. Using the distance,rtt information we have collected, we should be able to estimate the distance of this IP from every host that was derived from the rtt, we can approximate the location of the given IP address. However to be able to do this correctly, we need to have accurate location data of planet lab nodes.

This kind of approximation can also be used to find out errors in IP locations provided by Planet Lab.

# 6 Conclusions