Homework 7: Round-trip time measurements - Playing around with Planetlab

Raghavan KL

November 14, 2012

Abstract

In this homework, I measure and report round-trip times for 200 Planetlab hosts. I have used these planet lab machines to ping each other and collected their data, based on this data I have made estimations on rtt and distance between the machines around the world. Following are the steps involved and they are explained in detail in this document,

- Collection of data
- Extraction/Cleaning
- Visualization of data
- Analysis
- Inference

The whole process was carried out in two different time intervals one was on *Nov-12-2012 10.00PM(Night)* and other was on *Nov-13-2012 9.00AM(Morning)*.

1 Collection

After getting a planet lab account, I was able to login to any of the active planet machines(approx. 200).Then I passed a file to all the planet lab machine which consist of ping command to all machines. The ping data of all these machines I extracted and placed into my local log file. This step was optimized by running the commands in the background instead of waiting in the foreground, by this I managed to gather around 300,000 ping records. The distance between every other IP is calculated using the perl library which manipulates them based on the lattitude, longitude of IPs. A total record set of 200 * 200 IPs is calculated in the format of ***From IP, To IP, Distance*** and stored for future use. One time storage of this data would be a viable idea since this would be a static data and no need to calculate them again.

2 Extraction/Cleaning

2.1 Extracting ping data

The ping data which is gathered is then structured into a file consisting of the format ***From IP, To IP, RTT Time***. This is an cleaned dataset derived out the whole ping dataset which was gathered during the collection process.

2.2 Merging distance and Ping dataset

The stored distance data is used to map with the extracted ping information and new distance ping time file is structured in the following format ***From IP, To IP, Distance, RTT Time***.

3 Visualization of data

3.1 Scatter Plot

A scatter plot is developed taking the Distance(in Kms) in the X axis and RTT(Round Trip Time in ms) on the Y axis. From the given *Figures.1 and 2* we can observe that there is a linear increase on way section and there are many disturbances all around which might be due to the latency of ping or some machine which responded very late.

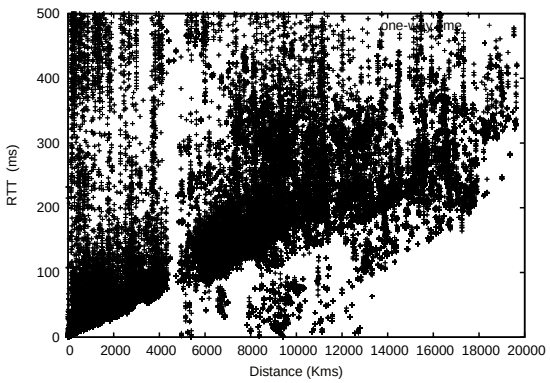

Figure 1: My round-trip time measurements on Night data.

3.2 CDF Graph

As CDF plot is made by taking the cummulative values of RTT in the Y axis and distances taken in the X axis. The *Figures. 3, 4* shows the CDFs drawn during two different time intervals.

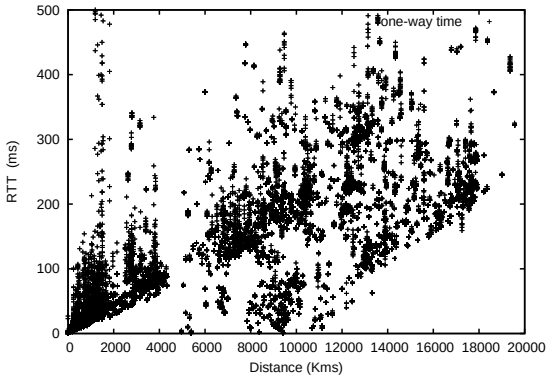

Figure 2: My round-trip time measurements on Morning data.

3.3 SD Curve

The mean speed of bits was calculated using the formulae

$$distance * 2/(rtt/1000)km/s$$

and the Standard Deviation calculated using this formulae

$$\sqrt{1/N \sum_{i=0}^{N} (X_i - mean)^2}$$

.

For the experiment which was run during night, the mean speed of bits was calculated as ***104535 km/s*** and the Standard Deviation calculated was ***307203*** this shows the disturbances in the distribution of the values. The *Figures. 5* shows the SDs drawn during the night experiment.

For the experiment which was run during morning, the mean speed of bits was calculated as ***118582 km/s*** and the Standard Deviation was around ***452566*** this again shows the disturbances in the distribution of the values.The *Figures.6* shows the SDs drawn during the morning experiment.

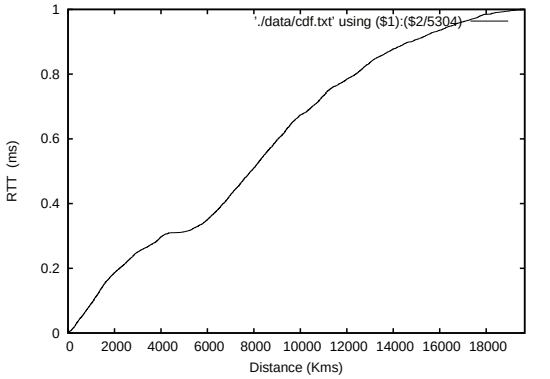

Figure 3: My CDF plot on Night collected data.

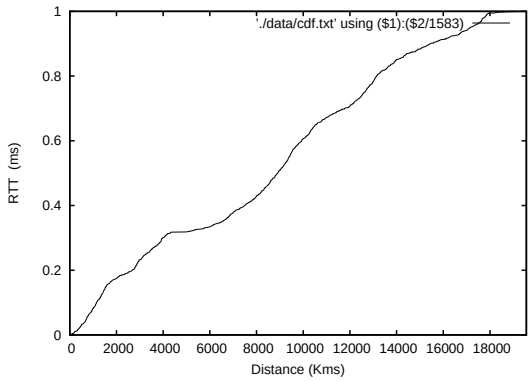

Figure 4: My CDF plot on Morning collected data

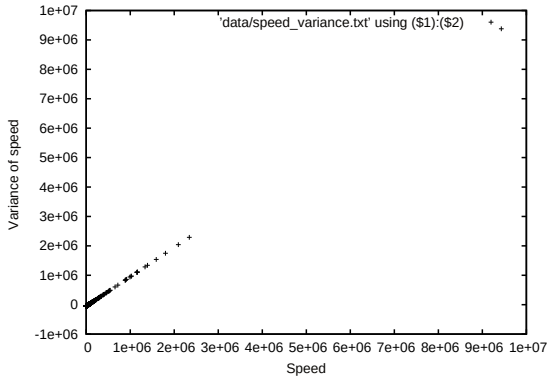

Figure 5: My SD plot on Night collected data

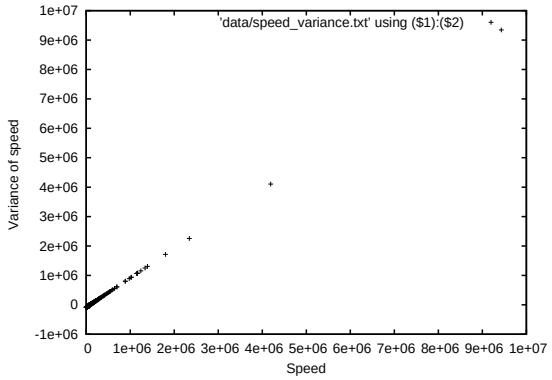

Figure 6: My SD plot on Morning collected data

4 Analysis

4.1 Max. Speed of bits on internet

Comparing to the maximum speed of light

$$3 * 10^8 m/s$$

the maximum speed of bits which I could infer from my dataset was maximum to the range of

$$1 * 10^8 m/s$$

which was below the speed of light it might be because of channel which carries the bits might be slow(bandwidth) or there might be many bottleneck connections which is shared among my channels.

4.2 Poorly connected countries

From the collected RTT dataset of all planet lab machines the *To IP* address which always had the less ping time from all other machines is sorted out and the last few records are tailed, on looking at this data there are few IPs which are expected to be in the poorly connected countries. Few such IPs are listed below,

- 138.4.0.120
- 130.216.1.22
- 130.216.1.23
- 163.117.253.23
- 163.117.253.22
- 156.62.231.242
- 193.136.124.226
- 202.112.28.100
- 190.227.163.141

4.3 Deducing IP location

The IP address given for deduction *68.86.95.9* is pinged from all the planet lab machines to all other machines, the RTT log is gathered out and this ping data is compared with the existing ping data set . The *From IP, To IP, Distance, RRT* matching with the existing data pool is collated and by ball parking this is infered to be a location matching to IP *128.252.19.18 with latitude 38.6479 and longitude -90.3015*

5 Challenges

There were few interesting challenges which I faced and almost succedded in solving them,

- To understand the awk, sed, many other bash commands and latex.
- Spawning background jobs to expedite the process of collecting ping data.
- Cleaning up the collected data and formatting them to process.
- Sorting data based on various parameters and logically extracting them.
- Understanding the math behind Standard Deviation, Cummulative Distribution and handling them.

6 Conclusions

This excerise on the whole helped me to understand the underneath concept of data travel around the internet, also a massive dataset was collected, cleaned and visualized with effective usage of shell scripting, few analysis was performed based on the dataset collected.