

Measurement of RTT (Round Trip Time) between any pair of active servers with error correction

Sri Harsha Chennavajjala
University Of Missouri, Kansas City
sc9v9@mail.umkc.edu

Teja Garidepally
University Of Missouri, Kansas City
tgww4@mail.umkc.edu

Raj Kiran Reddy Munnangi
University Of Missouri, Kansas City
rmyb9@mail.umkc.edu

Vinay Kumar Matoori
University Of Missouri, Kansas City
vm8b8@mail.umkc.edu

Uday Chowdary Malineni
University Of Missouri, Kansas City
um5m2@mail.umkc.edu

Aishwarya Rao Venkata
University Of Missouri, Kansas City
avrpn2@mail.umkc.edu

Pallavi Ramineni
University Of Missouri, Kansas City
pr3md@mail.umkc.edu

Swetha Chandra Karroti
University Of Missouri, Kansas City
sckqn8@mail.umkc.edu

FALL 2015, November 3, 2015

Abstract— The main aim of the project is to calculate the RTT (Round trip time) between any two nodes by collecting data and correcting error there by obtaining exact distance between the servers. Automated geolocation of IP addresses has important applications to targeted delivery of local news, advertising and other content over the Internet. Previous measurement-based approaches to geolocation employ active probing to measure delays among a set of landmark nodes with known locations. The location of a target IP address can be approximated by that of the nearest landmark, as determined by the delay measurements [1].

I. INTRODUCTION

The project mainly deals with calculating the RTT (Round trip time) between any two nodes by collecting data and correcting error there by obtaining exact distance between the servers.

Round Trip Time: In telecommunications, the round-trip delay time (RTD) or round-trip time (RTT) is the length of time it takes for a signal to be sent plus the length of time it takes for an acknowledgment of that signal to be received. This time delay therefore consists of the propagation times between the two points of a signal.

Basically RTT value depends on various factors including:

- The speed with which intermediate nodes and remote server function.
- The number of other requests handled by intermediate nodes and remote server.
- The number of nodes between source and destination.
- The physical distance between the source and destination.
- The data transfer rate of source internet's connection.
- The presence of interference in the circuit.
- The amount of traffic on LAN to which the end user is connected. [3]
- The nature of the transmission medium.

Other than above factors RTT depends mainly on two parameters.

Scheduling delay: It refers to a difference between a desired time of arrival or departure and the actual time.

Network delay: Network delay is an important design and performance characteristic of a computer network or telecommunications network. The delay of a network specifies how long it takes for a bit of data to travel across the network from one node or endpoint to another. It is typically measured in multiples or fractions of seconds. Delay may differ slightly, depending on the location of the specific pair of communicating nodes.

II. EXISTING SYSTEM

Geolocation of IP addresses has important applications to targeted delivery of local news, advertising and other content over the Internet. One approach to IP geolocation is to maintain a large distributed database containing mappings of IP addresses to geographical locations. Many sites rely on more or less accurate databases to determine the location of a customer for various reasons:

- Determining regional distribution of the clients, local news delivery, targeted advertising, restriction of content delivery based on regional policies, etc.;
- Prevention/reduction of Internet frauds such as credit card fraud, identity theft, spam and phishing;
- Application to intrusion detection.

Research is going on now to automate the geolocation of IP addresses. Padmanabhan and Subramanian have investigated three techniques [2]:

GeoTrack: Inferring the geographic location of an Internet host based on the DNS name of the host or another nearby node.

GeoCluster: clustering the IP address space into likely collocated clusters.

GeoPing: Pinging the host, with geolocation of the IP address performed by correlating ping delays.

The latter approach employs active probes to measure delays among a set of landmark nodes with known locations. Such delay measurements can be performed by a distributed network of servers. Such a network of servers can be self-calibrating and potentially able to detect when a target IP address has changed its geographical coordinates significantly.

However, inferring geographical location from Internet delay measurements may result in large errors due to the nonlinear relationship between geographical distance and “Internet” distance as determined from delay measurements. Given delay measurements among a set of landmark nodes with known locations, the location of a target IP address can be approximated by that of the nearest landmark. To improve geolocation accuracy, a variation of this approach uses multilateration with geographic distance constraints to obtain a continuous location space rather than the discrete set of landmark locations. Gueye improved upon GeoPing using an idea borrowed from sensor networks. Their Constraint-Based Geolocation (CBG) algorithm uses a multilateration algorithm to determine the probable location of the targets. Katz-Bassett proposed Topology Based Geolocation (TBG), which finds hosts along Internet paths using the traceroute utility and geolocates hosts simultaneously using CBG. All of these methods are based on deterministic algorithms, which can have unacceptable geolocation errors of more than 1000 km.

III. PROBLEM DESCRIPTION

The existing techniques are not up to the mark because of their way of approach will not always have the accurate and updated database. So we are using RTT phenomena to estimate the distance of nodes.

IV. DESCRIPTION OF OUR APPROACH

In this project are developing a RTT measurement system which measures the RTT between any pair of server through exchange of files with error correction.

The resources that we have used are

- Ubuntu Linux
- SCP/SSH
- MySQL
- gcc compiler
- PlanetLab
- Matlab
- R Language

Steps for data collection:

- PlanetLab is a global research network that supports the development of new network services.
- Since the beginning of 2003, more than 1,000 researchers at top academic institutions and industrial research labs have used PlanetLab to develop new technologies for distributed storage, network mapping, peer-to-peer systems, distributed hash tables, and query processing.
- After creating account in planet lab , a slice has been allocated to us
- Selected 89 nodes which are geographically diversified and added to our slice.
- Generated private key and public key using *ssh-keygen*, and uploaded the generated public key in to planetlab for authentication.
- Created 5 different sized files with names file32B, file1KB, file256KB, file512KB, file1MB in our local machine and transferred these files to all nodes using *scp* command.
- Installed gcc compiler in all machines to run our programs for communication between nodes.
- Our program starts each and every node as both client and server from our local machine.
- Every node acts a client and fetches 5 files from all other servers.
- Every node acts a server and sends files to all the requested clients.
- Time taken by every node to download files will be recorded in logs.
- All the generated logs will be loaded in to database *downloaddetails_bkp*.

Here the collected data will be analyzed in next section

V. DATA ANALYSIS

- Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and evaluate data.
- Here we are developing a model where we estimate the distance between source and destination based on the network delay.

In this part we first figure out the relationship between distance and network delay.

In Data Analysis we have done following:

- Calculation of DDR(Distance to Delay Ratio)
- Calculation of the Estimated Distance from source to host
- Calculation of Distance Error between Estimated and Actual Distances
- Plotting Graphs for Analysis
- Implementing Error Correction Methods to reduce the Distance Error.

Calculation of DDR: We use distance and network delay from the data which we collected to calculate the DDR values

DDR is calculated by taking the distance to delay ratio between for each combination of source and host.

Formula to calculate DDR:

$$\text{DDR} = \text{Distance}(\text{source}, \text{Host}) / \text{rtt}(\text{Source}, \text{Host})$$

Where $\text{rtt}(\text{source}, \text{host})$ is delay between source and Host.

Average DDR for each node is found by taking average of all DDR's for that node.

Query used to calculate Average DDR:

```
SELECT
    AVG(b.distance / a.avgdelay) AS
DDR
FROM
    downloaddetails_bkp a
    JOIN distofnodes b
WHERE
    a.source = b.source
    AND a.destination =
b.destination
    AND a.filename = 'file1KB.txt'
    AND a.source <> a.destination
GROUP BY a.destination;
```

In the above Query we joined two tables 'downloaddetails_bkp', 'distofnodes' and find the

average DDR for each node for downloading file size of 1KB. This average DDR gives us the data transfer rate of that node from all other nodes. In the same way average DDR is calculated for different size files (32B, 256KB, 512KB, 1MB). All the average DDR's are loaded into view so that they get automatically get updated when the other log data is inserted into database. The view of the DDR contains average DDR values for each node for different file size.

Calculation of the Estimated Distance from source to host:

Estimated distance to host is calculated by multiplying the DDR to the rtt (round trip time) or delay from that source to host.

This distance we predicted show how far the host is from the source.

Formula to calculate Estimated Distance:

$$\text{Estimated Distance} = \text{DDR} * \text{rtt}(\text{Source}, \text{Host})$$

Query used to calculate Estimated Distance:

```
SELECT given_delay*DDR32KB
FROM DDR
WHERE destination="ebb.colgate.edu";
```

In the above query we find the estimated distance from source 'ebb.colgate.edu' based on the given delay. This is the predicted distance value that host may be accurately located geographically.

Calculation of Distance Error between Estimated and Actual Distances:

Distance Error is the value of difference between geographical distance and the predicted distance.

Formula for calculating Distance Error:

$$\text{Distance Error} = \text{Actual distance} - \text{Estimated Distance}$$

This error value may be different for different combinations of nodes. We apply some error correction methods to reduce the distance error between estimated distance and geographical

distance so that the predicted distance may get closer to the actual distance where the host is present.

VI. DATA REPRESENTATION

We have plotted different kinds of graphs using the collected data.

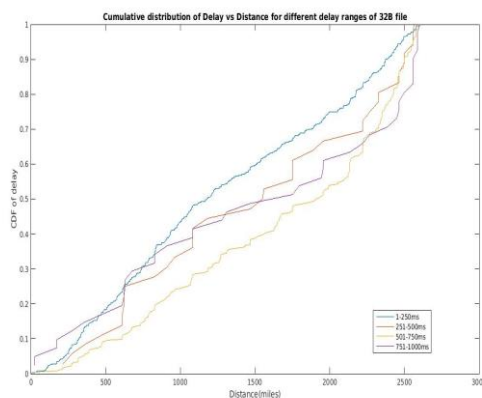
- CDF of delay vs distance.
- Distribution of measurements for 4 nodes in our planetlab set.
- Q-Q plot for 4 planetlab nodes.

CDF of delay vs distance

1. CDF of delay vs distance for 32B file

X-axis: distance

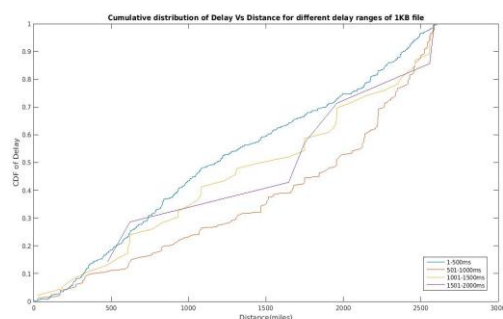
Y-axis: CDF of different delay ranges



2. CDF of delay vs distance for 1KB file

X-axis: distance

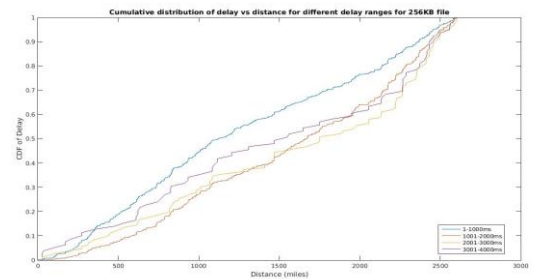
Y-axis: CDF of different delay ranges



3. CDF of delay vs distance for 256KB file

X-axis: distance

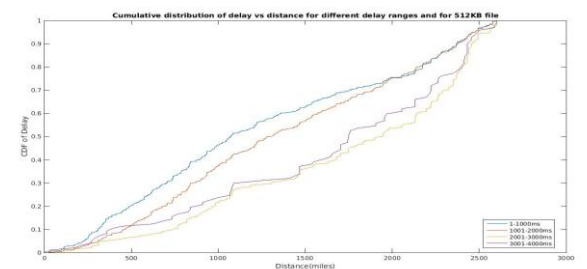
Y-axis: CDF of different delay ranges



4. CDF of delay vs distance for 512KB file

X-axis: distance

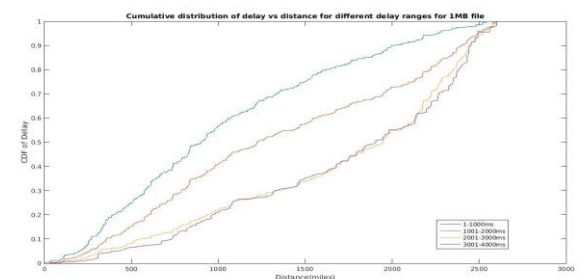
Y-axis: CDF of different delay ranges



5. CDF of delay vs distance for 1MB file

X-axis: distance

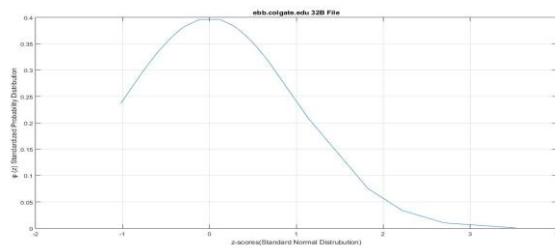
Y-axis: CDF of different delay ranges



Distribution of measurements

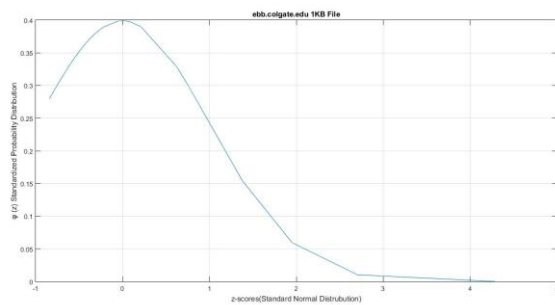
1. *ebb.colgate.edu* and 32B file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



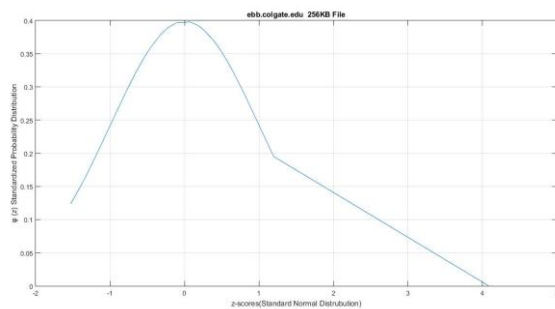
2. *ebb.colgate.edu* and 1KB file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



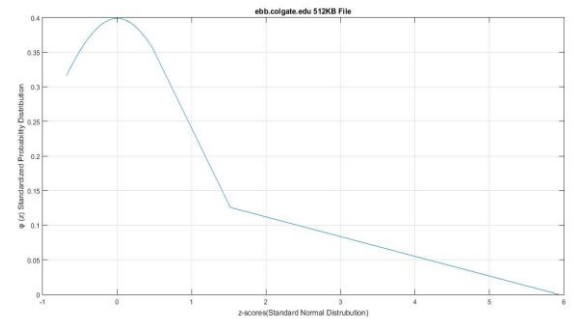
3. *ebb.colgate.edu* and 256KB file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



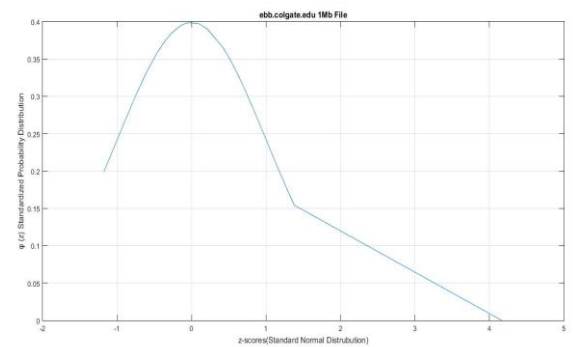
4. *ebb.colgate.edu* and 512KB file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



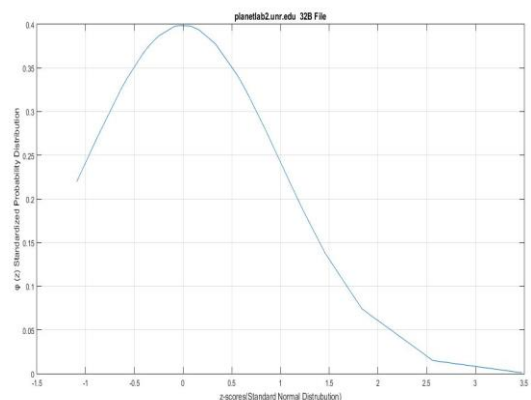
5. *ebb.colgate.edu* and 1MB file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



6. *planetlab2.unr.edu* and 32B file:

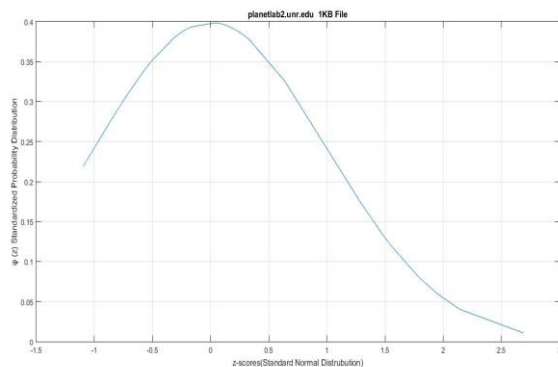
X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



7. *planetlab2.unr.edu* and 1KB file:

X-axis: z-scores for standard normal distribution

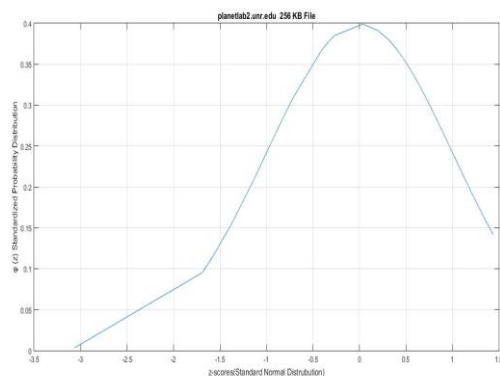
Y-axis: $\phi(z)$



8. *planetlab2.unr.edu* and 256KB file:

X-axis: z-scores for standard normal distribution

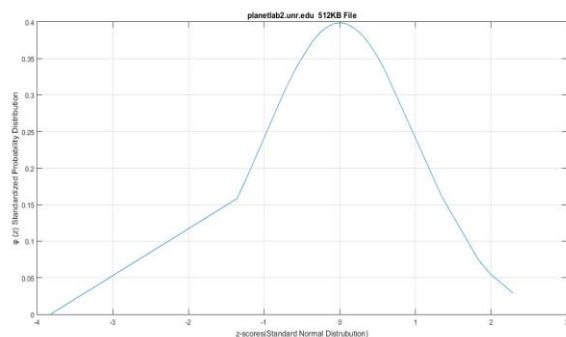
Y-axis: $\phi(z)$



9. *planetlab2.unr.edu* and 512KB file:

X-axis: z-scores for standard normal distribution

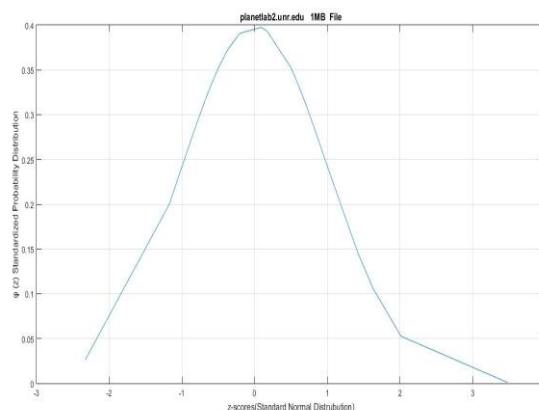
Y-axis: $\phi(z)$



10. *planetlab2.unr.edu* and 1MB file:

X-axis: z-scores for standard normal distribution

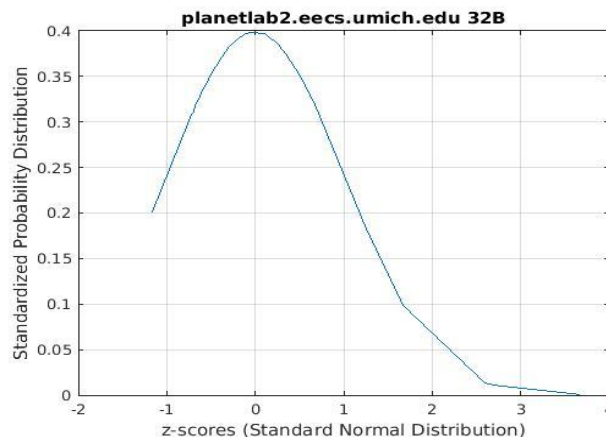
Y-axis: $\phi(z)$



11. *planetlab2.eecs.umich.edu* and 32B file:

X-axis: z-scores for standard normal distribution

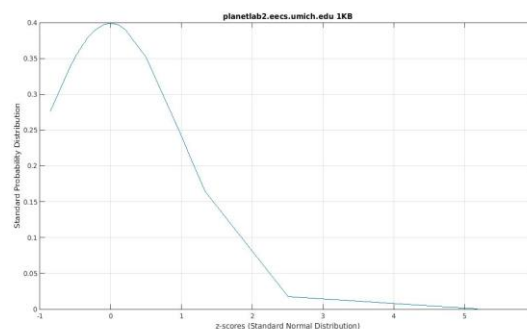
Y-axis: $\phi(z)$



12. *planetlab2.eecs.umich.edu* and 1KB file:

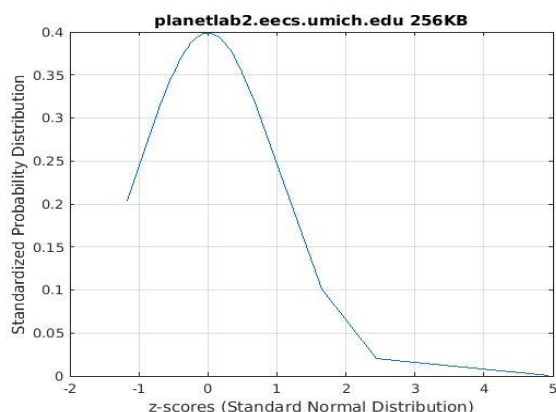
X-axis: z-scores for standard normal distribution

Y-axis: $\phi(z)$



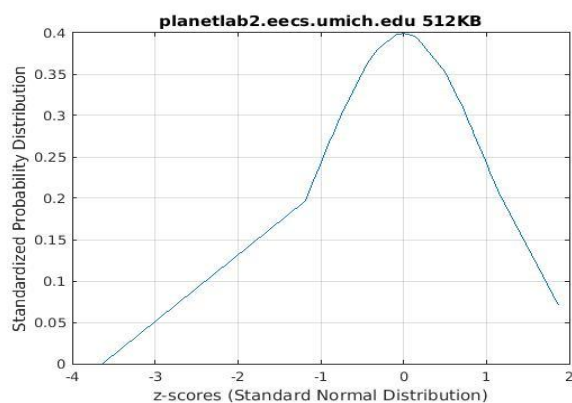
13. planetlab2.eecs.umich.edu and 256KB file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



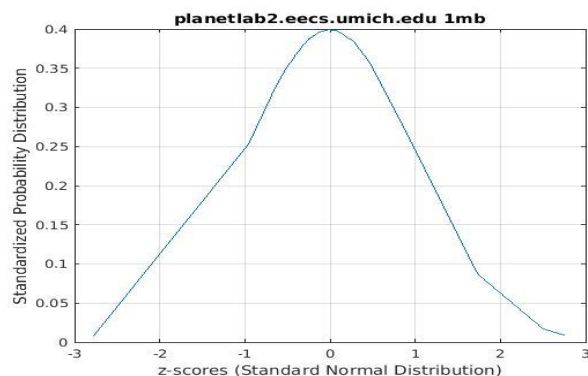
13. planetlab2.eecs.umich.edu and 512KB file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



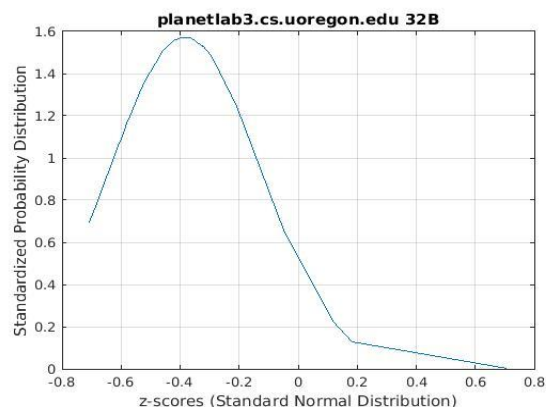
15. planetlab2.eecs.umich.edu and 1MB file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



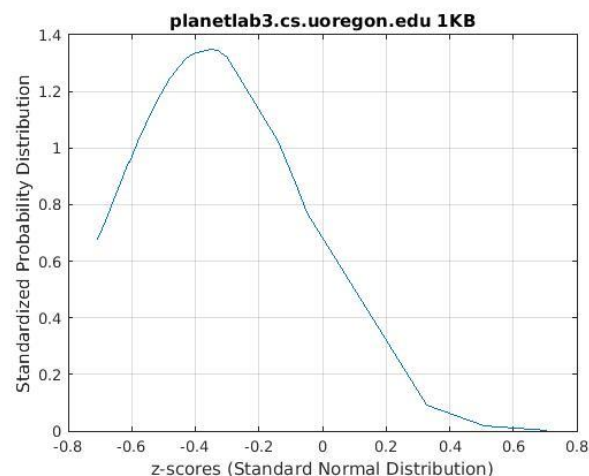
16. planetlab3.cs.uoregon.edu and 32B file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



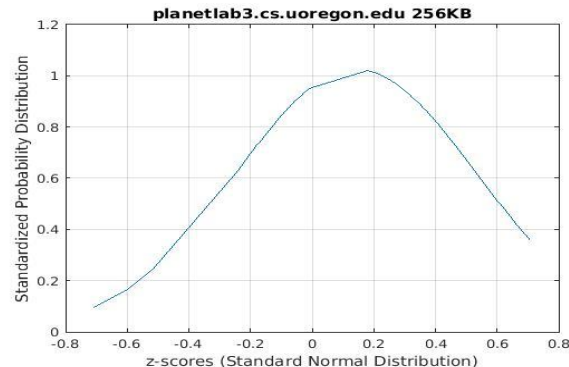
17. planetlab3.cs.uoregon.edu and 1KB file:

X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



18. planetlab3.cs.uoregon.edu and 256KB file:

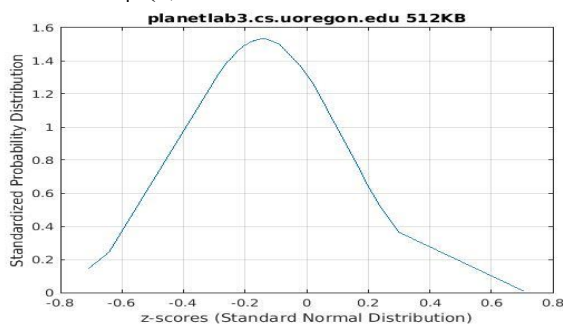
X-axis: z-scores for standard normal distribution
Y-axis: $\phi(z)$



19. planetlab3.cs.uoregon.edu and 512KB file:

X-axis: z-scores for standard normal distribution

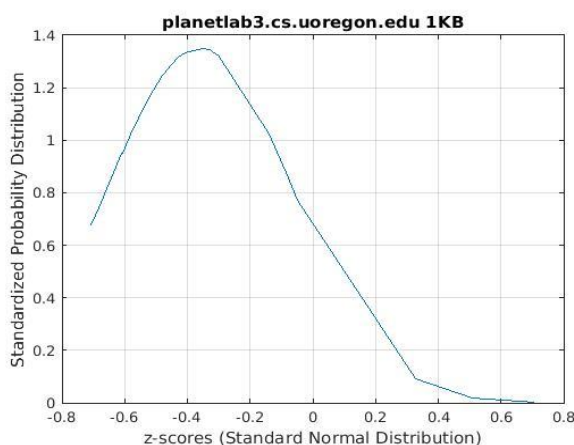
Y-axis: $\phi(z)$



20. planetlab3.cs.uoregon.edu and 1MB file:

X-axis: z-scores for standard normal distribution

Y-axis: $\phi(z)$



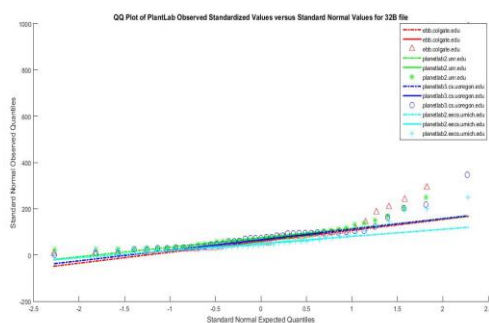
Q-Q plot

Q-Q plot for nodes ebb.colgate.edu,
planetlab2.unr.edu, planetlab3.cs.uoregon.edu,
planetlab3.eecs.umich.edu

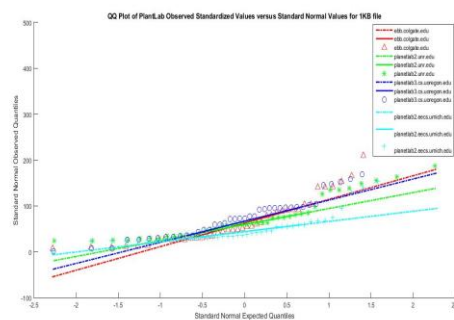
X-axis: Standard Normal expected quantities

Y-axis: Standard Normal observed quantities

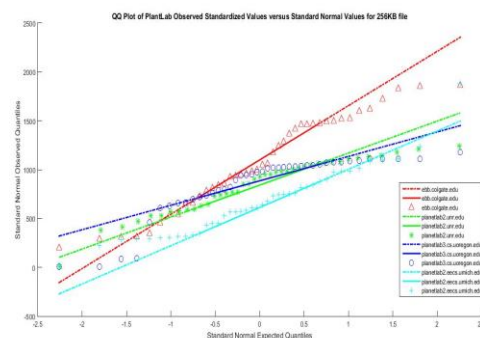
1. Q-Q plot for 32B file



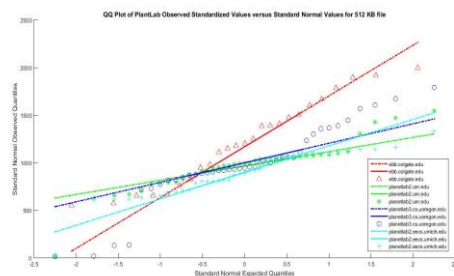
2. Q-Q plot for 1KB file



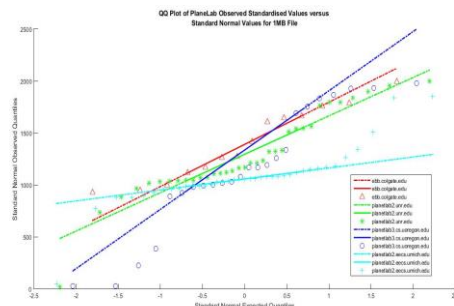
3. Q-Q plot for 256KB file



4. Q-Q plot for 512KB file



5. Q-Q plot for 1MB file



VII. ERROR CORRECTION TECHNIQUES

Any data collected will have some noise in it. To make the data noise free we make use of error correction techniques. These techniques detect the delays/errors in transmission and facilitate error free data. Statistical regression techniques give us a relationship between two variables, one dependent and the other independent. In other words, this regression analysis gives us an idea how a dependent variable varies when an independent variable is held fixed.

The data we collected has to be pruned and corrected. For this we have explored 3 regression techniques namely, Weighted Least Mean Squares (LMSQ), Linear Regression and K Nearest Neighbor.

Weighted Least Mean Squares (LMSQ)

Least squares is a standard method in regression analysis. It is used to the approximate solution of overdetermined systems. They are sets of equations in which there are more equations than unknowns. Least squares indicates that the overall solution reduces the sum of the squares of the errors made in the results of every single equation. Linear least squares is a way fitting a mathematical or statistical model to data. When we have idealized value provided by the model for any data point, it can be expressed linearly in terms of the unknown parameters. The resulting fitted model can be used to compile the data, to find the unobserved values from the same system and to understand the structure that may determine the system. This method is optimum when the sum, S , of squared residuals is a minimum. [2]

$$S = \sum_{i=1}^n r_i^2$$

The most important application of least squares is data fitting. It best fits in minimizing the sum of squared residuals, a residual which is the difference between an experimented value and the fitted value provided by a model. When the problem has substantial doubts in the independent variable, then simple regression and least squares methods have

problems. In such cases, the methodology essential for fitting errors-in-variables models may be measured.

A special case of this generalized least squares is called weighted least squares. It occurs when all the off-diagonal entries of Ω (the correlation matrix of the residuals) are null and the variances of the observations (along the covariance matrix diagonal) are unequal.

Hence, the expressions given above are based on the implicit hypothesis that the errors are not related with each other. Also we observe it with the independent variables and have equal variance. The Gauss–Markov theorem shows that, when this is so, $\hat{\beta}$ is a best linear unbiased estimator.

However, the measurements are uncorrelated but have different indecisions. For this a modified approach might be adopted. Aitken showed that when a weighted sum of squared residuals is minimized, $\hat{\beta}$ is the BLUE if each weight is equal to the reciprocal of the variance of the measurement.

$$S = \sum_{i=1}^n W_{ii} r_i^2, \quad W_{ii} = \frac{1}{\sigma_i^2}$$

The gradient equations for this sum of squares are

$$-2 \sum_i W_{ii} \frac{\partial f(x_i, \beta)}{\partial \beta_j} r_i = 0, \quad j = 1, \dots, n$$

In a linear least squares system give the modified normal equations,

$$\sum_{i=1}^n \sum_{k=1}^m X_{ij} W_{ii} X_{ik} \hat{\beta}_k = \sum_{i=1}^n X_{ij} W_{ii} y_i, \quad j = 1, \dots, m.$$

It is usually supposed that the response data has equal quality and has constant variance. If this is violated, your fit is unduly influenced by data of poor class. To improve this fit, you can use weighted least-squares regression where an additional scale factor like weight is included in the fitting process. Weighted least-squares regression minimizes the error estimate as,

$$S = \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2$$

Where w_i are the weights. The weight determines how much each response value

influences the final parameter estimates. High-quality data points influence the fit more than a low-quality data points. Therefore, weighting your data is commended if the weights are known, or if there is validation that they follow a particular format.

In weighted least mean squares, we calculate the error function at a specific location L for a given source, S , as follows:

For example we are given $rtt(S; H)$, the top K ($= 3$ by default) peaks of the probability distribution of the distance for a small rtt range around $rtt(S; H)$. All our delay measurements are Round Trip Time measurements and not one way delay. For each peak P among the K different peaks, let us compute the error to be $err(P; S; L) = (d(P) - d(S; L))^2 \cdot prob(P)$, where $d(P)$ and $prob(P)$ are referred to the distance value of the peak P and its probability density in the Probability Distribution Function. Among the K different error values, we compute the minimum value of error as the error measure of the location with respect to single source S . Therefore $error(S; L) = \min_P(err(P; S; L))$. The error measure with respect to all sources for a specific location L is the sum total of all the errors from different sources to destinations. This can be shown as, $error(L) = \sum_S error(S; L)$. The LMSQ algorithm accounts the location L with the least value of error (L). The reason for opting the top K peaks is to take into consideration the occurrence of multiple disjoint peaks in the probability distribution function. The $prob(P)$ term was used to evaluate the error inversely to the probability distribution around the peak values.

Weighted least squares regression is not related with a particular type of function used to define the connection between the process variables. On the contrary, weighted least squares reflects the behavior of the casual errors in the model. Also it can be used with functions that have either linear or nonlinear parameters. It is done by integrating additional nonnegative constants or weights linked with each data point, into the fitting criterion. The size of the weight specifies the accuracy of the information present in the accompanying observation. Enhancing the weighted fitting criterion to find the parameter estimations permits the weights to determine the contribution of each observation to the final parameter estimations. It is

significant to note that the weight for each observation is given comparative to the weights of the other observations; so different sets of absolute weights can have identical effects.

Weighted least squares is an efficient method that makes use of small data sets. It shares the ability to provide different types of easily interpretable statistical intervals for assessment, expectation, standardization and optimization. Moreover, the main advantage that weighted least squares has over the other methods is the ability to handle regression conditions in which the data points are of fluctuating quality. If the standard deviation of the random errors in the data is not persistent across all levels of the explanatory variables, using weighted least squares with weights that are inversely proportional to the variance at each level of the explanatory variables yields the most precise parameter estimates possible.

The major disadvantage of weighted least squares, is probably the fact that the theory behind this method is depended on the postulation that the weights are known accurately. This is almost never the case in real applications, so estimated weights must be used in its place. The effect of using estimated weights is difficult to evaluate, but experience specifies that small variations in the weights due to approximation do not often disturb a regression analysis or its explanation. Nevertheless, when the weights are assessed from small numbers of replicated remarks, the results of an analysis can be very badly and randomly affected. This is likely to be the case when the weights for extreme values of the predictor are estimated using only a few observations. It is important to be aware of this potential problem and to only use weighted least squares when the weights can be estimated precisely relative to one another.

K Nearest Neighbor

K Nearest Neighbor is a method that is used for data classification and regression. Here data is classified as training set and test set.

In KNN, basic regression algorithm is that, once a new point is determined, its distance to all its neighbors is calculated, sorted and nearest k -neighbors are determined. For a new data point, mean of its k nearest neighbors will be assigned.

Below is the command in R language that implements the regression for given set of data.

```
knn (train, test, k = 1, use. All = TRUE)
```

Arguments:

Train- a set of data classified based on a condition

Test- second set of the same data that does not classify the condition

k- Number of neighbors to be considered

use.all - handles ties (in case of votes)

Linear Regression

Linear Regression is one of the most popularly used statistical regression technique in practical applications. Models which depend linearly on their unknown parameters have their outputs easier to determine than the ones that are non-linearly dependent. It can be fitted in many ways, but commonly, linear regression fits in using the "least squares" approach.

Let Y be the dependent variable; let x1, x2... xk be the independent variables which should be predicted. The equation for predicting the value can be written as:

$$\hat{Y}_t = b_0 + b_1X_{1t} + b_2X_{2t} + \dots + b_kX_{kt}$$

where bi is the change in the predicted value of Y per unit of change in Xi, other things being equal. The additional constant b0, is the intercept, is the prediction that the model would make if all the X's were zero.

Least squares is used to estimate coefficients and intercepts that is, we set them equal to distinct values to minimize sum of squared errors within the data in which model would be fitted. And the model's prediction errors are typically assumed to be independently and identically normally distributed.

Applying linear regression to our data,

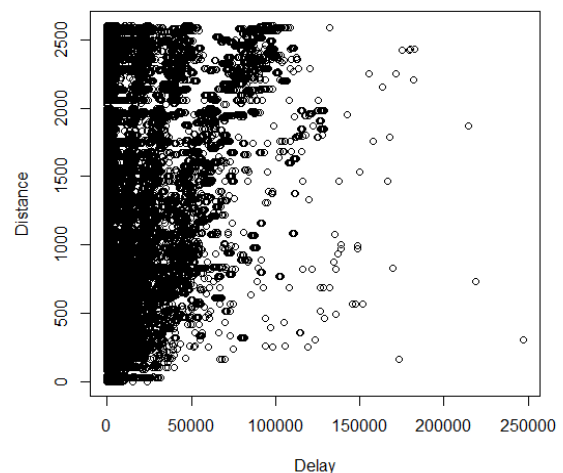
We have selected the data set to be the logset collected on 29th Oct for file size 1MB.

Step 1: Import the log file into R using the command:

```
f = read.csv("download1.csv")
```

Step 2: Plot a scattered graph with this uncorrected data using the dataset "f"

```
plot(f)
```



We see that the data is scattered. The scantily spread data is the noise that has to be corrected.

Step 3: Check for the structure of f
str(f)

```
'data.frame': 191627 obs. of 2 variables:
 $ Delay : Factor w/ 10004 levels "0","1","10","100",...: 1 10011 10034 11406 16466 16466 11644 3654 11644 11644 ...
 $ Distance: num 0 0 0 0 0 0 0 0 ...
```

Step 4: Check for the Delay and distances summary estimations
summary(f)

```
> summary(f)
      Delay      Distance
 0      : 4768   Min.    : 0.0
 49     : 1632   1st Qu.: 588.6
 42     : 1606   Median :1075.2
 60     : 1477   Mean   :1225.0
 34     : 1468   3rd Qu.:1954.2
 44     : 1419   Max.    :2601.9
 (Other):178157  NA's    :439
```

Step 5: Load required libraries for plotting linear regression

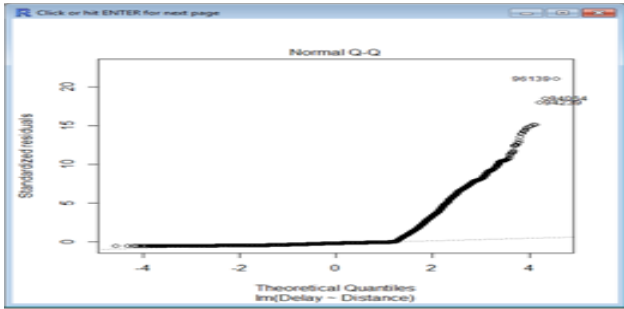
```
load=function() {+ library(MASS) }
```

Step 6: Generate a table from which graph has to be generated for given delay to distance fix (f)

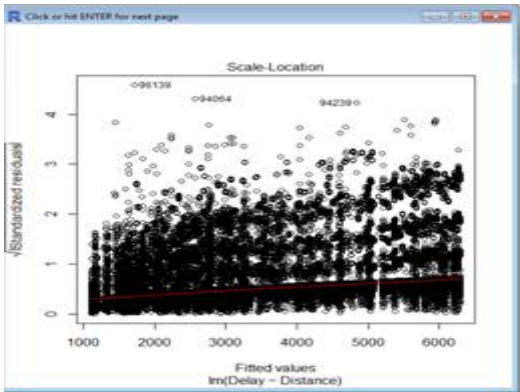
	Delay	Distance	var3	var4	var5	var6
1	0	0				
2	35	0				
3	36	0				
4	39	0				
5	71	0				
6	71	0				
7	4	0				
8	168	0				
9	4	0				
10	4	0				
11	70	0				
12	71	0				
13	71	0				
14	168	0				
15	71	0				
16	71	0				
17	4	0				
18	153	0				
19	4	0				

Step 7: Apply linear regression to the dataset
lm.fit=lm (Delay~Distance)

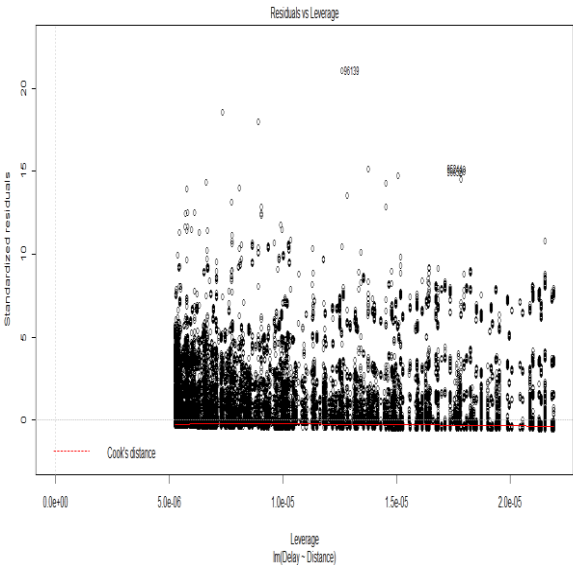
We get the following graphs upon regression[8]



The Q-Q plot gives us the graph based on actual data to estimated values. It gives us a straight line along which all the data is aligned.



This plot is used to know the non-linearity in data. In other words, it measures the error variances. We can see that it corrects the data to be concentrated around the "red" line.



The leverage vs residuals plot gives us the cooks's distance, which estimates how much error would be reduced, if a detected point would be deleted.

VIII. CONCLUSION

From the analysis of collected data, we can conclude that we have achieved good amount of error correction.

REFERENCES

- [1] Inja Youn, Brian L. Mark, Dana Richards, Statistical Geolocation of Internet Hosts. Dept. of Computer Science George Mason University Fairfax, Virginia 22030 USA Email: richards@cs.gmu.edu
- [2] Venkata N. Padmanabhan, Lakshminarayanan, Determining the Geographic Location of Internet, Microsoft research report November 2000.
- [3] https://en.wikipedia.org/wiki/Round-trip_delay_time
- [4] <http://searchnetworking.techtarget.com/definition/round-trip-time>
- [5] <https://tools.ietf.org/html/rfc1323>
- [6] Phillipa Sessini and Anirban Mahanti, Observations on Round-Trip Times of TCP Connections, Phillipa Sessini and Anirban
- [7] Jay Aikat, Jasleen Kaur, F. Donelson Smith, Kevin Variability in TCP Round-trip Times, Department of Computer Science University of North Carolina at Chapel Hill Chapel Hill, NC 27599-3175
- [8] <http://www.statemethods.net/graphs/line.html>
- [9] <http://www.statemethods.net/input/contents.html>
- [10] <http://www.statemethods.net/blog/?p=779>