**Harsh Shah**

[hshah@albany.edu](mailto:hshah@albany.edu)

**Report**

**Project 2: Internet Path Analysis using PlanetLab**

**Abstract:**

The project has basically given practical knowledge about the networking. In this paper I have taken measurements results on the internet path analysis using PlanetLab. The points that made in this report are based on small scale experiment. I have selected 5 different pairs of nodes in PlanetLab which were being accessed for conducting this experiment of running traceroute and ping every consecutive hour and the output information has been fetched and then analysed. With the fetched data, I have been analysing the values of RTT, number of hopes, numbers of packet loss, network outrages etc.

1. **Motivation: Why is Internet path stability important to measure?**

Internet is indistinguishable part of our day to day life. Internet is our main dependent source for day to day activities which we might not have realized. With the growing population, growing technologies, numbers of internet user are increasing. With this there’s also increased internet traffic, which may cause some problems like slow internet connection, delayed website response, unreachable website, Errror 404 and many other. Those problems are mostly caused by unresponsive browser, server overload, heavy traffic in the internet, or the internet routing instability. So, it is important to stable the internet. Also, internet is a network which contains many network. This makes more difficult to stable the internet. For stability, we need to study this field and have to apply appropriate algorithms.

1. **Introduction**

In order to measure the internet path stability I have used PlanetLab nodes to have a measurement among the nodes which are geographically spread in different range. Initial step of the project was to select the nodes that are in the working condition. Also for proper measurement and cover all around the world, I have chosen some trans-pacific as well trans-Atlantic nodes. Also I have selected some nodes that are present in USA and also taken node that are in Canada and Brazil.

**List of Nodes:**

**pl1.sos.info.hiroshima-cu.ac.jp (Trans-Pacific Node, Japan)**

**planetlab1.pop-mg.rnp.br (Brazil)**

**planetlab4.mini.pw.edu.pl (Trans-Atlantic, Poland)**

**ricepl-1.cs.rice.edu(USA)**

**planetlab1.cs.ucla.edu(USA)**

**planetlab1.cs.ubc.ca(Canada)**

**planetlab-03.cs.princeton.edu(USA)**

**planetlab1.cs.uoregon.edu(USA)**

**planetlab1.cs.uml.edu(USA)**

* **Traceroute** is a utility that records the route (the specific gateway computers at each hop) through the Internet between your computer and a specified destination computer. It also calculates and displays the amount of time each hop took.
* **Ping** is the name of a standard software utility used to test network connections. It can be used to determine if a remote device like a website or game server can be reached across the network and, if so, the connection's [latency](https://www.lifewire.com/latency-on-computer-networks-818119).

For internet measurements, I have used these tools. After the initial step of selecting node, the next step that I implemented is logged into every node and used these two tools to reach other node and get the details for output. I have collected data for almost 11 days. Due to conflicts in data in the initial period I was not able to collect all the data. Then in the second phase I have to analyse the data for that I am going to use Gawk and python script.

**Methodology:**

In this methodology, there are two phases. First phase includes collection of the data and second phase include analysis of the data. The main technical equipment for this experiment is PlanetLab Nodes. I have collected data of **Ping** and **Traceroute** between nodes for almost over a period of 2 weeks. For analysis, I have used gawk which is very powerful tool. And for the representation and to generate graph I have use Excel. Below are pairs of nodes that I have used for experiment

**Nodes Pair**

|  |  |
| --- | --- |
| 1 | planetlab1.cs.ucla.edu <--> pl1.sos.info.hiroshima-cu.ac.jp |
| 2 | planetlab1.pop-mg.rnp.br <--> planetlab4.mini.pw.edu.pl |
| 3 | planetlab1.cs.uml.edu <--> planetlab1.cs.ubc.ca |
| 4 | planetlab-03.cs.princeton.edu <--> planetlab1.cs.uoregon.edu |
| 5 | planetlab1.cs.ucla.edu <--> planetlab1.cs.uml.edu |

**3.1 Measurement Framework**

In this first framework, firstly I had selected 10 nodes and pair up them into 5. Also, I have checked that all the nodes that I have selected are in the working condition. Then with basic tools I have implemented the automatic script that will ping the other node in the pair as well as gather the data for the traceroute and also it stores the data of both the tools in a separate file on the node itself. I have collected the data for almost 11 days, which is sufficient to draw the conclusion. Here is the example of one of my shell script that I have implemented on every node:

Example:

**#!/bin/bash**

**while true; do**

**echo "-----------------------------------------------------------------" | tee -a hspair3t.txt >> hspair3.txt**

**echo "Date: $(date)" | tee -a hspair3t.txt >> hspair3.txt**

**echo "-----------------------------------------------------------------" | tee -a hspair3t.txt >> hspair3.txt**

**ping planetlab4.mini.pw.edu.pl -c 20 >> hspair3.txt**

**traceroute planetlab4.mini.pw.edu.pl >> hspair3t.txt**

**sleep 3600**

**done**

**3.2 Analysis Framework**

The second phase in the methodology is the final phase. In this phase, I have analysed the data that has been stored in the files at different nodes. For the internet stability, I have list down the parameters that I want and using tools like gawk and some mathematical functions I have analysed the data.

For calculations of min, max, avg values of all the reading for a particular node, I have imported the data of my text file into Excel and then performed some mathematical calculations for the output. For the path length, I have used gawk command. Using gawk command that is mentioned below, I have extracted the hop count (Path length).

cat /home/harsh/Desktop/trace/hspair7t.txt | gawk '(/traceroute/) {print a} {a=$1}' > /home/harsh/Desktop/hop/hspair7t.txt

This is just the example, I have implemented this for all the traceroute file.

I have implemented this in Linux using Virtual Machine to get the data.

After the execution of the above-mentioned command, it will give me the last hop count for every traceroute reading.

**Flow of my process:**

Measured Data 🡪 Stored it into the text file 🡪 Import the data in Excel 🡪 Create Tables for mathematical functions 🡪 Created chart for visualization.

**4 Analysis**

In the analysis part, I have stated some observations that are needed for Internet Path Stability.

**4.1 Path Length and Variations in measurements**

The path length is same as hop count which is given by traceroute. I have analysed the path length, min, max, avg and standard deviation for all the pair and the result is stated below in the table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pair Number | Pair | Minimum | Average | Maximum | Standard Deviation |
| 1 | planetlab1.cs.ucla.edu 🡪 pl1.sos.info.hiroshima-cu.ac.jp | 16 | 16.11059 | 18 | 0.447966 |
| pl1.sos.info.hiroshima-cu.ac.jp 🡪 planetlab1.cs.ucla.edu | 16 | 16.10138 | 18 | 0.439747 |
| 2 | planetlab1.pop-mg.rnp.br 🡪 planetlab4.mini.pw.edu.pl | 20 | 20.03225 | 22 | 0.223339 |
| planetlab4.mini.pw.edu.pl 🡪 planetlab1.pop-mg.rnp.br | 19 | 19.01843 | 21 | 0.191561 |
| 3 | planetlab1.cs.uml.edu 🡪 planetlab1.cs.ubc.ca | 18 | 18.04608 | 20 | 0.300764 |
| planetlab1.cs.ubc.ca 🡪 planetlab1.cs.uml.edu | 18 | 18.04608 | 20 | 0.300764 |
| 4 | planetlab-03.cs.princeton.edu 🡪 planetlab1.cs.uoregon.edu | 14 | 14.02030 | 15 | 0.150382 |
| planetlab1.cs.uoregon.edu 🡪 planetlab-03.cs.princeton.edu | 11 | 11 | 11 | 0 |
| 5 | planetlab1.cs.ucla.edu 🡪 planetlab1.cs.uml.edu | 18 | 18 | 18 | 0 |
| planetlab1.cs.uml.edu 🡪 planetlab1.cs.ucla.edu | 18 | 18 | 18 | 0 |

**From the observation in above table, we can see that in the pair 4 the path length varies by a big margin. The average path length for** planetlab-03.cs.princeton.edu 🡪 planetlab1.cs.uoregon.edu is 14.02 where average path length for planetlab1.cs.uoregon.edu 🡪 planetlab-03.cs.princeton.edu is 11. So this is an exception case where node select different route to travel back.

There are mostly two reasons behind it. Change in server or load-balancing.

**Analysis on ping data:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pair Number | Pair | Minimum | Average | Maximum | Standard Deviation |
| 1 | planetlab1.cs.ucla.edu 🡪 pl1.sos.info.hiroshima-cu.ac.jp | 113.3414 | 113.6661 | 115.8962 | 0.829717 |
| pl1.sos.info.hiroshima-cu.ac.jp 🡪 planetlab1.cs.ucla.edu | 112.737 | 113.062 | 114.6629 | 0.69653 |
| 2 | planetlab1.pop-mg.rnp.br 🡪 planetlab4.mini.pw.edu.pl | 300.5654 | 301.1239 | 302.3737 | 0.909532 |
| planetlab4.mini.pw.edu.pl 🡪 planetlab1.pop-mg.rnp.br | 298.84 | 299.4648 | 300.4573 | 0.82645 |
| 3 | planetlab1.cs.uml.edu 🡪 planetlab1.cs.ubc.ca | 73.41511 | 74.46194 | 79.10286 | 1.510147 |
| planetlab1.cs.ubc.ca 🡪 planetlab1.cs.uml.edu | 73.3248 | 74.43089 | 79.0749 | 1.507872 |
| 4 | planetlab-03.cs.princeton.edu 🡪 planetlab1.cs.uoregon.edu | 67.55158 | 67.72674 | 68.18644 | 0.314147 |
| planetlab1.cs.uoregon.edu 🡪 planetlab-03.cs.princeton.edu | 67.48505 | 67.61637 | 68.08734 | 0.293053 |
| 5 | planetlab1.cs.ucla.edu 🡪 planetlab1.cs.uml.edu | 70.15861 | 70.90465 | 76.14852 | 1.579219 |
| planetlab1.cs.uml.edu 🡪 planetlab1.cs.ucla.edu | 70.14987 | 70.64021 | 74.06118 | 1.033968 |

From the above observation, we can see that with the increase in geographical distance the average, minimum and maximum increases. And we can see the changes in the path length over the different measurements because the time is depending, on the distance travelled. Or in the other words we can say that the values of maximum, minimum and average is directly proportional to the distance.

**4.2 Detection of Outrages**

|  |  |  |  |
| --- | --- | --- | --- |
| Pair Number | Pair | Nick Name | Outrages |
| 1 | planetlab1.cs.ucla.edu 🡪 pl1.sos.info.hiroshima-cu.ac.jp | Node 1 | 197 |
| pl1.sos.info.hiroshima-cu.ac.jp 🡪 planetlab1.cs.ucla.edu | Node 1R | 36 |
| 2 | planetlab1.pop-mg.rnp.br 🡪 planetlab4.mini.pw.edu.pl | Node 2 | 2 |
| planetlab4.mini.pw.edu.pl 🡪 planetlab1.pop-mg.rnp.br | Node 2R | 445 |
| 3 | planetlab1.cs.uml.edu 🡪 planetlab1.cs.ubc.ca | Node 3 | 166 |
| planetlab1.cs.ubc.ca 🡪 planetlab1.cs.uml.edu | Node 3R | 219 |
| 4 | planetlab-03.cs.princeton.edu 🡪 planetlab1.cs.uoregon.edu | Node 4 | 0 |
| planetlab1.cs.uoregon.edu 🡪 planetlab-03.cs.princeton.edu | Node 4R | 245 |
| 5 | planetlab1.cs.ucla.edu 🡪 planetlab1.cs.uml.edu | Node 5 | 219 |
| planetlab1.cs.uml.edu 🡪 planetlab1.cs.ucla.edu | Node 5R | 581 |

We can calculate the total numbers of outrages for each and every node with the help of the gawk tool. By this tool, we can figure out at what hop count the problem occurred and couldn’t processed further. From the chart, we can see that in the pair 5 where both of the node are present in USA still they both have highest number of Outrages in their path. On the other side, Pair -1 (Trans-Atlantic) and Pair-2 (Trans-Pacific) have less number of outrages in their path.

From above, we can say that outrages are not depends on the geographical distance. For example, the distance for Trans-Pacific node (Pair 2) is quite high compare to other node so it is not related to distances. There are two type of outrages.

1. Temporary Outrages
2. Long-Term Outrages

**Temporary Outrage**

When a sequence of probes are lost in traceroute, it may be either due to temporary network loss or heavy congestion in the network. In this I have experienced many temporary outrages. The above-mentioned figure is temporary outrages on every node.

**The traceroute doesn’t provide the details of the cause of the outrage.**

**Long-Term Outages**

Long-term outrages which I came across is access denied by the node due to change in credentials. The planetlab keeps updating their certificates due to certain issues or sometime for maintenance they disable the node so these are the type of Ling-Term Outrages. I have performed this experience for shorter period of 14 days so I haven’t faced any Long-term Outages issues while my experiment. In the initial stage only I got outages by nodes because they have problem with some authentication.

**4.3 Core vs Edge in Outages**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Edge | Edge % | Core | Core % |
| Pair 1 | Node 1 | 190 | 96.93 | 6 | 3.07 |
|  | Node 1R | 0 | 0 | 36 | 100 |
| Pair 2 | Node 2 | 0 | 0 | 2 | 100 |
|  | Node 2R | 0 | 0 | 445 | 100 |
| Pair 3 | Node 3 | 145 | 87.34 | 21 | 12.65 |
|  | Node 3R | 1 | 0.45 | 218 | 99.54 |
| Pair 4 | Node 4 | 0 | 0 | 0 | 0 |
|  | Node 4R | 0 | 0 | 245 | 100 |
| Pair 5 | Node 5 | 0 | 0 | 219 | 100 |
|  | Node 5R | 144 | 24.53 | 437 | 75.47 |

The outages that fell under the local ISP falls under the Edge outages. While the outages that are not fall under the local ISP and falls other than the local ISP are called Core outages.

**4.4 Different continents links**

In this experiment, we will check the significance difference in the reliability of continent links. From this experiment we can check if there any significant delay is there or not. So we have to analyse the packet loss among the different nodes.

Also, there are no significant differences in the reliability of continental link vs inter-continental link.

|  |  |  |  |
| --- | --- | --- | --- |
| Pair Number | Pair | Nick Name | Average % error |
| 1 | planetlab1.cs.ucla.edu 🡪 pl1.sos.info.hiroshima-cu.ac.jp | Node 1 | 0.355505 |
| pl1.sos.info.hiroshima-cu.ac.jp 🡪 planetlab1.cs.ucla.edu | Node 1R |
| 2 | planetlab1.pop-mg.rnp.br 🡪 planetlab4.mini.pw.edu.pl | Node 2 | 0.126147 |
| planetlab4.mini.pw.edu.pl 🡪 planetlab1.pop-mg.rnp.br | Node 2R |
| 3 | planetlab1.cs.uml.edu 🡪 planetlab1.cs.ubc.ca | Node 3 | 0.068807 |
| planetlab1.cs.ubc.ca 🡪 planetlab1.cs.uml.edu | Node 3R |
| 4 | planetlab-03.cs.princeton.edu 🡪 planetlab1.cs.uoregon.edu | Node 4 | 0 |
| planetlab1.cs.uoregon.edu 🡪 planetlab-03.cs.princeton.edu | Node 4R |
| 5 | planetlab1.cs.ucla.edu 🡪 planetlab1.cs.uml.edu | Node 5 | 0.045872 |
| planetlab1.cs.uml.edu 🡪 planetlab1.cs.ucla.edu | Node 5R |

**4.5 Fluttering**

Fluttering refers to rapidly-oscillating routing. In this experiment, for every pair of node there’s no fluttering.

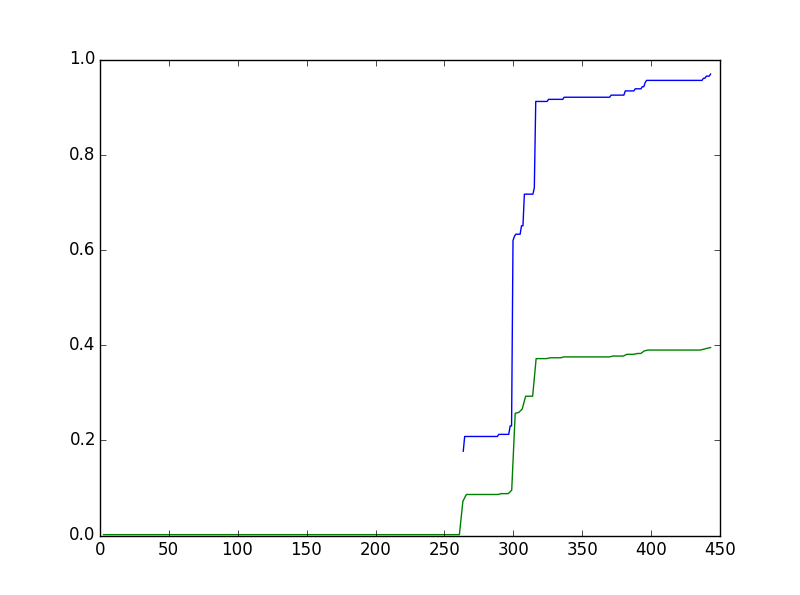
**4.6 Inconsistencies**

There’s no inconsistencies between any node.

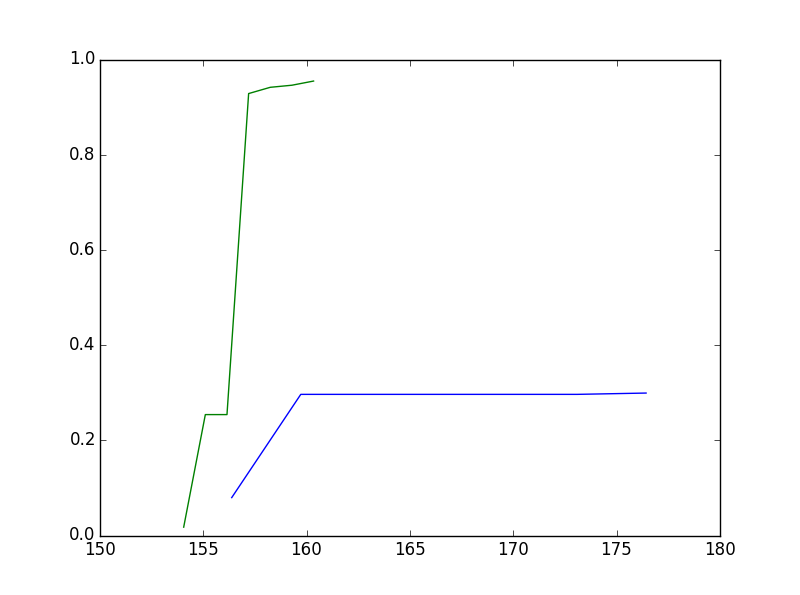
**4.7 Triangular Routing**

There is no triangular routing in the fetched data. As I did this experiment of the limited time of 2 weeks so there’s no triangular routing.

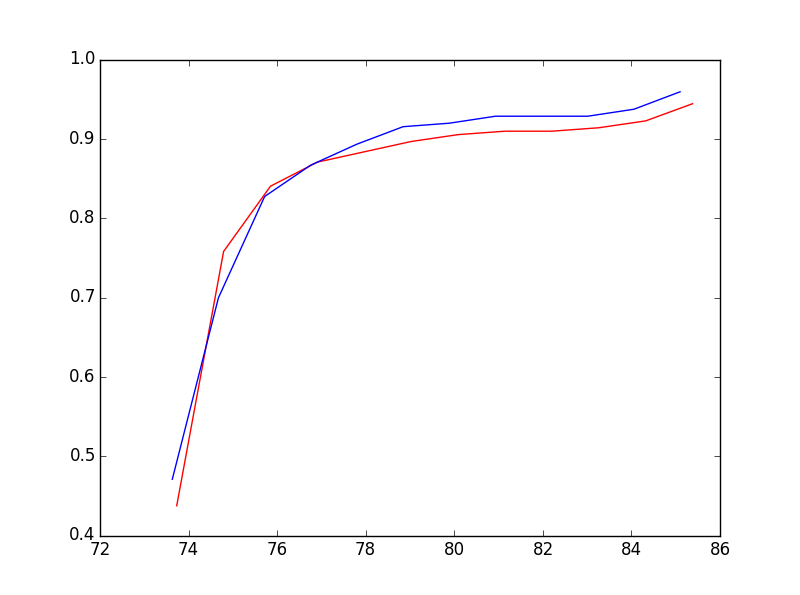
**4.8 Average Latency Pairwise**

****

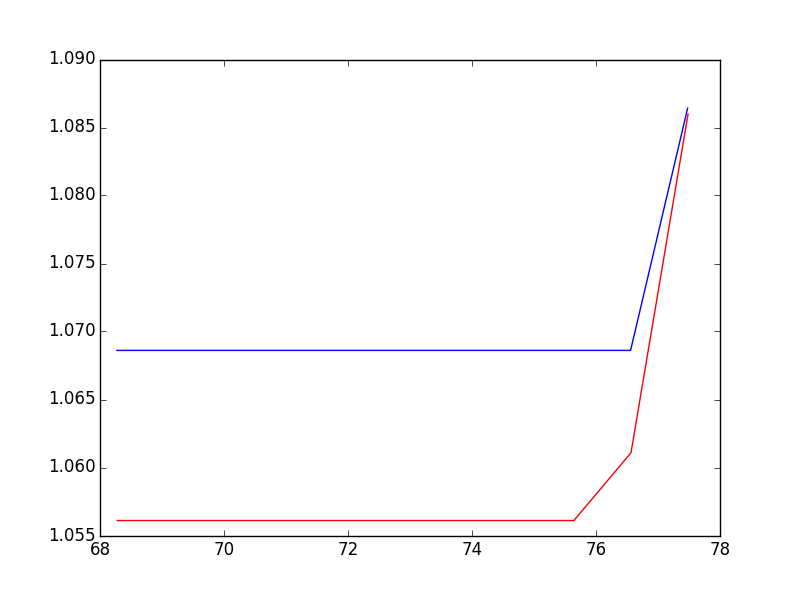
This is the average latency graph for my pair-1. Here Blue colour indicates the path Node-1🡪Node1R and Green colour indicates Node1🡨Node1R.

****

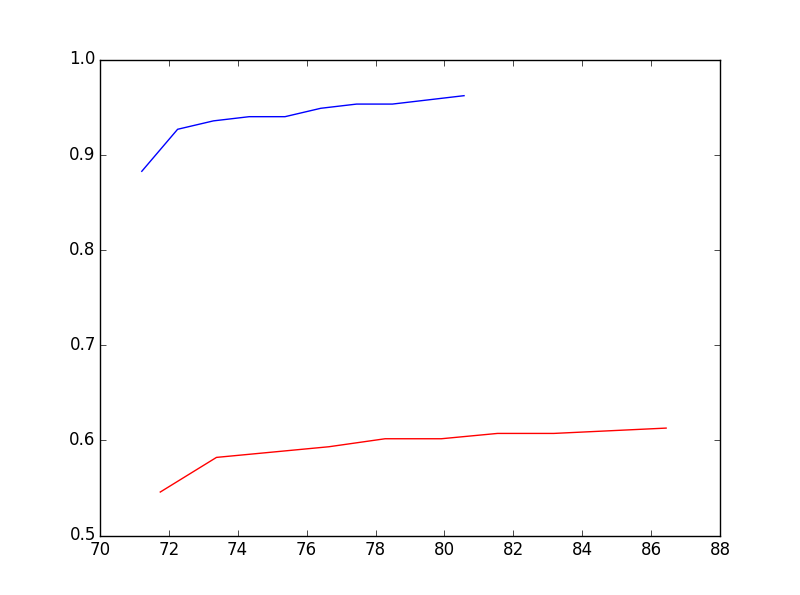
This is the average latency graph for my pair-2. Here Blue colour indicates the path Node-2🡪Node2R and Green colour indicates Node2🡨Node2R.



This is the average latency graph for my pair-3. Here Blue colour indicates the path Node-3🡪Node3R and Red colour indicates Node3🡨Node3R.

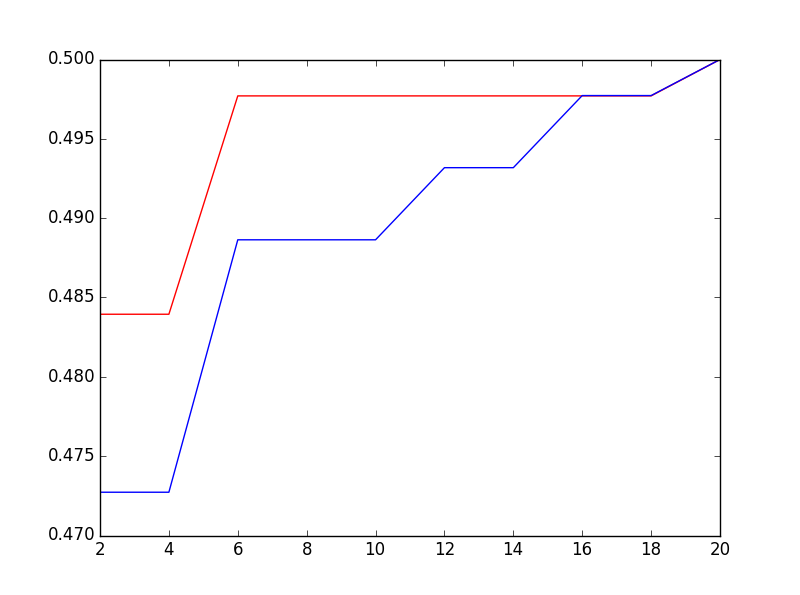


This is the average latency graph for my pair-4. Here Blue colour indicates the path Node-4🡪Node4R and Red colour indicates Node4🡨Node4R.

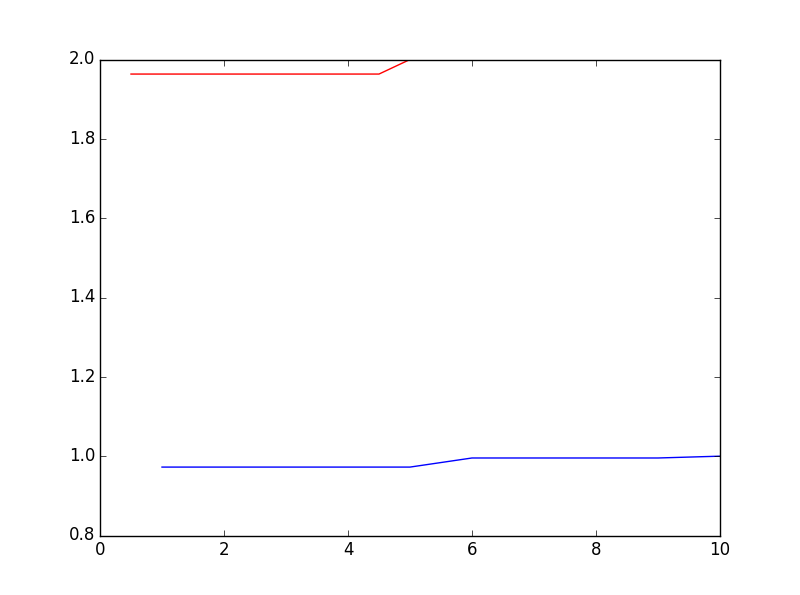


This is the average latency graph for my pair-5. Here Blue colour indicates the path Node-5🡪Node5R and Red colour indicates Node5🡨Node5R.

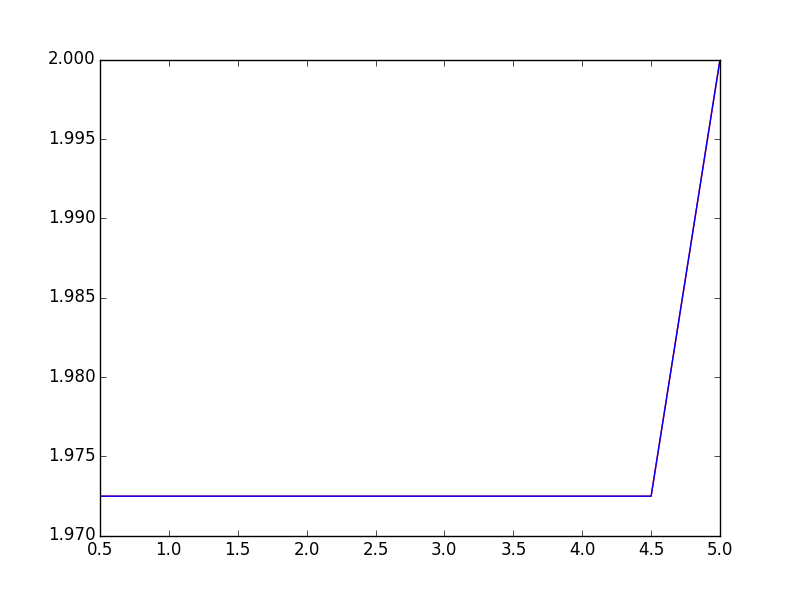
**6.2 Average Packet loss in Pairwise**

****

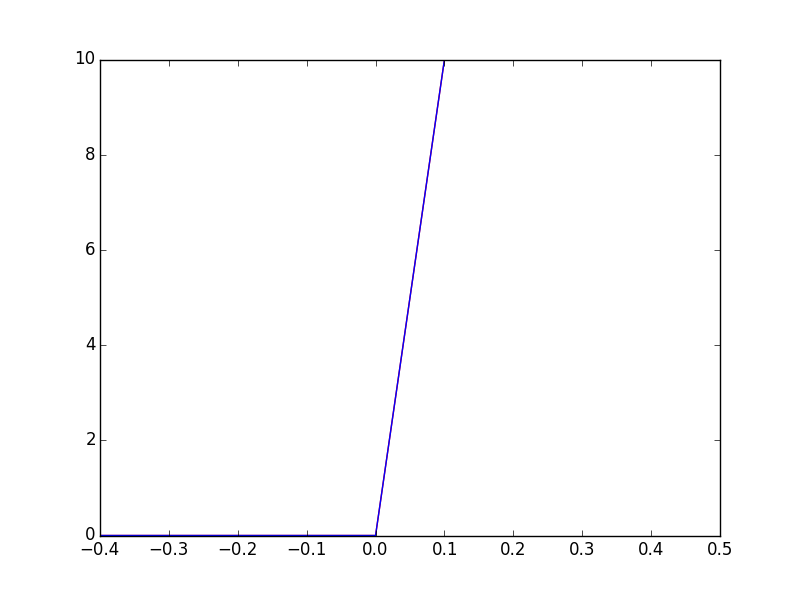
This is the average packet loss graph for my pair-1. Here Blue colour indicates the path Node-1🡨Node1R and Red colour indicates Node1🡪Node1R.

****

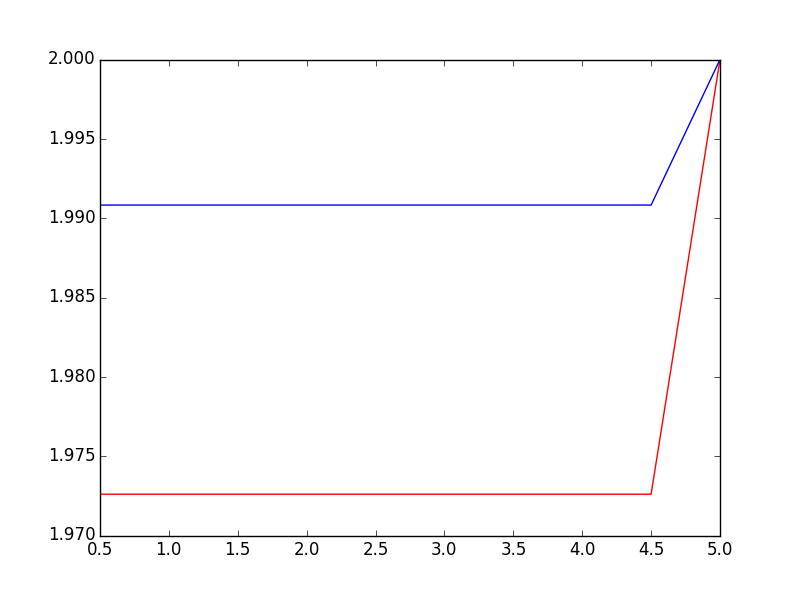
This is the average packet loss graph for my pair-2. Here Blue colour indicates the path Node-2🡨Node2R and Red colour indicates Node2🡪Node2R.



This is the average packet loss graph for my pair-3. Here Blue colour indicates the path Node-3🡨Node3R and Red colour indicates Node3🡪Node3R. (Here, Red Colour is not in the graph because the average packet loss for this node is zero).



This is the average packet loss graph for my pair-4. Here Blue colour indicates the path Node-4🡨Node4R and Red colour indicates Node4🡪Node4R. (Here, Red Colour is not in the graph because the average packet loss for this node is zero).



This is the average packet loss graph for my pair-5. Here Blue colour indicates the path Node-5🡨Node5R and Red colour indicates Node5🡪Node5R. (Here, Red Colour is not in the graph because the average packet loss for this node is zero).

**Conclusion**

I have conducted my analysis on 10 nodes that have been geographically spread, almost over the period of two weeks. This study includes the path length and variations over these measurements. The next I study is the Outrages. And we also covered the type of the Outages (i.e. Temporary Outages & Permeant Outages.) Mainly, the focus is on the Temporary outages. Also, we covered fluttering. Our statistical methodology helps us to know the probability of the latency at a given path.

Based on the collected data, we cannot say strongly whether internet stability has increased or decreased. For a particular answer, we need more data and we have to do this analysis on the big scale. But from the gathered data we can say that packet loss of around 0.03% and temporary outage of 7% on whole hopes seems to be stabilized internet routing.