Name : Viral Patel

Email : vdpatel@albany.edu

Motivation : The tremendous growth of internet in the last few decades proved that as internet increases the scalability increases. For scalability, We use routers. Routers decides the paths in which the packets or the data has to go from source to destination. In this Project, we are doing the measurements of path between two nodes. We uses PING and traceroute commands to get the date for the analysis of path and from which we will be able to understand that in a Long traverse of packet how much is the failure rate, about how many percentage the packet failure happens, What is the min, max, avg, dev value for the average packets. In short, We do analyze the path between two nodes which are geographically two far from each other. In this project, we are doing the three geographically distributed path measurements. One is between USA -> Europe [transatlantic] , Second is USA -> Asia, China, Russia, Australia etc. [transpacific] and third is between North - > South USA or Canada and nearby areas.

Methodology :

- The measurement framework I programmed in PYTHON.
- Here, I started taking data from the pair of nodes for the 5 pairs.
- I do have many back up nodes too so that if some node stop working then I can use another.
- For the transatlantic nodes, I had backup for extra 3 nodes but due to some error I was not able to take the data and for transatlantic I have few data only.
- So, In the analysis , I used 6 pairs of nodes. One is transatlantic, two are transpacific and others are between local nodes.
- First step for this project was to gether the data between two pairs.
- I first uploaded my SSH key of Laptop to the planetlab account.
- Then, I am using Git Bash software as the middleware to collect all the data which is collected on the nodes.
- The first script I created in Python was to check How many nodes from the pool of nodes on planetlab are alive.
- Then I choose pairs from that list according to the uptime of the nodes.
- Then, I made the script which will run in the background in all the nodes which basically contains the ping and traceroute command output. This script runs every 1 hour.
- After that, I had created the python script to analyze the output data which will take the input as ping data file and generate the csv file which will contain the min, max, avg , dev for all the ping.
- One more script I wrote to analyze the output of the traceroute which will take traceroute file as input and generate csv which will contain that how many times the packet is dropped in between.
- One more script I wrote to count the total no. of packet failures in Each node.

Analysis :

- I have created the script.py file which will continuously run on the every node and contain the ping and traceroute output.
- On the each node, I am generating the data by the directory name as timestamp and for every hour.
- This script will generate output of ping and traceroute day by day.
- From the generated text files, I have created the script to parse through ping output and generate new file which "data.txt" contains the min, avg, max, dev.
- I have created the csv file which contains the min, max, avg, dev from the above generated data.txt.
- So, I am putting the equation to find the average of all the four values and the table is as shown below.

	Min	Max	Avg	Dev
planetlab5.eecs.umich.edu	130.5438	130.6391	130.982	0.37
pl2.prakinf.tu-ilmenau.de	130.5662	130.6624	131.0417	0.368542
planetlab04.cs.washington.edu	171.4347	173.1868	177.6934	1.976478
PLANETLAB-N2.WAND.NET.NZ	171.4452	173.9728	180.2398	2.841421
planetlab4.cs.uoregon.edu	178.6651	180.3065	185.0597	2.048474
pl1.6test.edu.cn	178.7377	180.8094	186.7304	2.543193
PLANETLAB1.CS.PURDUE.edu	34.78108	51.96957	142.5496	27.46327
PLONK.CS.UWATERLOO.CA	34.78168	54.98575	158.2366	32.19167
PLANETLAB2.UTDALLAS.EDU	41.02157	41.18827	42.06094	0.353989
planetlab-3.cmcl.cs.cmu.edu	40.5249	40.71349	40.71349	0.366483
PLANETLAB1.CS.UCLA.EDU	24.58488	24.91569	25.68833	0.393605
planetlab4.cs.uoregon.edu	24.66139	24.9973	25.67436	0.404332

- As you can see from above table, the transatlantic and transpacific node's packets are getting too much time to travel then the one which are local.

- For Path failure, Again I made two scripts. First in which, I am writing all the lines in which one or more drop of packets are there.
- After getting this list I made another script for total count of all the failure packets by each nodes.
- In my Library inside the directory of all the node, I will have one of the three directory. Local, Transatlantic or Transpacific. Also, It will contain the script.sh file inside of all three directories which is the script which I ran on each nodes to get the data.
- Inside that directory I have two more directories named ping and traceroute which contains the ping output and traceroute output respectively.
- Inside Ping Directory, script.py files contain the script to generate the csv output to the data.txt. ping.txt contain the full output of ping. Csvscript.py generates the csv file from the data.txt file.
- Data.csv contain all the required values of min, max, avg and dev.
- The same applied to the directory traceroute. In addition, traceroute directory contains more files named count.py which will generate the total no of failure packets [no of stars] from the data.txt and writes it to the count.txt. csvcount.csv contain the data of count.txt. edge_failure_script.py file contain the script to check the nodes which are at edge and has packet failures and write it to edgeFailure.txt. csvedge.csv contains the csv data of the edgeFailure.txt file.

Node	Path	Total No of Failure
planetlab5.eecs.umich.edu	Transatlantic	0
pl2.prakinf.tu-ilmenau.de	Transatlantic	0
planetlab04.cs.washington.edu	Transpacific	0
PLANETLAB-N2.WAND.NET.NZ	Transpacific	0
planetlab4.cs.uoregon.edu	Transpacific	40
pl1.6test.edu.cn	Transpacific	2
PLANETLAB1.CS.PURDUE.edu	Local	0
PLONK.CS.UWATERLOO.CA	Local	203
PLANETLAB2.UTDALLAS.EDU	Local	51
planetlab-3.cmcl.cs.cmu.edu	Local	3
PLANETLAB1.CS.UCLA.EDU	Local	0
planetlab4.cs.uoregon.edu	Local	0

- Of the Failures I detected below is the status.

Node	Path	Total No of packet Failure	Edge Failure	Core Failure
planetlab5.eecs.umich.edu	Transatlantic	0	0	0
pl2.prakinf.tu-ilmenau.de	Transatlantic	459	0	459
planetlab04.cs.washington.edu	Transpacific	0	0	0
PLANETLAB-N2.WAND.NET.NZ	Transpacific	1554	0	1554
planetlab4.cs.uoregon.edu	Transpacific	994	0	994
pl1.6test.edu.cn	Transpacific	19	5	14
PLANETLAB1.CS.PURDUE.edu	Local	7032	391	6641
PLONK.CS.UWATERLOO.CA	Local	668	148	520
PLANETLAB2.UTDALLAS.EDU	Local	867	0	867
planetlab-3.cmcl.cs.cmu.edu	Local	3	0	3
PLANETLAB1.CS.UCLA.EDU	Local	0	0	0
planetlab4.cs.uoregon.edu	Local	1557	0	

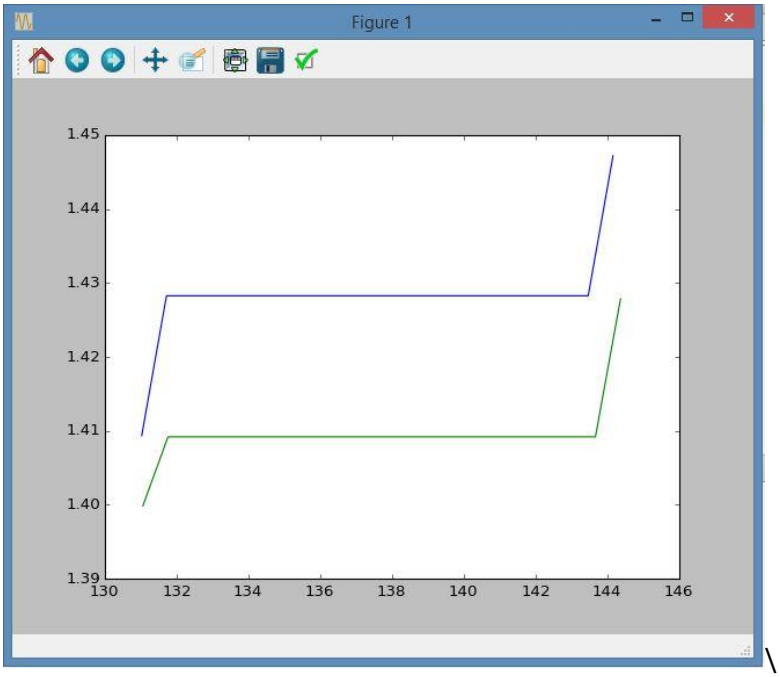
In above table the total no of packet failure include all the failures. Like , In one line if we have two * then it will count as 2 packet failure.

- The edge failure I calculated using the script edge_failure_script.py which will check the failure data file data.txt and check whether it is the first node which is reason for failure of packets then it will count that as edge failure. The result of this count will be generated in edgeFailure.txt.
- I cant see any significant differences in reliability of continental links vs inter-continental links. Because as you can see from the above table that the most failure outages are from the united states only so there's no difference for the continental or inter-continental links.
- When I did traceroute from PLANETLAB1.CS.PURDUE.edu = PLONK.CS.UWATERLOO.CA, I can see some fluttering from line 6536 and one more before that. It's not often. How I decide the fluttering?
- Moreover, When I did traceroute from PLANETLAB2.UTDALLAS.EDU -> PLANETLAB-3.CMCL.CS.CMU.EDU, at line number 1249 the path is changing the direction mean it is going through different hops.
- I created the below script to count the hops between two nodes. If there is sudden path change in the output then it is fluttering.
- Actually, I made the script to count the hopcount from the traceroute.txt file. I took the reference from the http://stackoverflow.com/questions/13680769/counting-jumpno-of-lines-between-first-two-string-occurrences-in-a-file to make the script. In this script I am getting the answer as hopcount + 7[for extra pattern.]. Mean, If I am getting the hopcount as 15 by script then actual hopcount will be 15-7 = 8. The script is named hopcount.py and generated hopcount.py after execution.

- I can see some inconsistencies. When I do traceroute from PLANETLAB1.CS.PURDUE.edu → PLONK.CS.UWATERLOO.CA, from first to second node it takes 24 hopes but from second to first it takes 21 hopes.
- No, I checked manually but couldn't find any triangular routing.
- Below are the Graphs for each of the latency averages I calculated in the above generated data.csv.

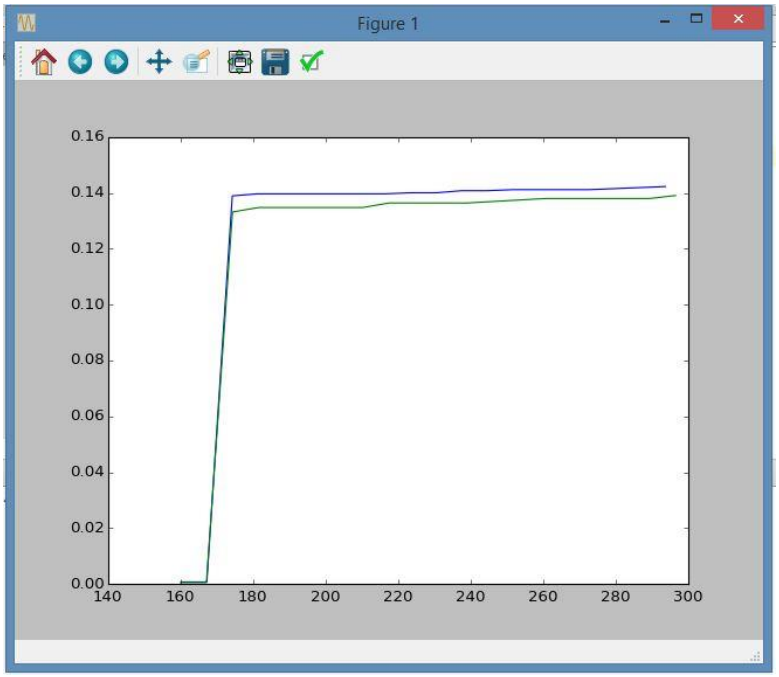
Latency Graph ::

1) planetlab5.eecs.umich.edu-pl2 → prakinf.tu-ilmenau.de

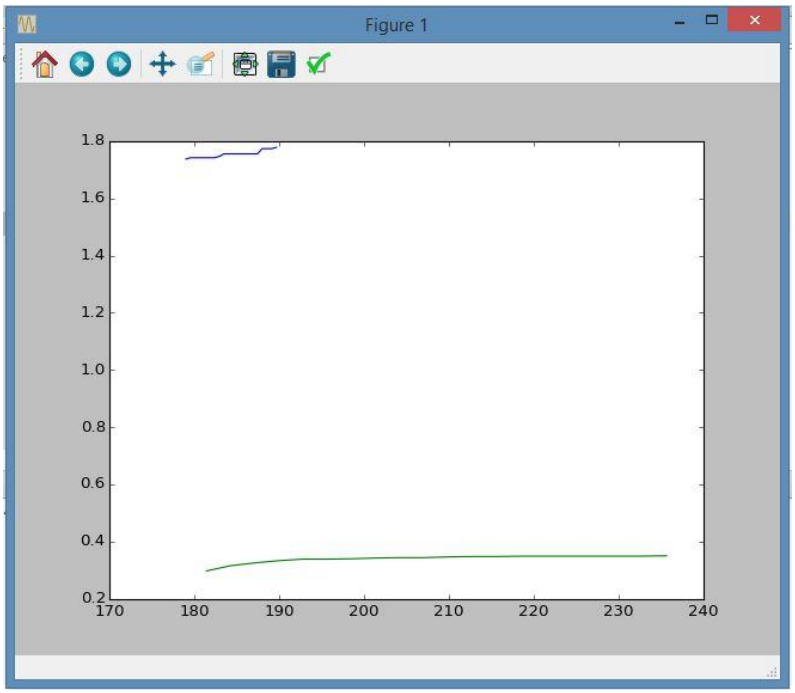


\

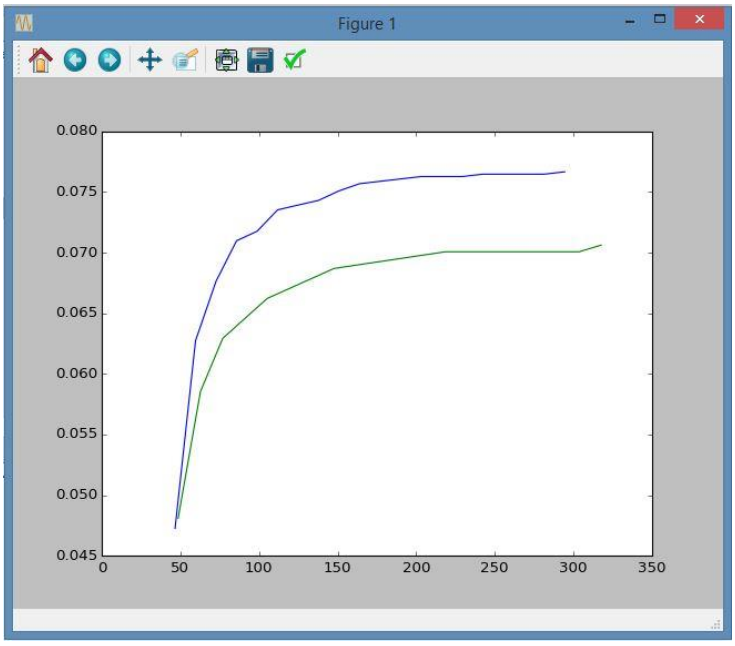
2) planetlab04.cs.washington.edu → PLANETLAB-N2.WAND.NET.NZ



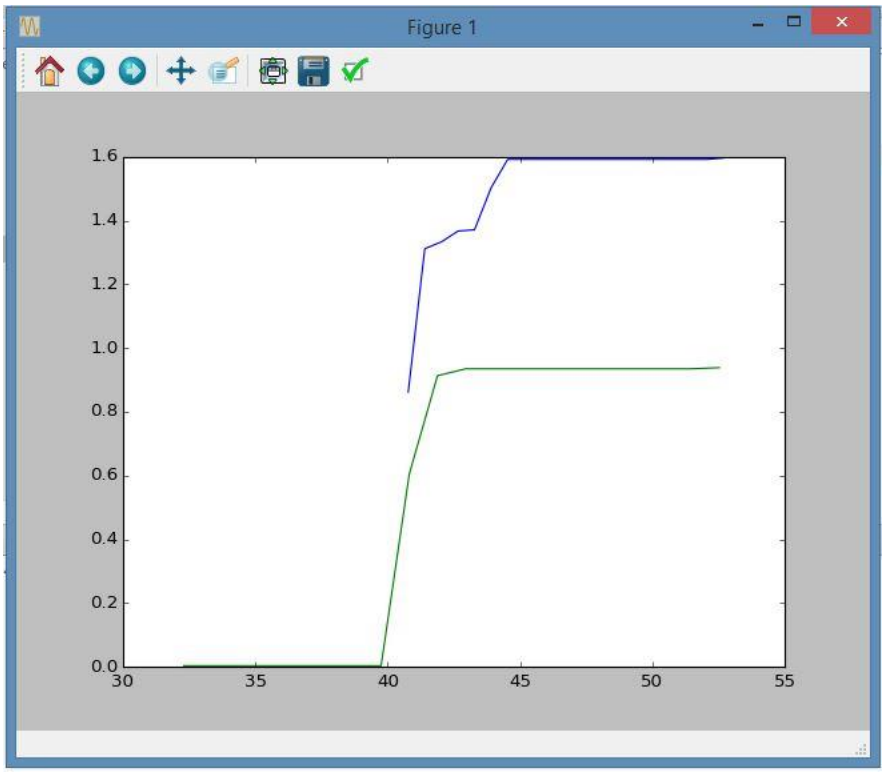
3) planetlab4.cs.uoregon.edu → pl1.6test.edu.cn



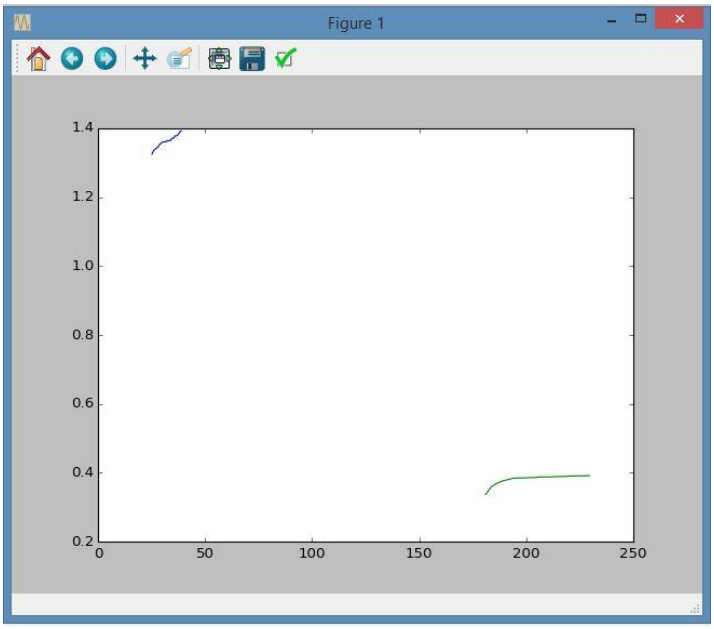
4) PLANETLAB1.CS.PURDUE.edu → PLONK.CS.UWATERLOO.CA



5) PLANETLAB2.UTDALLAS.EDU → planetlab-3.cmcl.cs.cmu.edu



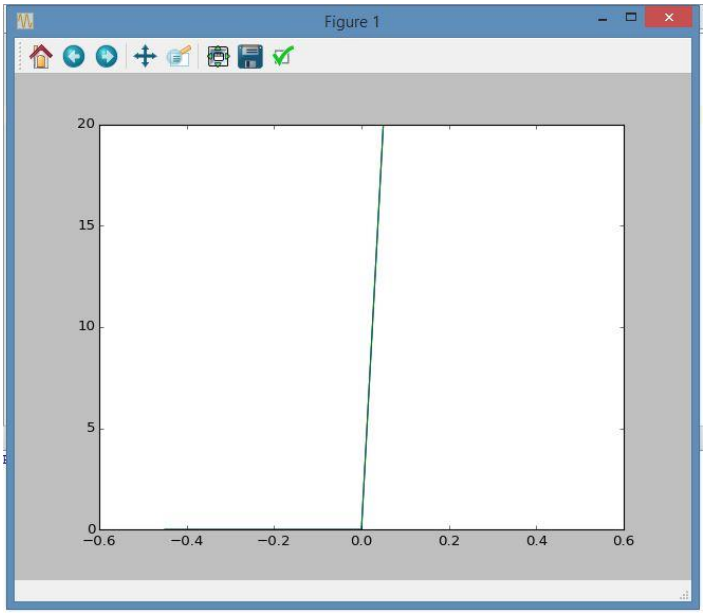
6) PLANETLAB1.CS.UCLA.EDU → planetlab4.cs.uoregon.edu



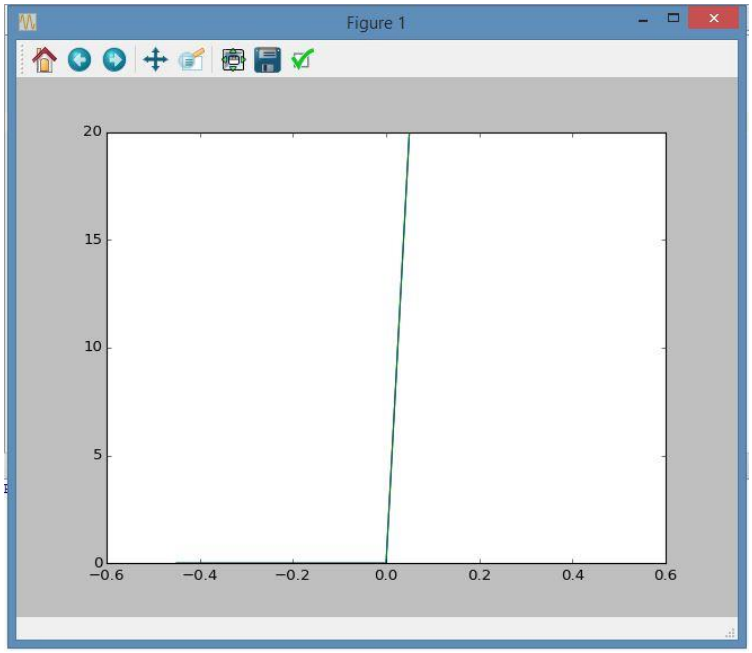
Packet Loss Graph ::

For generating packet loss each ping directory has one file names packetLoss.py which ultimately checks the percentage packet loss in the ping output and write it to packetLoss.txt and packetLoss.csv file using csvpacketLoss.py. From this csv file I have created the text file and make a python script to plot the graphs as shown below. The script and all text are there in the Graph folder.

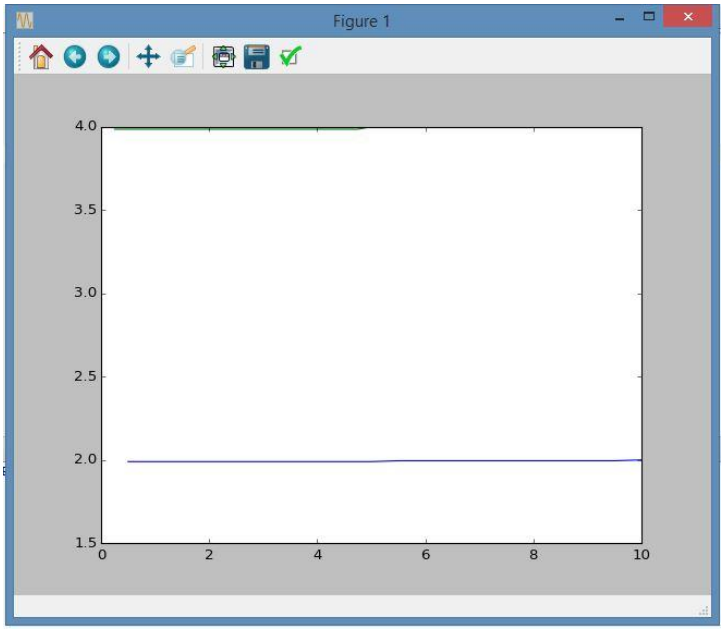
1) planetlab5.eecs.umich.edu-pl2 → prakinf.tu-ilmenau.de



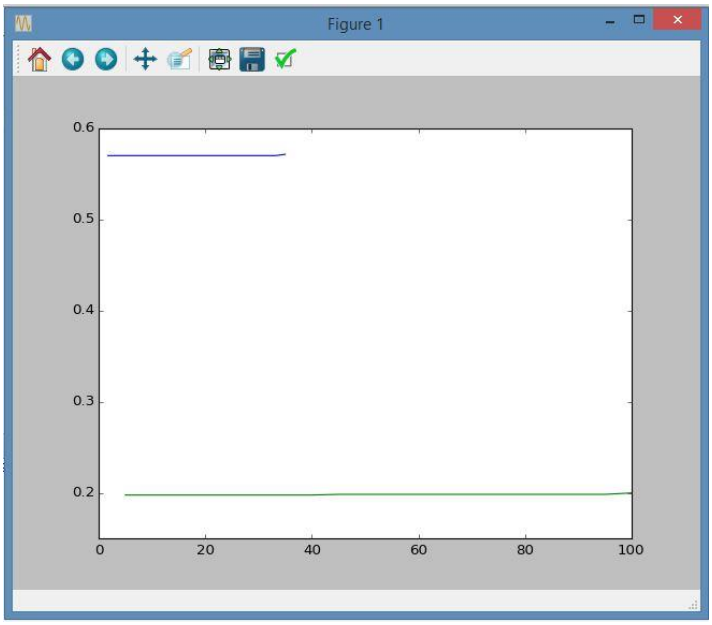
2) planetlab04.cs.washington.edu → PLANETLAB-N2.WAND.NET.NZ



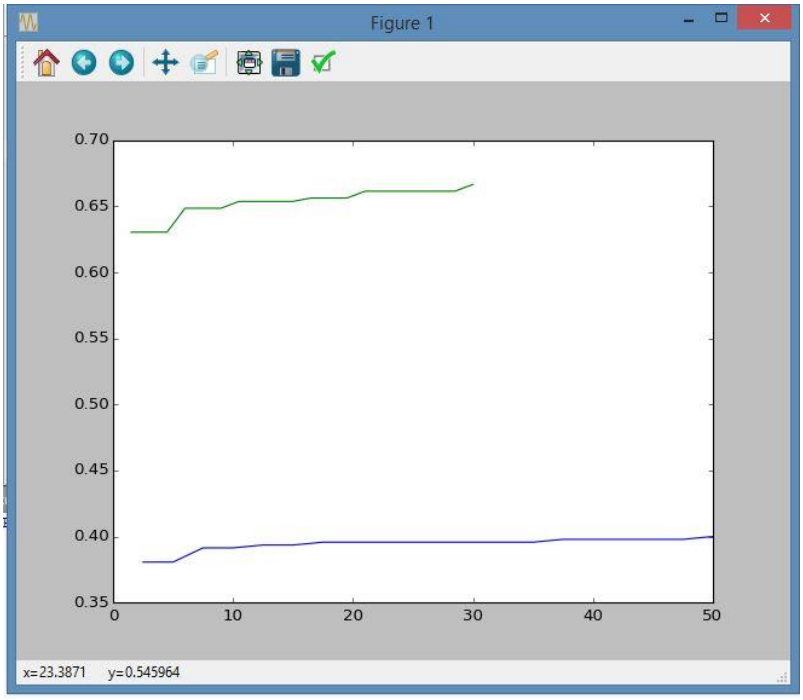
3) planetlab4.cs.uoregon.edu → pl1.6test.edu.cn



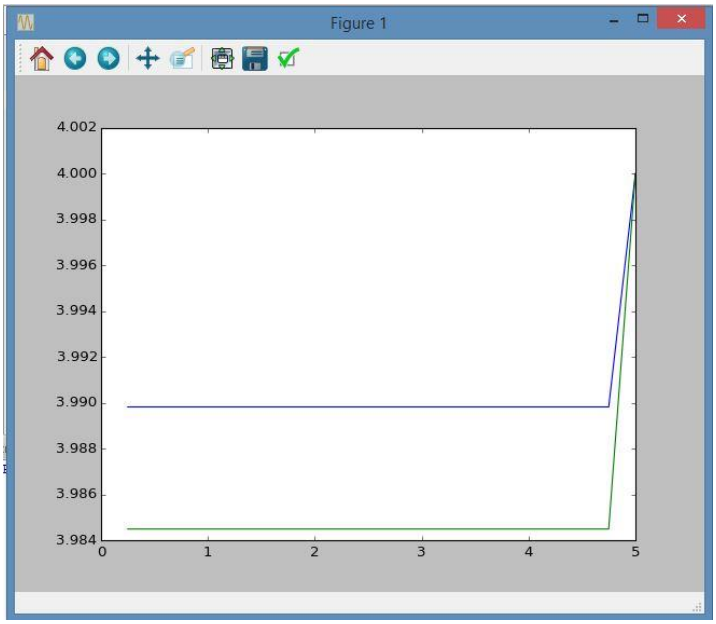
4) PLANETLAB1.CS.PURDUE.edu → PLONK.CS.UWATERLOO.CA



5) PLANETLAB2.UTDALLAS.EDU → planetlab-3.cmcl.cs.cmu.edu



6) PLANETLAB1.CS.UCLA.EDU → planetlab4.cs.uoregon.edu



Conclusion ::

According to my observation, the instability has been decreased over the years. There is less failure of packets and much shorter path to reach the destination than earlier.