Making the Most of Multihoming

Aditi Thanekar
College of Computing,
Georgia Tech.
aditi@cc.gatech.edu

Constantine Dovrolis
College of Computing,
Georgia Tech.
dovrolis@cc.gatech.edu

Warren Matthews
Office of Information Technology,
Georgia Tech.
warren.matthews@oit.gatech.edu

CONTENTS

1. Introduction

Given the increasing number of stub networks that resort to multihoming for reliability and availability, it is worth investigating whether multihoming is actually being utilized in a way that it provides the benefits it promises.

Multihoming is the connection of a stub/edge network to more than one Internet provider. It gives a network administrator the ability to make a choice - one that will help reap the maximum benefits from the network setup. This can only be ascertained by a conscientious analysis of the network's egress paths.

Most networks that focus on reliability and availability use multihoming as a form of redundancy. Educational institutions for example purchase services from multiple ISPs. In addition to commercial service providers, these networks are usually also connected to academic networks like NLR and Abilene.

This report is an effort to quantify the different attributes of the paths leaving a multi-homed educational network. We analyze paths between the participating universities: Georgia Tech, NCREN, Ohio and TAMU. We present results that characterize, correlate and compare these paths. These in turn may be used to provide guidelines for a network administrator to make an informed decision. Our testbed is restricted to these universities due to the use of destination based routing. In order to perform routing using destination based routing, we need static routes at the edge routers (Section 2) and hence active cooperation from the routers of all remote networks.

Section 2 describes the setup for destination based and source based routing. In section 3, we characterize the paths between two networks on the basis of capacity, jitter and available bandwidth. We compare these paths in Section 4 in terms of available bandwidth and jitter. Section 5 further correlates these paths.

2. Setup

In the initial setup, the gateway routers at the three universities at North Carolina, Ohio and Georgia Tech were setup for destination based routing: the upstream ISP is chosen depending on the destination address of the packet. Eventually, Georgia Tech's gateway router was updated to provide for filter based forwarding, and hence have the ability to perform source based routing, where the router chooses an upstream ISP based on the source address of a packet.

2.1 Destination based routing:

The *initial* setup includes 3 universities: North Carolina (NCREN), Ohio (OAR) and Georgia Tech (GT). At each university, there exists one machine with five interfaces, ie. having 5 virtual addresses. The routing tables at each router have five hard coded entries for every machine's addresses, specifying the upstream IP address through which the packets should be sent to the destination.

For example, the routing table entries at North Carolina contain the following five hard coded routes to Georgia Tech:

Destination at GT	Upstream ISP
143.215.252.50	NLR
143.215.252.90	Abilene
143.215.252.91	Qwest
143.215.252.92	L3
143.215.252.93	L3

Table 2.1 Hard coded routing table entries at NC for five addresses at GT

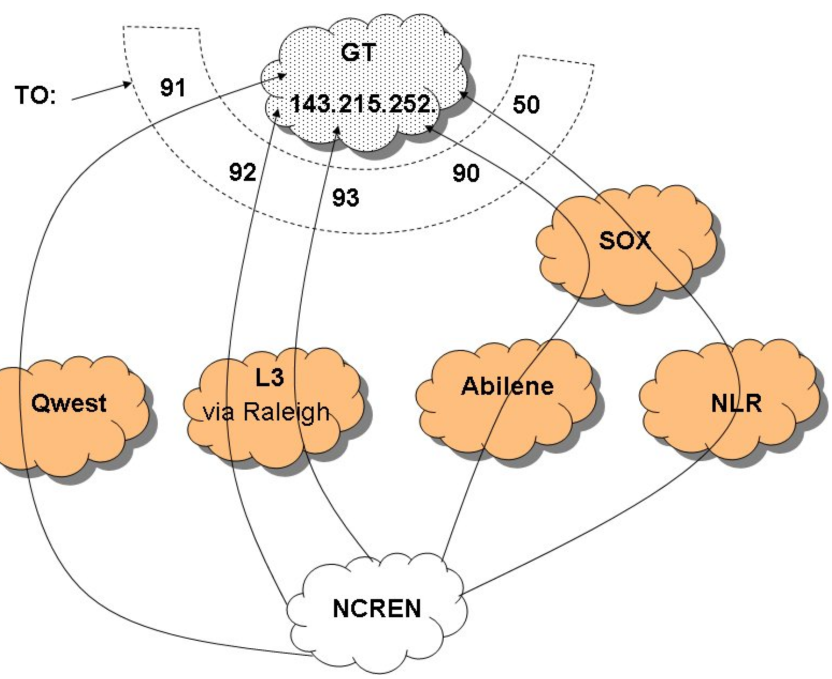


Fig. 2.1 Routes from North Carolina into Georgia Tech

Fig. 2.1 illustrates the routes taken by packets sourced at North Carolina destined to the five IP addresses at Georgia Tech. Traceroutes were run between each pair of nodes to each of the hard coded addresses to reveal the following routes:

Table 2.2 : GT -> NC

Upstream ISP	Address	Path
Abilene	128. 109.41.26	GT - SOX – Abilene - NC
Qwest	128. 109.41.27	GT – Qwest - NC
L3	128. 109.41.28	GT - L3 (thru charlotte) - NC
Cogent	128. 109.41.29	GT - Cogent - Qwest - NC
NLR	128. 109.41.30	GT - SOX - NLR - NC

Table 2.3 : GT -> OH

Upstream ISP	Address	Path
Abilene	192.148.235.70	GT - SOX – Abilene - OH
Qwest	192.148.235.71	GT – Qwest - OH
L3	192.148.235.72	GT - L3 (thru washington) - OH
Cogent	192.148.235.73	GT - Cogent - Qwest - OH
NLR	192.148.235.74	GT - SOX - NLR - OH

Table 2.4: Ohio -> NC

Upstream ISP	Address	Path
Abilene	128.109.41.26	OH – Abilene - NC
Qwest	128. 109.41.27	OH - Qwest - NC
Qwest	128. 109.41.28	OH - Qwest - NC
Qwest	128. 109.41.29	OH - Qwest - NC
Abilene	128. 109.41.30	OH – Abilene - NC

Table 2.5: Ohio -> GT

Upstream ISP	Address	Path
Abilene	143.215.252.50	OH – Abilene – SOX – GT
Qwest	143.215.252.90	OH – Abilene – SOX – GT
Qwest	143.215.252.91	OH - Qwest - GT
Qwest	143.215.252.92	OH - Qwest - GT
Abilene	143.215.252.93	OH - Qwest - GT

Table 2.6: NC -> Ohio

Upstream ISP	Address	Path
Abilene	192.148.235.70	GT - SOX – Abilene - OH
Qwest	192.148.235.71	GT – Qwest - OH
L3	192.148.235.72	GT - L3 (wash) – Qwest - OH
Abilene	192.148.235.73	GT – Abilene - OH
Abilene	192.148.235.74	GT – Abilene – OH

Table 2.7: NC -> GT

Upstream ISP	Address	Path
NLR	143.215.252.50	NC – NLR – SOX – GT
Abilene	143.215.252.90	NC – Abilene – SOX – GT
Qwest	143.215.252.91	NC – Qwest – GT
L3	143.215.252.92	NC – L3 (Raleigh) – GT
L3	143.215.252.93	NC – L3 (Raleigh) – GT

2.2. Source based Routing:

Initally, all three universities were set up for destination based routing. As a result, the test bed is only restricted to addresses that are hard coded into the routing tables. This situation can be significantly improved by introducing "source based routing", or filter based routing. IN this setup, filters at the gateway router route traffic on the basis of the source address of the packet.

These filters have been set up only at Georgia Tech, hence we now have NC and OH with destination based routing as shown above, and Georgia Tech with source based routing (as demonstrated in the Fig 2.2)

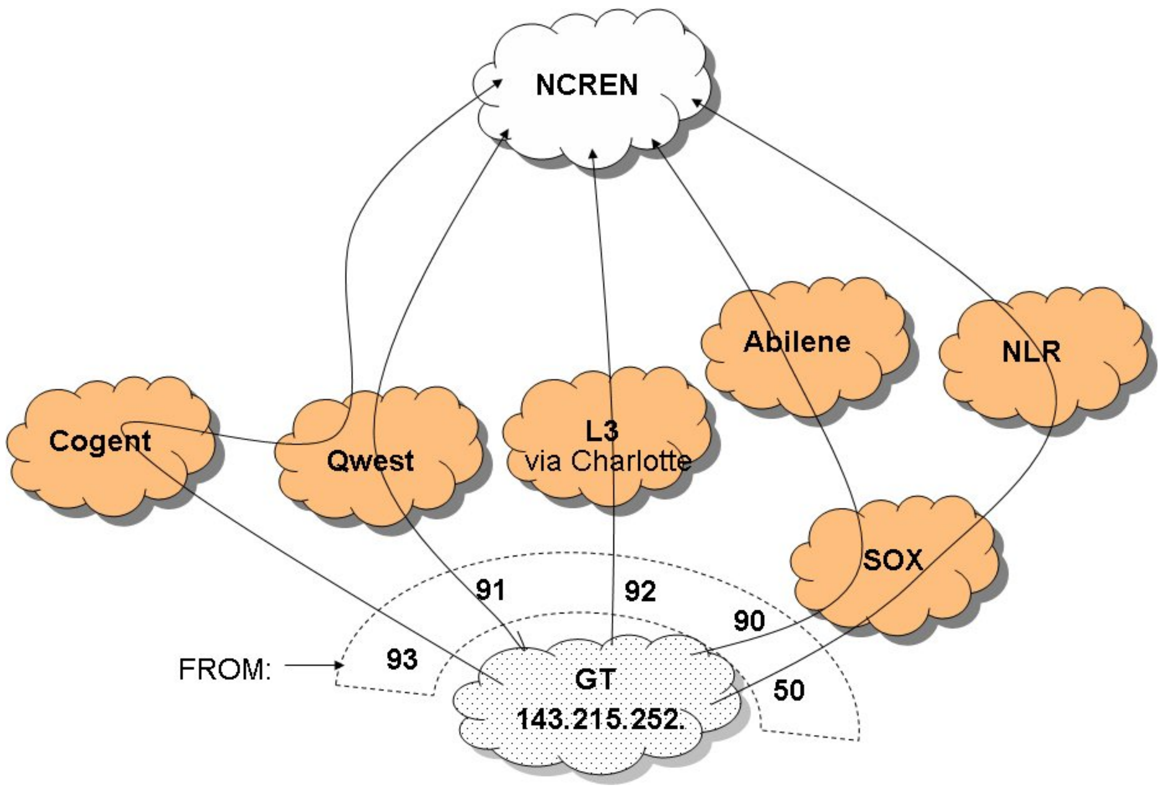


Fig 2.2: Routes from Georgia Tech into North Carolina using source based routing

It is worth noticing that the addresses corresponding to the upstream ISPs in destination based routing correspond to the same upstream ISPs in source based routing. For example, in table 2.1, we saw the upstream ISPs for all routes into GaTech using destination based routing. Table 2.8 shows the upstream ISPs for routes out of Georgia Tech using source based routing:

Source at GT	Upstream ISP
143.215.252.50	NLR
143.215.252.90	Abilene
143.215.252.91	Qwest
143.215.252.92	L3
143.215.252.93	Cogent

Table 2.8 Upstream ISPs determined by source address at Georgia Tech

A significant consequence of this fact is that we now necessarily see symmetric paths. For example, if a TCP connection is setup with source address = 143.215.252.91 and destination address as one of the address at North Carolina, then:
- By source based routing at GT: The data packets are sent through upstream ISP Qwest
- By destination based routing art NC: The acknowledgment packets are sent through upstream ISP Qwest.

3. Characterizing the path

3.1 Capacity Estimates:

Pathrate was run on all paths between NC, GT and OH. Pathrate measurements were done twice a day on weekdays. The output of Pathrate is a range: we calculate the mean of all lower bound and upper bound readings. The following tables show the mean taken over all 10 pathrate readings.

Table 3.1: GT -> NC

Upstream ISP	Estimated Capacity (Mbps)
Abilene	~1000
Qwest	400-441
L3	92 - 99
Cogent	345 -381
NLR	~1000

Table 3.2: NC-> GT

Upstream ISP	Estimated Capacity (Mbps)
Abilene	~1000
Qwest	472 - 481
L3	93 – 98
NLR	~1000

The capacities of all paths in and out of Ohio were observed to be in the range of approximately 98 Mbps, suggesting all these paths share a common bottleneck, which appears to be a Fast Ethernet link.

3.2 Available Bandwidth:

Pathload was run between each pair of machines at NC, GT and OH to determine the available bandwidth along these paths. The available bandwidth from NC to GT across the five upstream ISPs is as shown.

Table 3.3: GT -> NC

Upstream ISP	Estimated Capacity (Mbps)	Available Bandwidth (Mbps)
Abilene	~1000	760.23 – 825.46
Qwest	400-441	340.21 – 382.12
L3	92 - 99	20.73 – 59.30
Cogent	345 -381	282.67 – 315.79
NLR	~1000	680.32 – 734.53

Pathload measurements from NC to GT were unsuccessful due to the problem of interrupt coalescence. Also, available bandwidth on the four paths: OH → GT, GT → OH, OH → NC, NC → OH was consistently observed to be 98 Mbps.

Due to the diversity in readings for the paths from GT to NC, From this point, we restrict out observations to the paths from Georgia Tech to North Carolina through the 5 upstream ISPs: Qwest, L3, Cogent, Abilene and NLR.

Figures 3.3 and 3.4 show the distribution of available bandwidth measurements across the three commercial ISPs: Qwest, Cogent and L3.

Figures 3.5 and 3.6 show the distribution of available bandwidth measurements across the commercial ISPs in comparison to the two academic networks: Abilene and NLR.

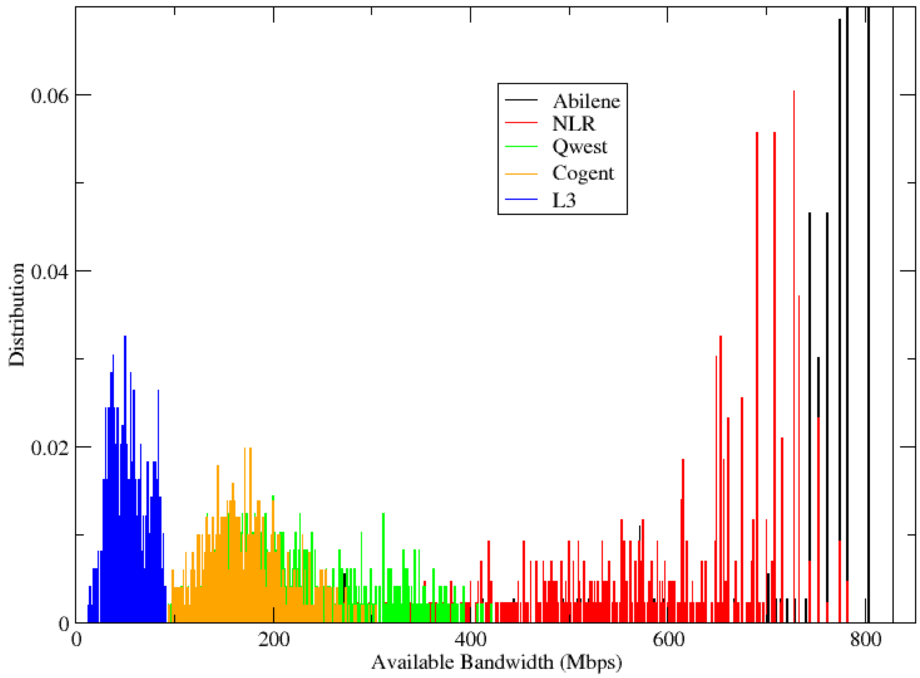


Fig 3.3: Distribution of available bandwidth through Abilene, NLR and commercial ISP's from GT to NC.

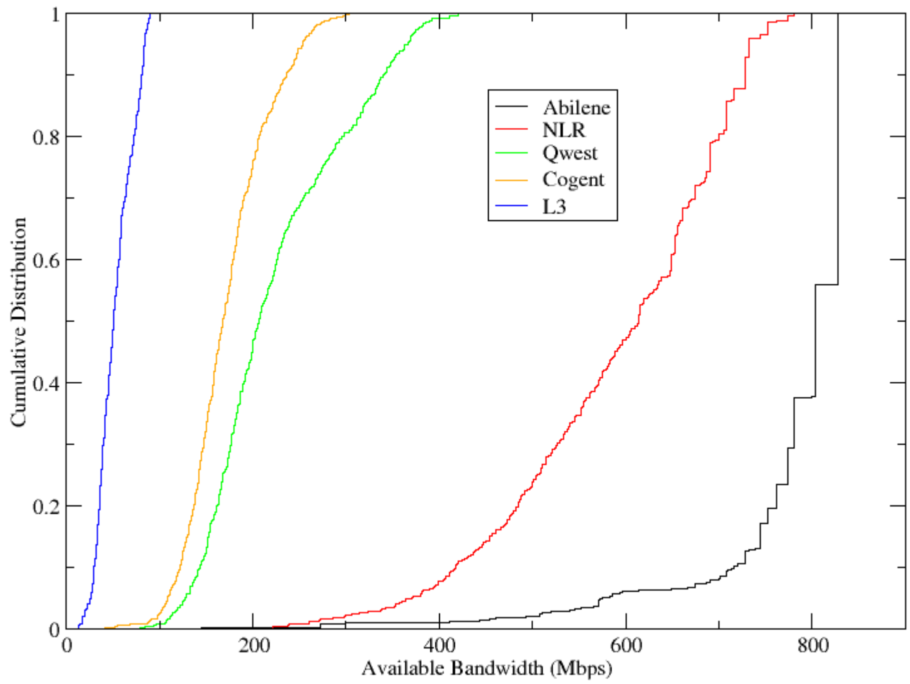


Fig 3.4: CDF of available bandwidth through Abilene, NLR and commercial ISP's from GT to NC.

Pathload returns an lower and a upper limit for the available bandwidth. Figures 3.6 and 3.7 show that the CDFs of the lower and upper limits of the available bandwidth are consistent with the mean available bandwidth readings.

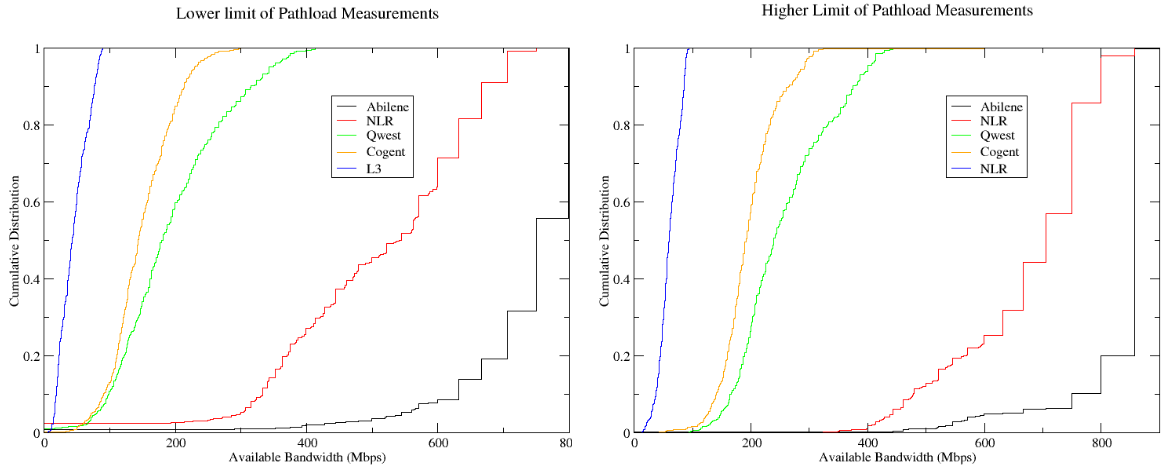


Fig 3.5 & 3.6: CDF of available bandwidth through Abilene, NLR and commercial ISP's from GT to NC.

With the upper and lower estimates of the available bandwidth, we can also determine the variance in the available bandwidth across each of the upstream ISPs:

$$\text{Relative Variation Range} = \frac{(\text{UpperLimit} - \text{LowerLimit})}{\text{mean}}$$

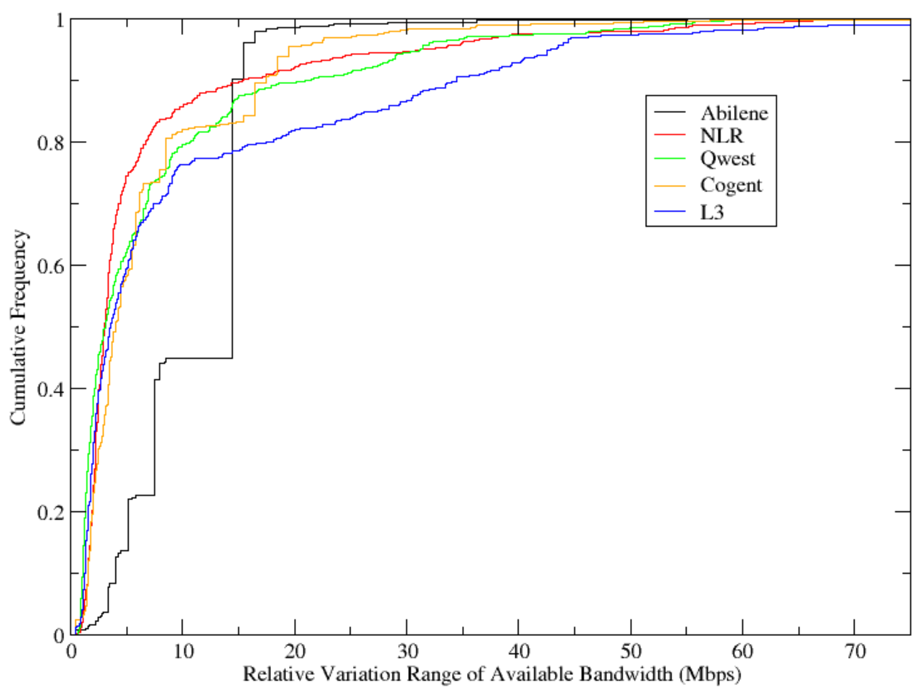


Fig 3.7: CDF of the relative variation range of available bandwidth through Abilene, NLR and commercial ISP's from GT to NC.

3.3 Jitter

Jitter is the variation in the time between packets arriving, caused by network congestion, timing drift, or route changes, among other reasons. For a single jitter measurement, we collect successive ping measurements taken every 100 ms for 30 seconds, and use these RTTs to calculate jitter.

Inter packet delay variance is determined using the formula:

$$\frac{\sum |d_{i-1} - d_i|}{k-1}$$

This value was normalized by division by mean. The plots resulting from these calculations are as shown:

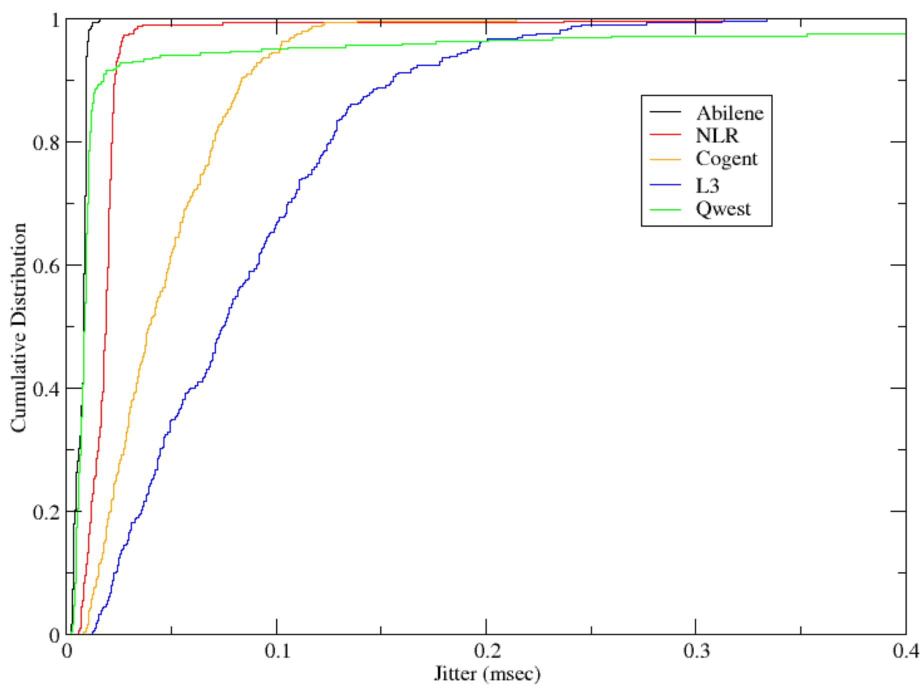


Fig 3.8: CDF of Jitter values across the 5 upstream ISPs from GT to NC

It is observed that Level3, with the least available bandwidth shows maximum jitter. Cogent also appears to reflect the same ranking as its available bandwidth. However, as we head towards paths with higher available bandwidth , the jitter tends to 0, and it difficult to rank these paths.

We can thus conclude that
- Low available bandwidth means occasional queuing delays and jitter.
- But if available bandwidth is very high, the queuing delays and jitter are practically zero most of the time.

3.4 TCP Throughput

Iperf[?] is used to calculate the TCP throughput across each path. Iperf accepts as a command line argument, the size of the window at both the sender and receiver. This window is as large as possible (in our case, 200 MB) in order to determine the maximum throughput, without letting the buffer limit the flow.

In the initial stages, iperf did not perform as expected, and therefore, we develop a home grown application that will try to saturate the path, and then measure the throughput achieved.

TCP streams were set up from Georgia Tech to NC, and the rate was calculated at the receiver end, by aggregating the number of packets received every 100 ms. The specifications for the sender and the receiver are as follows:

Sender:
The packet size and buffer settings at the server are as follows:
Maximum application block size : 16KB
Socket Buffer size : 200 MB.
The time for which this stream will be sent (in ms) is accepted as a command line argument

Receiver:
The socket buffer size at the client too is set up at 200MB.
The client is responsible for calculating the rate at which it receives the packets. It does this by counting the number of bytes that are received at the client every 100 ms

The increased buffer sizes ensure that flow is not limited by buffer size. Increased application block sizes ensure that path is saturated.

It is ensured that the buffer sizes at both, the sender and the receiver are actually are set to 200 MB by the OS by setting the following Unix variables at GT and NC:

```
net.core.rmem_max = 300 MB
net.core.wmem_max = 300 MB
```

This application was run across the paths through the five upstream ISPs. We also have with us mean available bandwidth measurements taken for the same time of the day as the time of these runs. Figures 3.9 to 3.13 represent the timeline of the rates calculated every 100 ms over a period of 60 seconds. The magenta line in each of the figures shows the average throughput over 60 seconds.

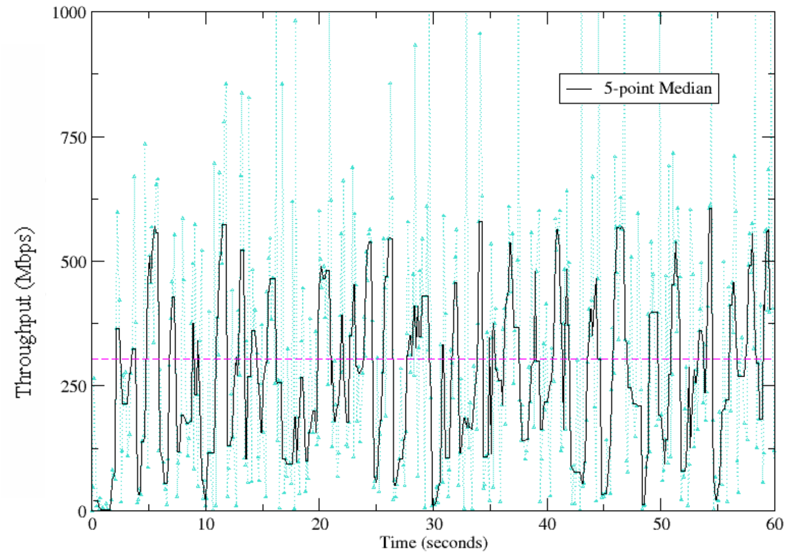
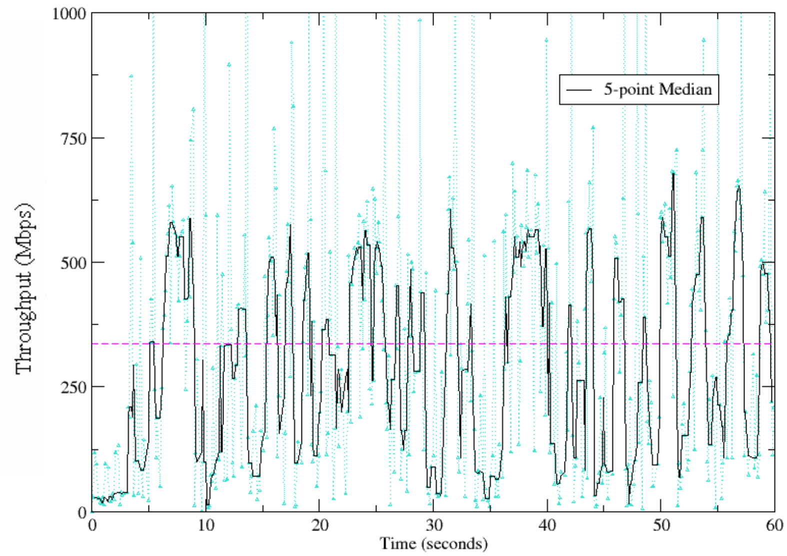
Fig 3.9: Rates from GT to NC through Abilene	
Fig 3.10: Rates from GT to NC through NLR	

Fig 3.11: Rates from GT to NC through Qwest

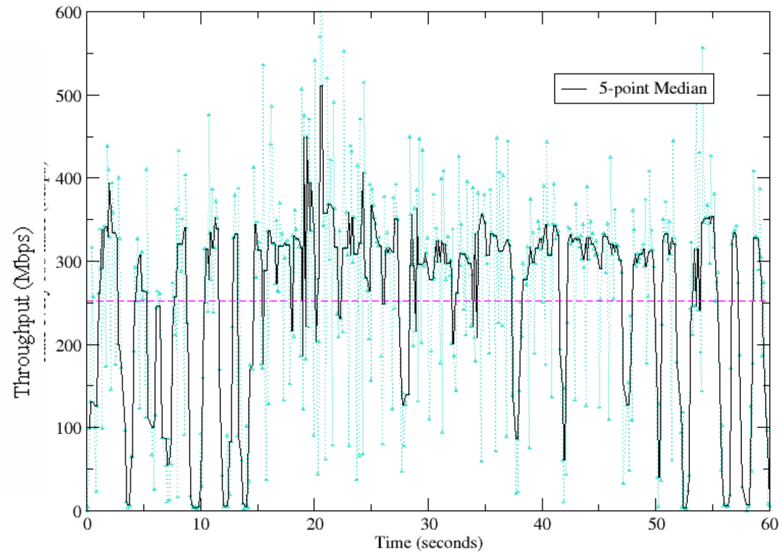


Fig 3.12: Rates from GT to NC through Cogent

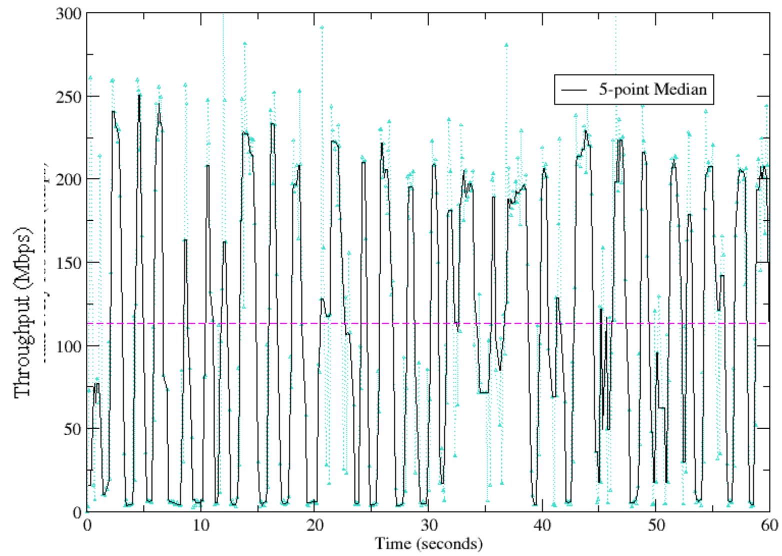
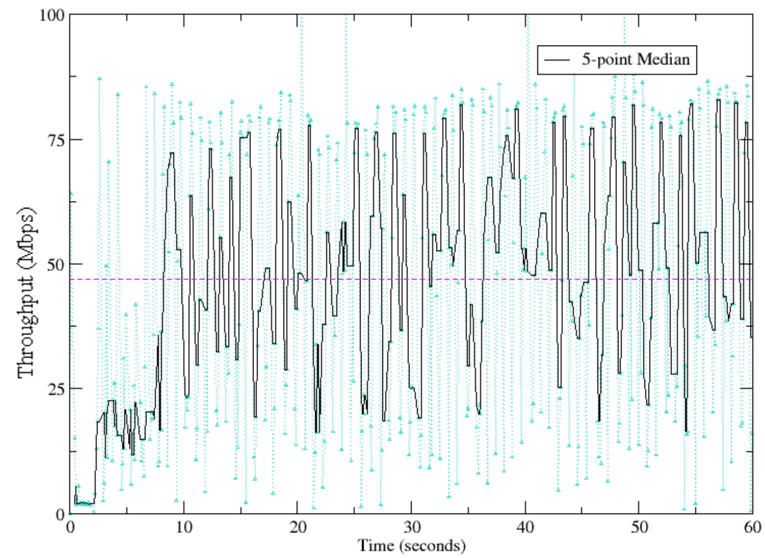


Fig 3.13: Rates from GT to NC through L3



The observations from these figures can be summarized in Table 3.4:

Table 3.4: ++

	Abilene	NLR	Qwest	Cogent	L3
Average Available Bandwidth	621 - 634	589 - 642	322 - 341	253 - 367	82 - 93
Buffer size* / RTT	8230	18518	8438	8928	12821
Average TCP Throughput	300	320	250	120	48
Maximum TCP Throughput	600 ± 50	550 ± 100	320 ± 25	250 ± 50	75 ± 10

*Buffer Size = min(Sender buffer, Receiver Buffer)
++All values are given in Mbps

The table shows that the five flows are never limited by buffer size, since the values for Buffer size / RTT are very high. It is seen the maximum TCP throughput values are approximately equal to the available bandwidth. The only limitation now is the available bandwidth. According to TCP Congestion Control, as soon as the link gets congested, ie throughput > Available bandwidth, the TCP window is cut down, resulting in reduced throughput. Now, the window begins to increase again, only to be cut down when the link gets congested once more. This constant variation results in an average throughput that is lower that the available bandwidth.

In order to achieve average throughput ~ available bandwidth, buffer sizing is recommended. The buffer should be set to an optimal value such that the connection can utilize the complete available bandwidth, yet maintain constant throughput by not causing losses that would indicate congestion in the link.

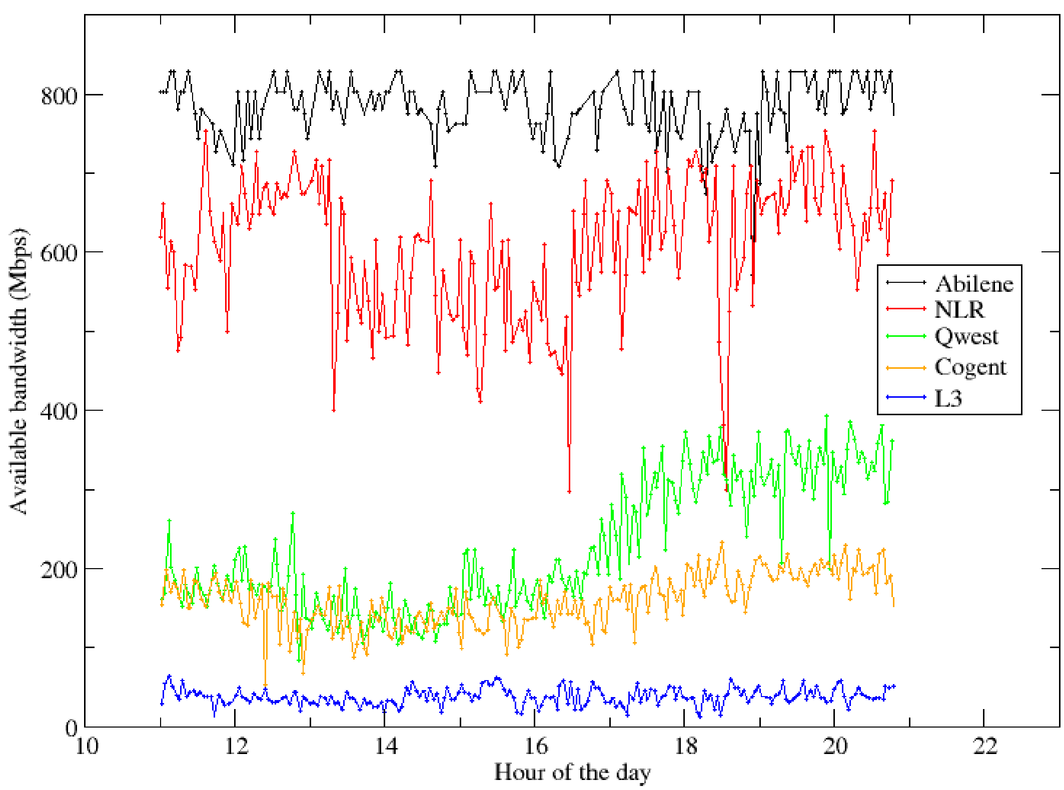
4. Correlating the paths

Although connections are set up through different upstream ISPs, two or more paths may share at least one link. In the worst case, one of these links is a bottleneck. Internet Routing Control involves switching to the 'best path'. When this switch is made, it would be advisable to avoid a path that shares a bottleneck with the current path. We therefore try to make out if the available bandwidth across two paths demonstrates similar variation in available bandwidth, thus indicating that they share a common link.

4.1 Method of Data Collection:

Pathload measurements were interleaved between the three ISPs, Qwest L3 and Cogent; and successive measurements were taken for three hours.

The following graph shows the average bandwidth measurements for 10 hours from 11:00 am to 21:00pm on Nov 11, 2006.



4.2 Correlation Coefficient

If two paths share the same bottleneck, they will appear to be “correlated” in the the above graph. By observing the correlation coefficient, it would be possible to determine if two paths are correlated. The formula for the correlation coefficient is given by:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{N}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{N}\right)}}$$

The value of r was determined for each pair among the five ISPs between GT and NC:

	Day 1	Day 2	Day 3
Abilene – NLR	0.489	0.081	0.008
Abilene – Cogent	0.318	-0.091	0.249
Abilene – Qwest	0.221	-0.066	0.105
Abilene – L3	0.359	0.048	0.018
NLR – Cogent	0.505	0.265	0.127
NLR – Qwest	0.314	0.309	0.146
NLR – L3	0.396	-0.03	0.093
Cogent – Qwest	0.592	0.649	0.794
Cogent – L3	0.256	0.067	0.175
Qwest – L3	0.196	0.124	0.675

It is seen that the correlation coefficient between Cogent and Qwest is significantly and consistently high, as compared to the other pairs. This leads us to believe that Cogent and Qwest share one or more common links. On observing the physical setup, it is confirmed that the two ISPs do share a significant chunk of their paths. Traceroutes show that traffic is handed over from Cogent to Qwest before it enters the NC network. It is seen that the two paths share five hops in common once they leave the Georgia Tech network.

We also see NLR-Cogent with a correlation coefficient of 0.5 on one day, but readings on other days are not consistent with this value.

5. Comparing Paths

5.1 Method of data collection:

- Scripts were set up to take available bandwidth and jitter measurements from North Carolina to Georgia Tech through the 3 ISPs: Qwest, Cogent and NLR.

  These measurements were taken from 9:00 am to 5:00 pm on weekdays, and a sequence was set up as follows:

  1. Pathload through Abilene
  2. Ping through Abilene
  3. Pathload through NLR
  4. Ping through NLR
  5. Pathload through Qwest
  6. Ping through Qwest
  7. Pathload through Cogent
  8. Ping through Cogent
  9. Pathload through L3
  10. Ping through L3

  The objective of taking these measurements was to determine if conclusions can be drawn about the correlation between the available bandwidth and jitter measurements.

- For throughput, however, less number of measurements were taken since we saturate the path in order to determine the throughput. In this case, measurements were taken every hour, interleaved among ISPs in the same way as given above, but with iperf measurements instead of pings.

5.2 Available bandwith and jitter:

Pathload is an intrusive technique to determine the available bandwidth of the path. We therefore make attempts to determine if we can find an alternate metric that would give an indication of the available bandwidth on a path. Jitter calculation is far less intrusive, and we determine correlation coefficients between corresponding available bandwidth and jitter readings. As seen in Table 5.1, these readings show that the available bandwidth and jitter are uncorrelated.

	Day 1	Day 2
Abilene	0.035	0.007
NLR	-0.022	-0.077
Qwest	-0.191	0.124
Cogent	-0.383	-0.504
L3	0.122	-0.79

Table 5.1: Correlation between Available Bandwidth and corresponding Jitter values

5.2.2. Ranking:

Since each individual reading of available bandwidth and jitter does not show any correlation, we rank the paths by both available bandwidth and jitter, and observe if these ranks are consistent.

Method to determine ranking:

For every 10 minutes, we determine the ***median*** of the available bandwidth and jitter measurements. Results of this median calculation are shown in figures 5.1 and 5.2. The two graphs have the indx number as the X axis, since for the purpose of calculating the correlation coefficient, we use index numbers to associate available bandwidth measurements with the corresponding jitter measurement.

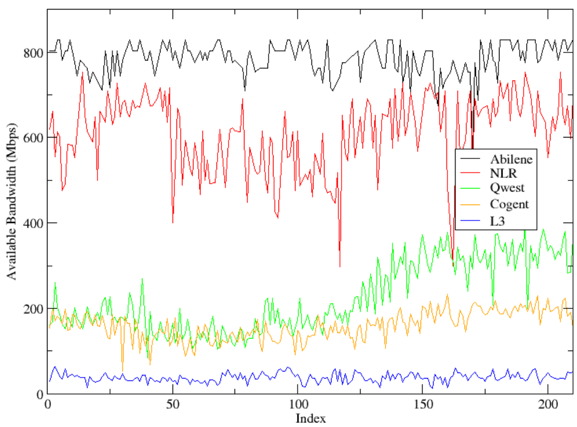


Fig 5.1a: Available bandwidth readings by Index

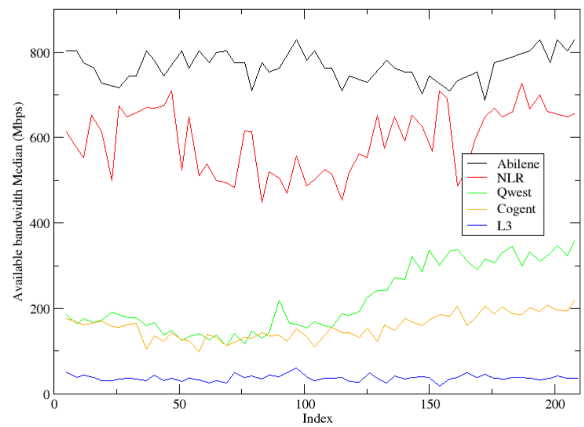


Fig 5.1b: Available bandwidth median over 10 min: displayed by Index

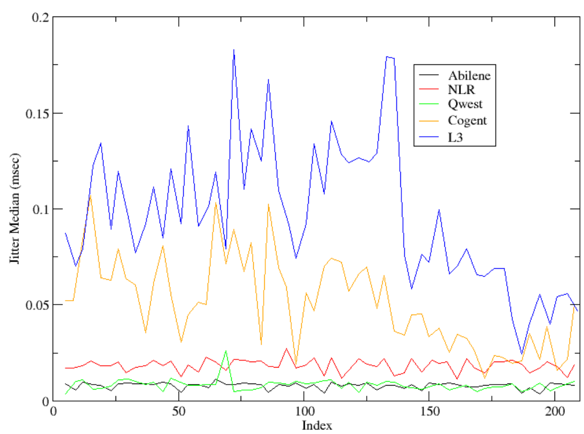


Fig 5.2: Jitter median over 10 minutes.

The readings for rankings can be summarized in Tables 5.3 and 5.4. The path with the higher available bandwidth is ranked higher, whereas the path with the lower jitter is ranked higher.

Table 5.2: Ranking for available bandwidth:

Rank \ ISP	Abilene	NLR	Qwest	Cogent	L3
First	89	1.09	7.69	2.18	0
Second	1.09	86.8	4.39	5.49	0
Third	0	0	82.42	7.60	9.89
Fourth	9.89	0	5.49	82.42	2.18
Fifth	0	12.08	0	0	87.92

Values represent percentage of times a particular ISP holds a particular rank

Table 5.3: Ranking for Jitter:

Rank \ ISP	Abilene	NLR	Qwest	Cogent	L3
First	49.43	0	34.83	0	0
Second	33.71	4.49	39.32	5.61	1.12
Third	10.11	39.32	20.22	8.99	6.74
Fourth	6.74	39.33	2.25	28.09	8.99
Fifth	0	16.85	0	48.31	11+7

Values represent percentage of times a particular ISP holds a particular rank

It is observed that the ranking of the paths for jitter does not agree with the ranking for available bandwidth. For example, in terms of Available bandwidth, Qwest is second for 86.8% of the times, but in terms of Jitter, it is second for only 4.49% of the time.

Thus, the ranking for jitter gives no indication of the ranking for available bandwidth. Therefore, using this approach, the jitter measurements cannot be used to substitute the more intrusive available bandwidth measurements.

5.3 Available bandwidth and Throughput:

These measurements were taken every hour from 9:00 am to 5:00 pm on weekdays, and a sequence was set up as follows:

1. Pathload through Abilene
2. iperf through Abilene
3. Pathload through NLR
4. iperf through NLR
5. Pathload through Qwest
6. iperf through Qwest
7. Pathload through Cogent
8. iperf through Cogent
9. Pathload through L3
10. iperf through L3

Fig 5.3 shows the CDF of the throughput values. These measurements are spaced at one hour, and readings have been taken for three days, giving us 24 readings. As a result these values may not be conclusive, warrant a larger number of measurements taken more frequently. It is seen that 50% of the throughput readings are consistent with the available bandwidth readings, that is, the throughput seen in L3 is the least and the throughput for Abilene is highest.

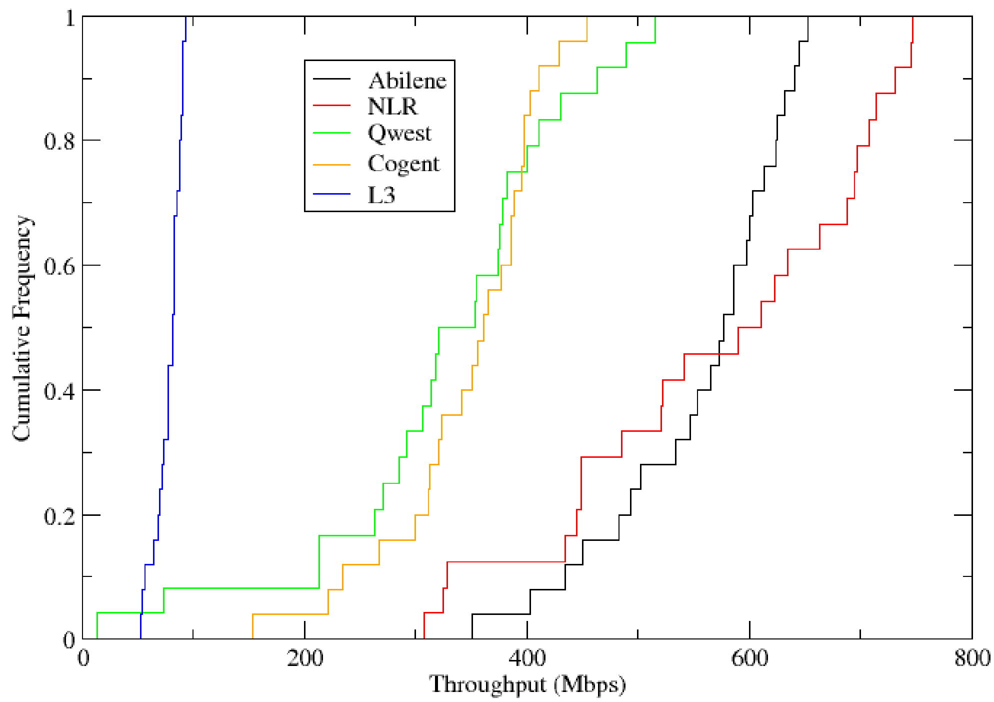


Fig. 5.3: Throughput measurements from GT to NC through 5 upstream ISPs

6. References

[1] Gao R., Zegura E.W., Dovrolis C., “Avoiding Oscillations due to Intelligent Route Control Systems” *In the Proceedings of IEEE INFOCOM. Barcelona, Spain, April 2006.*

[2] OWAMP: http://e2epi.internet2.edu/owamp/

[3] Akella A., Maggs B., Seshan S., Shaikh A., Sitaraman R. “A Measurement-Based Analysis of Multihoming” *In the proceesings of ACM SIGCOMM 2003, Karlsruhe, Germany.*

[4] Akella A., Seshan S., Shaikh A. “Multihoming Performance Benefits: An Experimental Evaluation of Practical Enterprise Strategies” *In USENIX Annual Technical Conference 2004, Boston, MA.*

[5] Dovrolis D., Ramathan P., Moore D., “Packet Dispersion Techniques and Capacity Estimation”, *In IEEE/ACM Transactions on Networking, December 2004.*

[6] PATHRATE: http://www.cc.gatech.edu/fac/Constantinos.Dovrolis/pathrate.html

[7] Jain M., Dovrolis D., “Pathload: A Measurement Tool for End-to-end Available Bandwidth” *In Proceedings of the 3rd Passive and Active Measurements Workshop, March 2002, Fort Collins CO.*

[8] PATHLOAD: http://www.cc.gatech.edu/fac/Constantinos.Dovrolis/pathload.html

[9] IPERF: http://dast.nlanr.net/Projects/Iperf/