School of Engineering and Computing Sciences

Geo-location in Large Networks with Congestion

REU fellow: Nadine Moukdad[1], Faculty mentor: Anand Santhanakrishnan[2]

Affiliation: 1. Fordham University 2. School of Engineering and Computing Sciences, NYIT

NYIT Research Experience for Undergraduates (REU)

Contact: nmoukdad@nyit.edu, asanthan@nyit.edu

May 31st – August 5th, 2016

Abstract

Geo-location of computers is important to ensure safe communication across networks. However, geo-location by IP address and delay-based geo-location, which only consider propagation delay, are neither reliable nor accurate[1]. We develop a real-time algorithm that improves the accuracy of geo-location of computers.
This research proposes a renewal theory based algorithm to geo-location. Our approach takes into account the delay suffered by packets due to network congestion in addition to propagation delay. *This is the first research on geo-location that takes into account network congestion.*

Data Collection

Planet-Lab[2] platform [Figure A] used to:

- Send probing ping packets from source nodes to destination nodes whose location is known *a priori*
- Extract round trip time (RTT) measurements from probing ping packets [Figure B]

Data Set:

- RTTs of ping packets sent continuously for 3 days between source nodes and destinations at different locations
- Locations: Canada, Princeton University, Rutgers University, Temple University, University of Washington, Colgate University, Australia, Hiroshima, New Zealand, Poland, and Spain

Methodology

- Measure the RTT between nodes by sending probing ping packets
- Extract statistical properties of the RTT
- Use statistical properties to develop algorithms based on renewal theory that correlate distance and RTT
    - Preliminary results revealed that the RTTs are exponentially distributed. Therefore, it is possible to fit the RTTs to an exponential PDF [Figure B]. The geo-location algorithm developed in this research uses this property.

Acknowledgement

The project is funded by National Science Foundation Grant number CNS-1559652 and New York Institute of Technology.

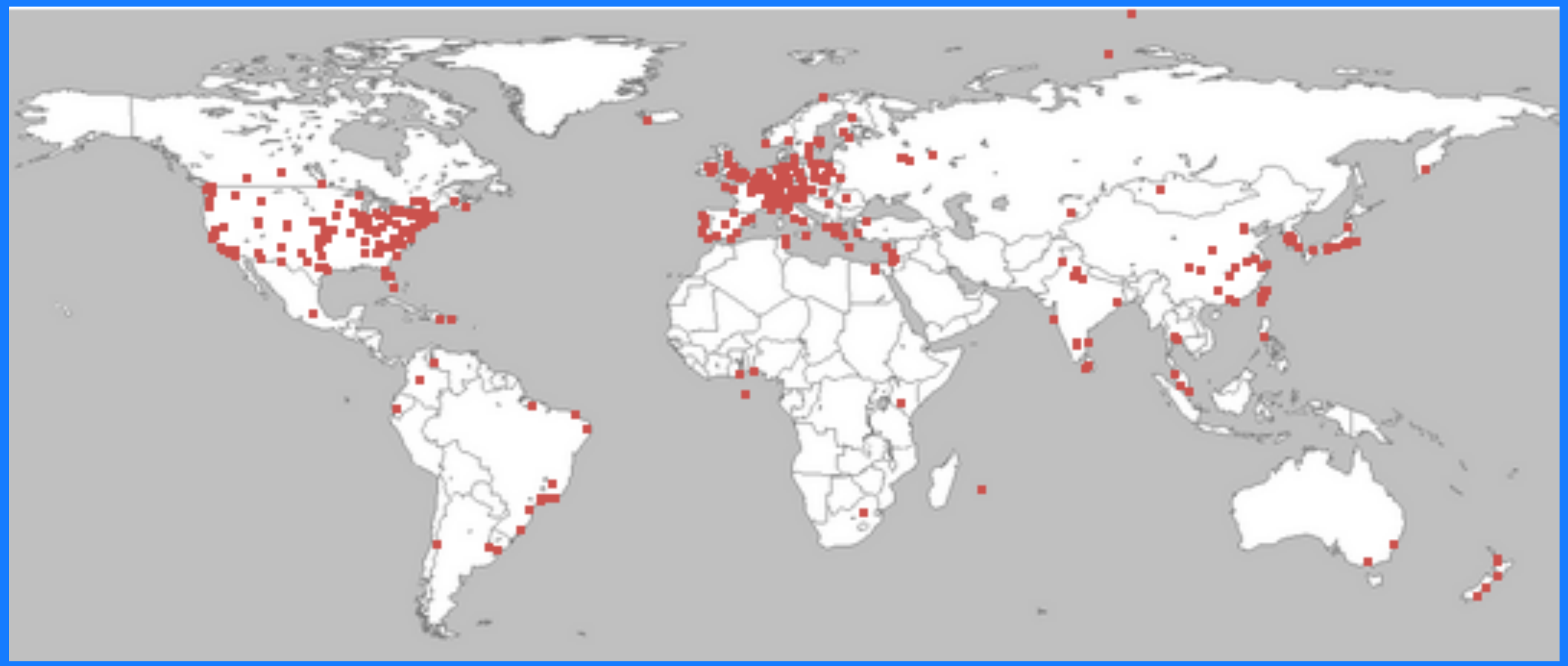

Figure A. Planet-Lab Node Distribution: 1353 nodes at 717 sites
Planet-Lab is a research platform used as a computer networking test-bed.

```
PING planetlab4.mini.pw.edu.pl (194.29.178.14) 56(84) bytes of data.
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=1 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=2 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=3 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=4 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=5 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=6 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=7 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=8 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=9 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=10 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=11 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=12 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=13 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=14 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=15 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=16 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=17 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=18 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=19 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=20 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=21 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=22 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=23 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=24 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=25 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=26 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=27 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=28 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=29 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=30 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=31 ttl=50 time=344 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=32 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=33 ttl=50 time=343 ms
64 bytes from planetlab4.mini.pw.edu.pl (194.29.178.14): icmp_seq=34 ttl=50 time=343 ms
```

Figure B. Raw data generated by sending probing ping packets. The last column represents the RTT in milliseconds from Australia to Poland.

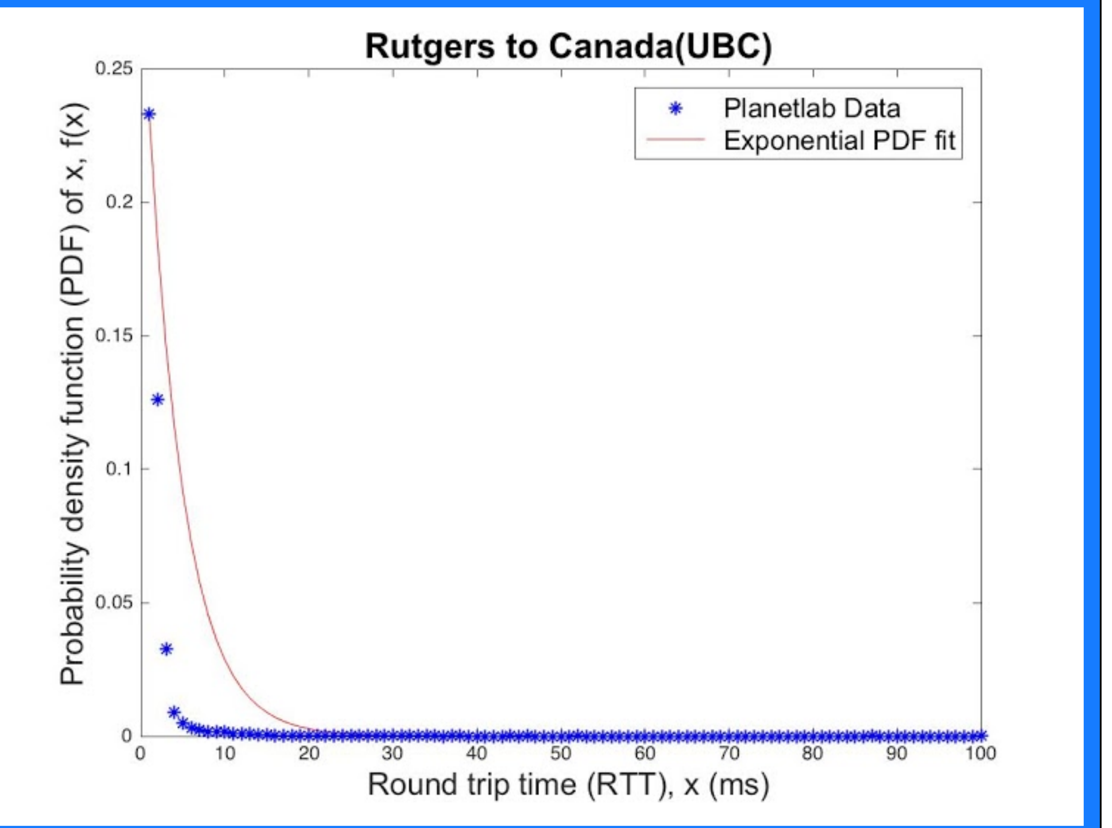

Figure C. An exponential probability density function for the RTTs from Rutgers University to the University of British Columbia.

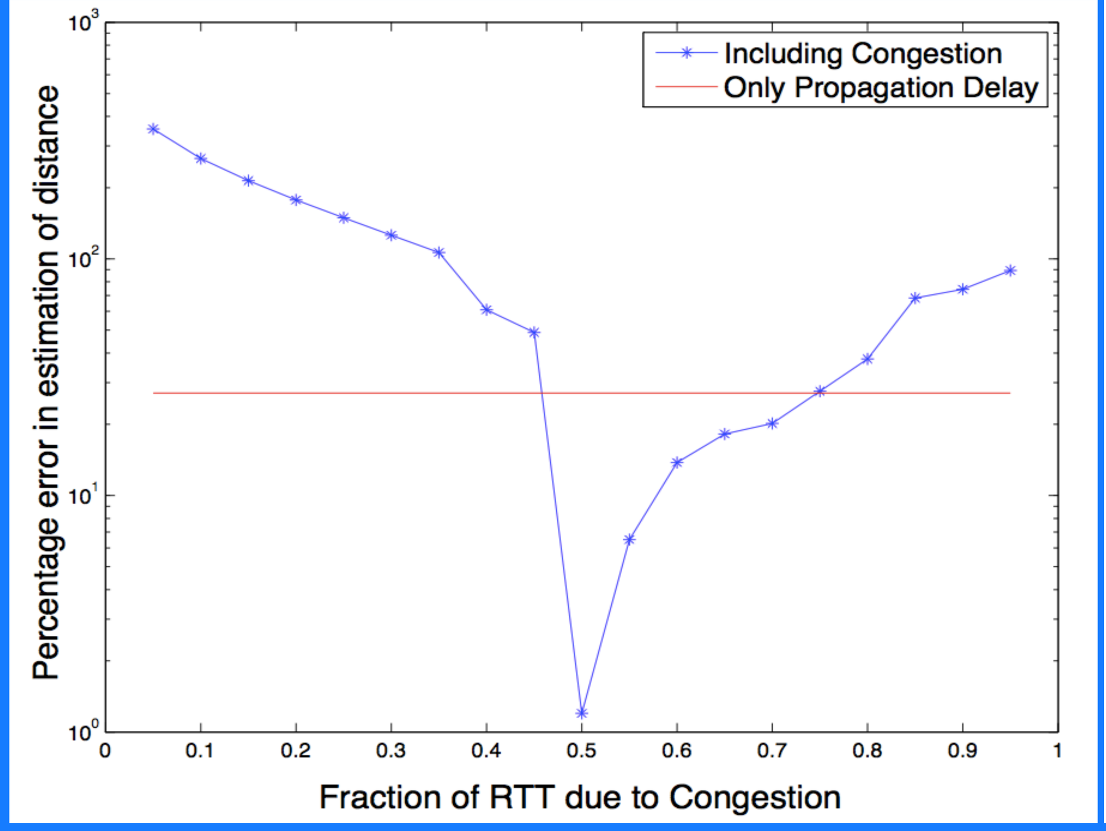

Figure D. Estimation error for estimated distances between Monash University in Australia to University of British Columbia in Canada.
Distance = 7,990 miles

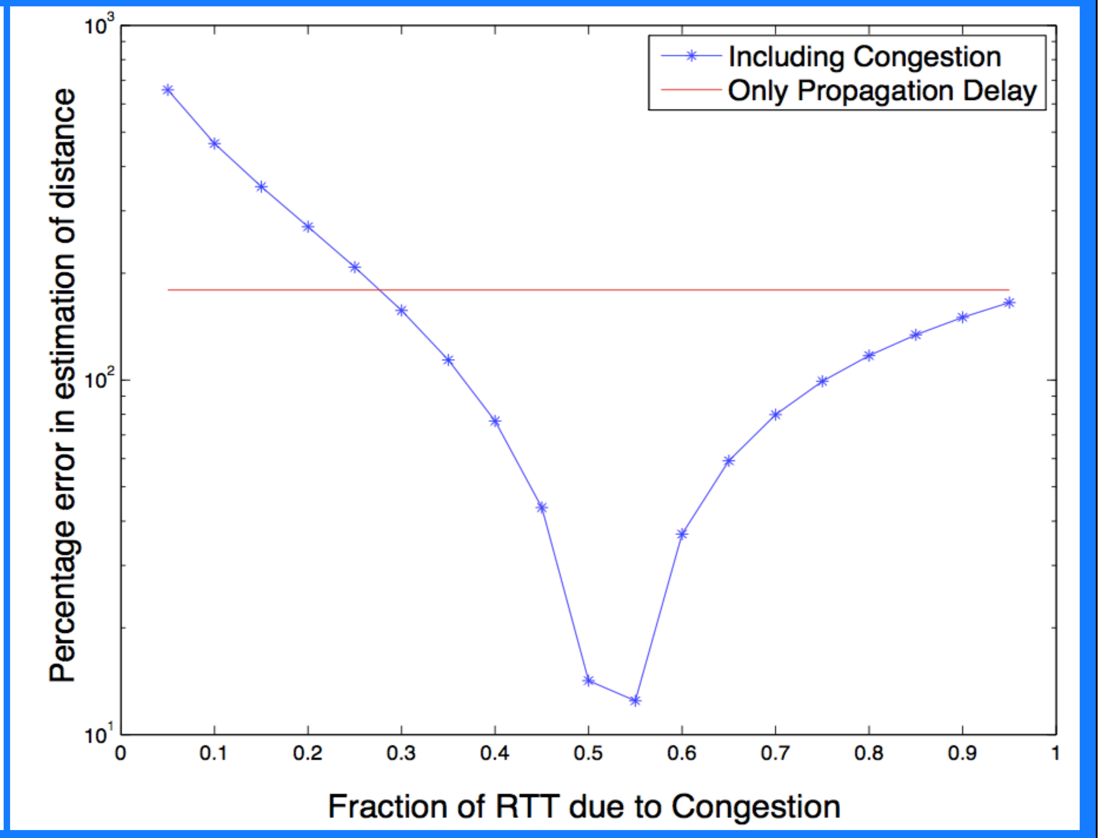

Figure E. Estimation error for estimated distances between Monash University in Australia to Hiroshima, Japan.
Distance = 5,081 miles

Algorithm

1. Calculate the mean RTT between a source-destination pair. Let mean RTT be X.
2. Let confidence interval k = 0.95 and decrement to 0.05 in steps of 0.05. 1-k = fraction of the RTT considered to be due to congestion. i.e. if k = 0.95, there is 5% confidence that the RTT is due to congestion.
3. For each value of k, multiply X by the negative natural logarithm of k. Let value equal Y. Y = Queuing delay.
4. X -Y = Propagation delay.
5. Use (X-Y) to estimate distance in meters when accounting for queuing delay. Use X to estimate distance in meters when considering only propagation delay.
6. Compare both values with the actual distance and determine estimation error.

Results

Depending on the percentage of RTT that is taken to be delay due to congestion, the estimation error in distance is greater when considering propagation delay alone and, in some cases, is worse when queuing delay is considered. The experimental results reveal that the estimation error in distance is smallest when queuing delay is taken to be 50-55% of the RTT for multiple source-destination pairs [Figure D]. This is true for small as well as large distances between source and destination pairs. For near distances [Figure E], considering queuing delay outperforms considering propagation delay alone for a large variation in queuing delay. For larger distances [Figure D], the variation in queuing delay is smaller in order to outperform propagation delay alone.

Future Work

In the future, more data will be collected from more source-destination pairs. Estimation errors will also be collected in periods of low condition and periods of high condition. The algorithm will be developed further according to renewal theory, use of Markov chains, stochastic processes, and queuing theory.

References

[1] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, andY. Chawathe. Towards ip geolocation using delay and topology measurements. In Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, IMC'06, pages 71–84, New York, NY, USA, 2006. ACM.

[2] www.planet-lab.org