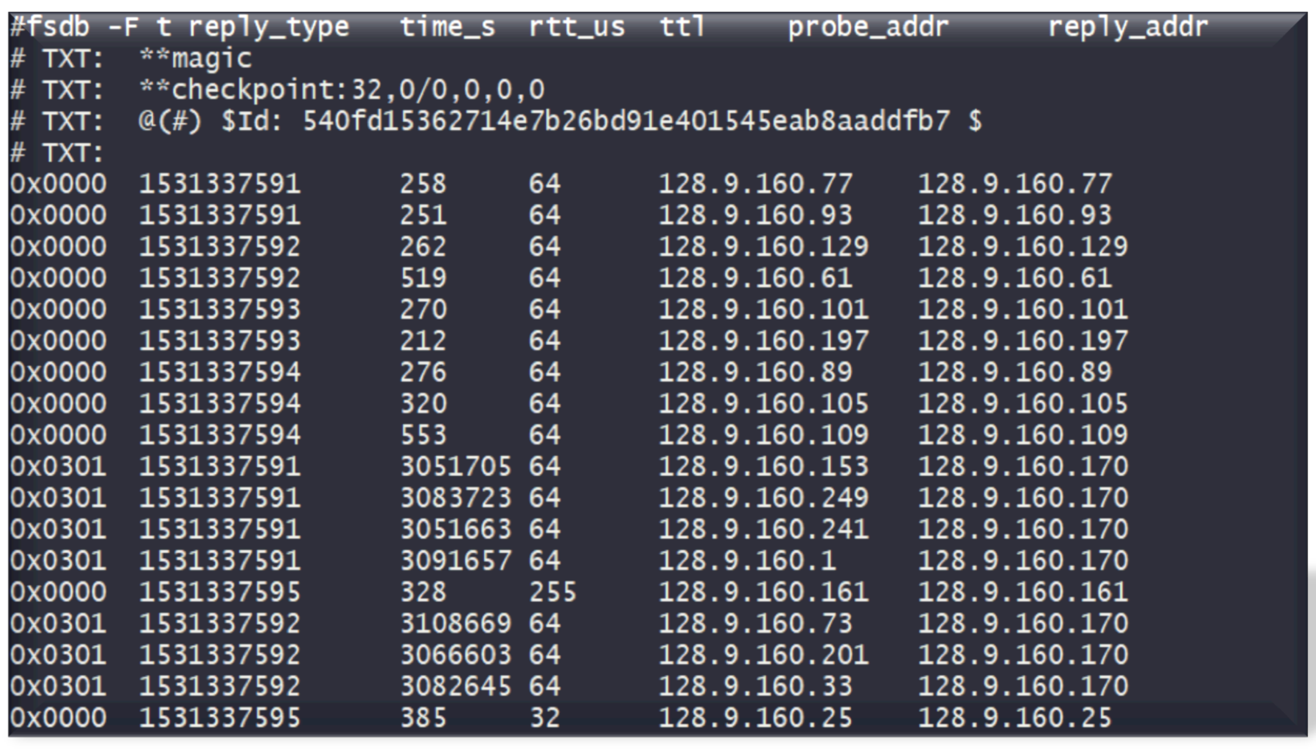
Improving Accuracy of RTT Estimation of the ICMPTrain Algorithm

Christopher O. Morales, Yuri Pradkin and John Heidemann

Problem Statement

- Accurate estimation of RTTs in internet-wide scans would help improve accuracy of geolocation.
- ICMPTrain algorithm probes all IPv4 addresses in the Internet to establish which ones are live
- Currently ICMPTrain probes each address once. If we probe more, we can be more accurate.
- Our goal: design adaptive probing to
  - a) Minimize error
  - b)While minimizing number of probe
- Approach:
  - Modify ICMPTrain to probe adaptively.
  - Measure how many probes per address are enough for correct estimate, and what is the error.

ICMPtrain



```
#fsdb -F t reply_type   time_s   rtt_us   ttl      probe_addr       reply_addr
# TXT:  **magic
# TXT:  **checkpoint:32,0/0,0,0,0
# TXT:  @(#) $Id: 540fd15362714e7b26bd91e401545eab8aaddfb7 $
# TXT:
0x0000  1531337591      258     64      128.9.160.77     128.9.160.77
0x0000  1531337591      251     64      128.9.160.93     128.9.160.93
0x0000  1531337592      262     64      128.9.160.129    128.9.160.129
0x0000  1531337592      519     64      128.9.160.61     128.9.160.61
0x0000  1531337593      270     64      128.9.160.101    128.9.160.101
0x0000  1531337593      212     64      128.9.160.197    128.9.160.197
0x0000  1531337594      276     64      128.9.160.89     128.9.160.89
0x0000  1531337594      320     64      128.9.160.105    128.9.160.105
0x0000  1531337594      553     64      128.9.160.109    128.9.160.109
0x0301  1531337591      3051705 64      128.9.160.153    128.9.160.170
0x0301  1531337591      3083723 64      128.9.160.249    128.9.160.170
0x0301  1531337591      3051663 64      128.9.160.241    128.9.160.170
0x0301  1531337591      3091657 64      128.9.160.1      128.9.160.170
0x0000  1531337595      328     255     128.9.160.161    128.9.160.161
0x0301  1531337592      3108669 64      128.9.160.73     128.9.160.170
0x0301  1531337592      3066603 64      128.9.160.201    128.9.160.170
0x0301  1531337592      3082645 64      128.9.160.33     128.9.160.170
0x0000  1531337595      385     32      128.9.160.25     128.9.160.25
```

ICMPTrain sending multiple pings.

- Sends multiple pings to multiple ip addresses.
- I used ICMPTrain to use pinging.
- I ping 20 times each ip to collect statistics.
- I get IP address from random ip from a hitlist.

Challenges Measuring RTT

Problem: Multiple factors affect

- Bandwidth.
- Queue Delays.
- Physical Distance (Propagation Delay).

- Solution: Having multiple probes from each IP can be more accurate on finding true RTT.

$$E_i = R + Q_i$$

- Minimum RTT filters out queueing delay component of noise.

Conclusions

- We learn that probing 10 times gives you good accurate measurements.
- Simulate this results on an adaptive algorithm that we created.
- We believe that adaptive will be more efficient and just as accurate.
- Plans to run simulation to evaluate accuracy of adaptive algorithm.

Experiments:

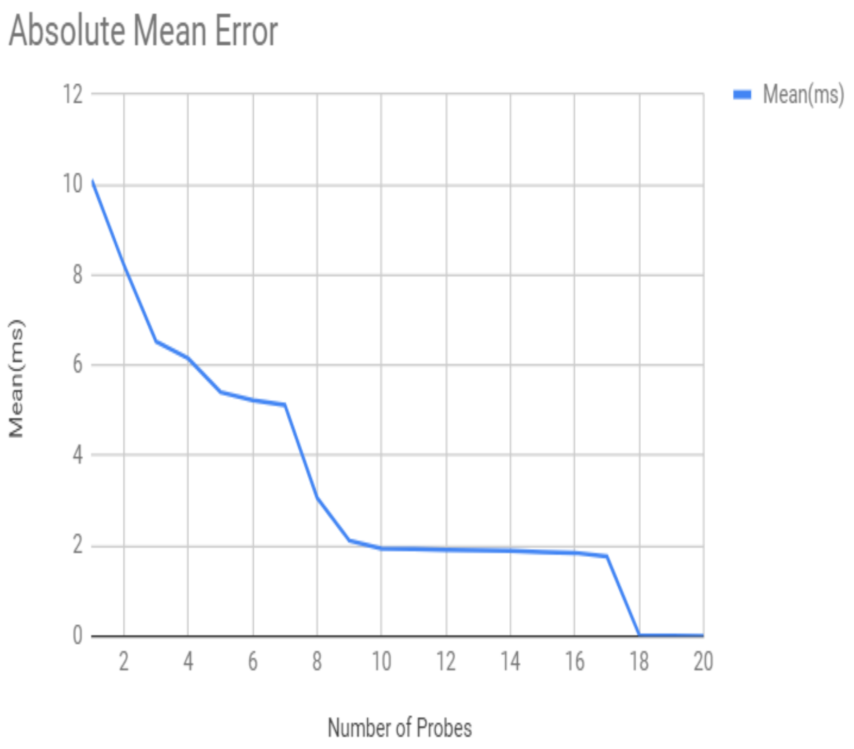
RTT Accuracy Vs. Cost

Hypothesis: Do showing more Probe helps?

Methodology:
EstimateRTT(target_ip, N_probes) = min(observation(1..N_probes, target_ip))

1) Selected random 1000 targets from hitlist
2) Ping each 20 times with a wait time of 5 seconds.
3) Discarded targets with incomplete observation, fewers than 20 results.
4) Calculated the ground true value from 20 probes.
5) GroundTruthRTT(IP) = EstimateRTT(rtt, rarget_ip, 20)
6) Evaluate how much more probes help: EstimateRTT(target_ip, i) for I in 1 to 20 probes

- This is the error, compared to our ground truth for each IP if we probe it 'j' times:
- AbsoluteError(IP , j) = estimate(rtt, target_ip, j) - groundThurthRTTMin(rtt (IP, k)) - GroundTruthRTT(IP)
  k = 1...j



Expected Results: Maybe using few observation is as good as GroundTruthRTT.

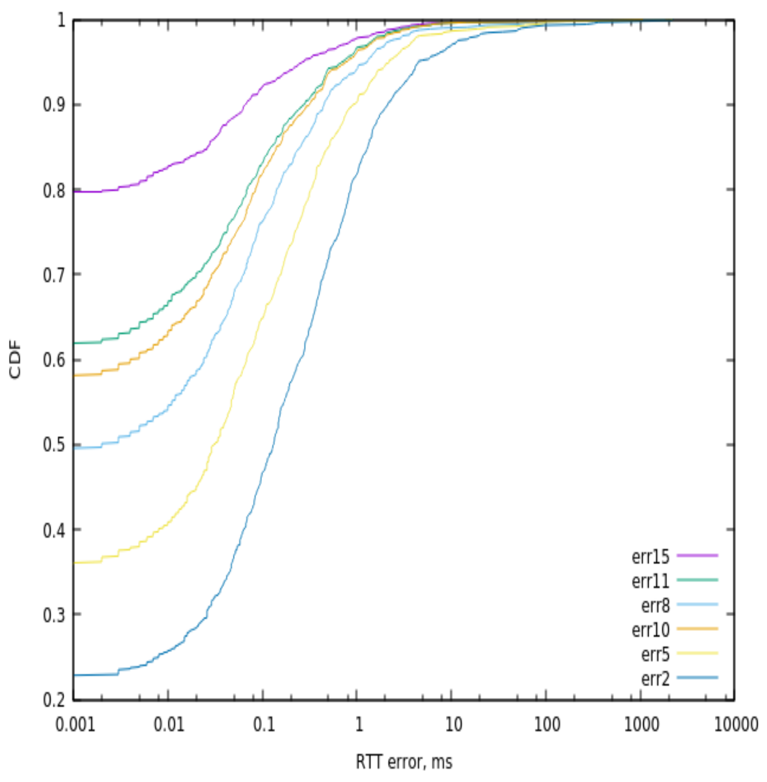
Discussion:
Giving from the graph, the hypothesis is true. Using 9 observation is as good as GroundTruthRTT.

Distribution of RTT Accuracy

Hypothesis: How accurate is more Probe?

Methodology:
Estimatertt(target_ip, N_probes) = min(observation(1..N_probes, target_ip))
GroundTruthRTT(IP) = estimate(rtt, target_ip, 20)
ErrN(N_probes,target_ip) = EstimateRTT() - GroundTruthRTT(target_ip)
CounterTotalAmountError = size(ErrN)
CDF(N_probes) = (count (sort(ErrN(N_probes, target_ip)) / CounterTotalAmountErrorRTT)

1) Selected random 1000 targets around the internet.
2) Ping each 20 times with a wait time of 5 seconds.
3) Discarded targets with incomplete observation, fewers than 20 results.
4) Calculated GroundTruthRTT value from 20 probes.
5) Calculated the ErrN from first probe...last.
6) Sort the ErrN from minimum to maximum.
6) Finishing by calculating CDF which have a counter of the ErrN of each probe(1..20) errors and divided by a counter of total amount of error.



Expected Results: Maybe, for finding the best possibility of finding error is the 10 distribution of estimate.

Discussion:
The hypothesis is true. The distribution of estimate using 10 input, observation has a much higher fraction of finding the X percent of population.

If interested contact <Christopher Morales>
Christopher.Morales8@upr.edu

Work performed under REU Site program
supported by NSF grant #1659886

USC Viterbi
School of Engineering
Information Sciences Institute