Sarah Whelan
slw96
December 2, 2015
Networks 325

Project 2 Report

1. **Explain the technique you use to measure the hop distance from your machine to destination using a single probe.**

In order to use a single probe or packet to determine hop distance I utilized a single UDP packet and ICMP messages. On the single UDP packet I set the TTL field to a number I knew in advance – in this case 32. In theory any number would have worked (that fit within the bits of the field) but as Rami Al-Dalky pointed out and Professor Rabinovich relayed certain equipment will reset this field if the number is too high. This makes sense as a packet that stays alive for a long time without getting to its destination wastes resources. Professor Rabinovich also pointed out that it is rare to have hop counts that are larger than 32. This information is what I used to determine that 32 would be an acceptable value with which to set the TTL. With the TTL a known value I sent the UDP packet to the destination in the targets.txt file and to a port that would likely be unused. The port number I choose to use was an official traceroute port number as this is some variation on that. My reasoning was that most likely any port number would work as long as it was one that was not commonly used. Additionally in the interest of security or performance many proxies and servers may not wish to respond with ICMP messages on ports they are not using as it can be a drain on the network and server and not responding may protect against some more naïve attacks that just scan on ports. However some people may configure their equipment to respond to the Traceroute/Ping ports so as not to lose the ability to debug/check network on their own servers etc. Either way I tried my tool with several different ports manually and got similar response rates for every port I tried. I also set a message in the body of the packet an 8 byte string: abcdefgh. The message was just an additional precaution to ensure the responses I was getting were for the packets I was sending. I also determined my ip address and the ip address that I had send the packet to. Python uses the first DNS record that returns so I had to save the ip I used in order to check the returning packet against the original ip and not just a different first DNS record.

After sending my packet with known TTL to a port that should be unreachable I waited for a response. Initially I had a non-blocking socket I would only poll so many times but this had the same behavior as a blocking socket with a time out and the blocking socket code was cleaner so I went with that but both would have worked. Once I received my response I verified that it was in response to the packet I send (more on this in question 2) and knowing that what I was receiving was a packet with a set of IPv4 headers, then ICMP headers, then the IPv4 headers of the packet that caused the port unreachable ICMP message to be created at the destination and the first/all of the 8 bytes of my message. From this I pulled the TTL header field from the inner IPv4 headers (my original) and subtracted this from my known original TTL of 32 to get the number of hops that the packet took before it got to the destination. RTT was also measured with a single probe in that I got the time just before the packet was sent and again just after the response was received to determine how long the packet was travelling.

I noticed that in using my tool if I was on the case campus or it was a busy time for internet at my house (ie everyone else was awake too) that I'd get a lot of destinations not responding but if I tried the tool a bunch of times one after another I'd get a response so this tool was harder to test when a lot of the UDP packets were getting dropped. The internet activity also had a drastic effect on the RTT I was measuring but not on the hop count.

2. **Explain how you will match ICMP responses with the probes you are sending out**

To match ICMP responses to the probes I am sending out I checked specific header fields and the message field to ensure that the response was for the packet I originally sent.

Specifically I checked the IP source and destination addresses on both set of IP headers and if my message was returned I checked that as well (I was getting some packets where my message was not returned along with the headers but all of the other header fields were in the right spots). Additionally I checked the ICMP headers to ensure it was for the right error. Checked the source and destination IPs on the inner IPv4 packet for matching my ip and the ip I sent the original packet to. Checked the source and destination IPs on the outer IPv4 to match the ip I sent the original packet to and my ip address, respectively. If I got my message back I checked that it matched. Also checked if both the type and code of the ICMP message was 3 indicating it was host/port unreachable. I thought about setting and checking the identification field of the original packet but that should be used only for fragmentation and is apparently not allowed. (https://tools.ietf.org/html/rfc6864 and https://en.wikipedia.org/wiki/IPv4#Identification)

3. **list all possible reasons you can think of for not getting the answer when probing an arbitrary host**

- UDP packet was dropped on the way to the destination.
- ICMP message was dropped on the way back from the destination
    - Both of the above would be more likely in high traffic scenarios
- proxy dropped packet (either direction)
- firewalls
- Destination did not want to respond with ICMP
    - configured to not respond to ping/traceroute like things to avoid ping flooding/other attacks
- No internet connection
- Destination not found
- Server down
- For specifically my tool – timeout after 3 seconds
- There are many things that could go wrong I think that main culprits for lack of response for me were intermittent packet dropping and proxy/firewall/server settings to not respond

Data

List of original sites:

wikipedia.org
bongacams.com
wikihow.com
google.com.pe
kohls.com
souq.com
lemonde.fr
houzz.com
mi.com
tiexue.net

Only the ones in bold responded so to get at least 10 sites I substituted some other sites shown below in the table.

The following data was collected in one run of the tool at night and these were some of the best RTTs I received. In the runs the RTT would vary dramatically in periods of high traffic but the hops would not change as much. As this was one run the network conditions were the same for all of them and I choose to do this instead of averaging as I would have been averaging with much higher RTTS/possibly different hops (hop changes could be tangentially related to traffic but could change normally).

Data (Collected 11pm December 1st):

Host	RTT (milliseconds)	Hops	Distance(km)
case.edu	0.247002	3	0
google.com.pe	13.886929	9	3470
google.com	14.21308582	9	3470
news.ycombinator.com	18.4021	8	3484
etsy.com	18.975019	14	1370
kohls.com	22.56012	14	874
target.com	34.42192186	11	874
twitter.com	41.708946	8	3484
callmedrew.com	66.869974	15	3296
espn.com	70.734024	18	3315
facebook.com	88.037968	14	3478
dropbox.com	99.351168	14	3484
lemonde.fr	106.021166	11	1370

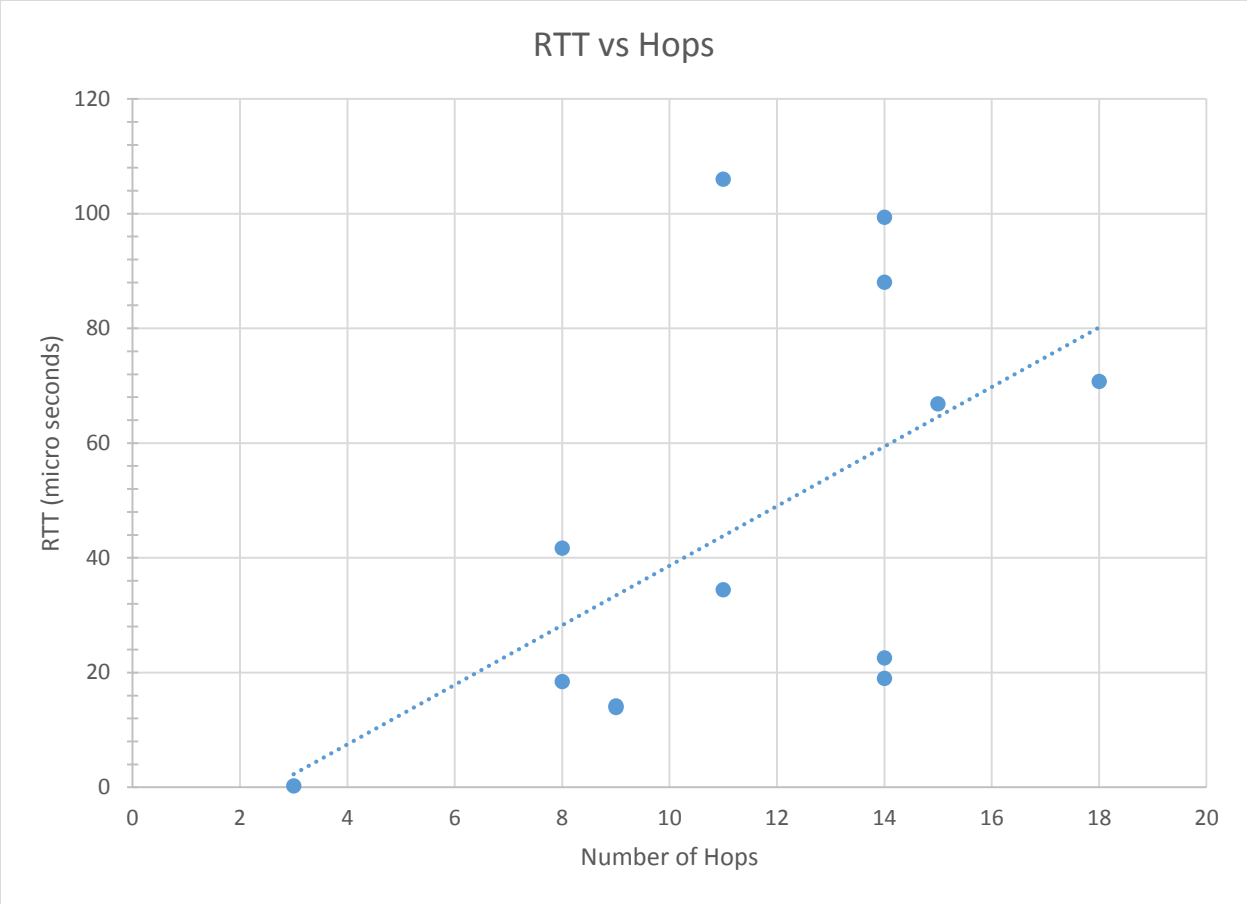
The above table does not include the non-bolded sites from the list of original targets that did not respond (or the packet was dropped). The data is ordered by RTT shortest to longest.

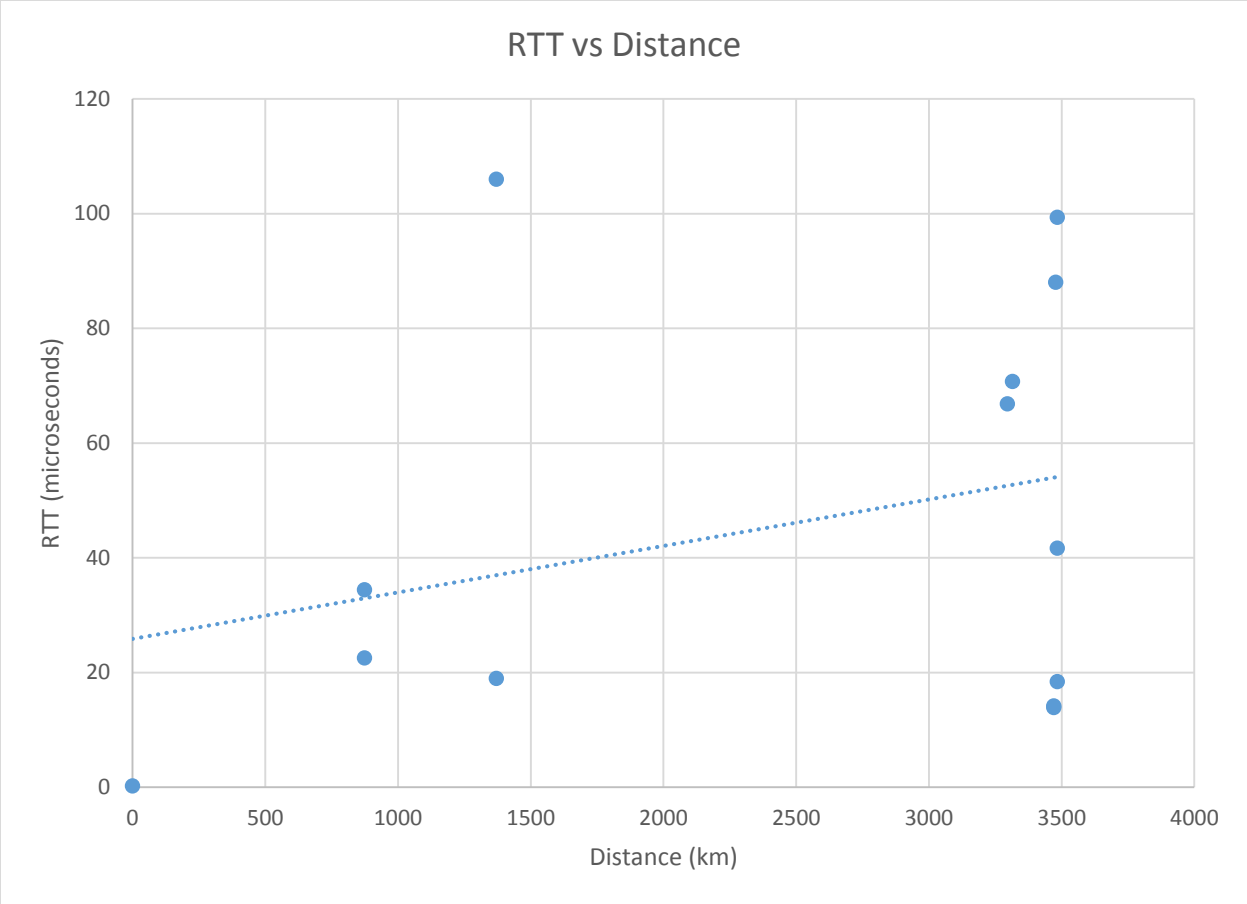
Correlation Coefficients (Calculated in Excel - Data.xlsx):

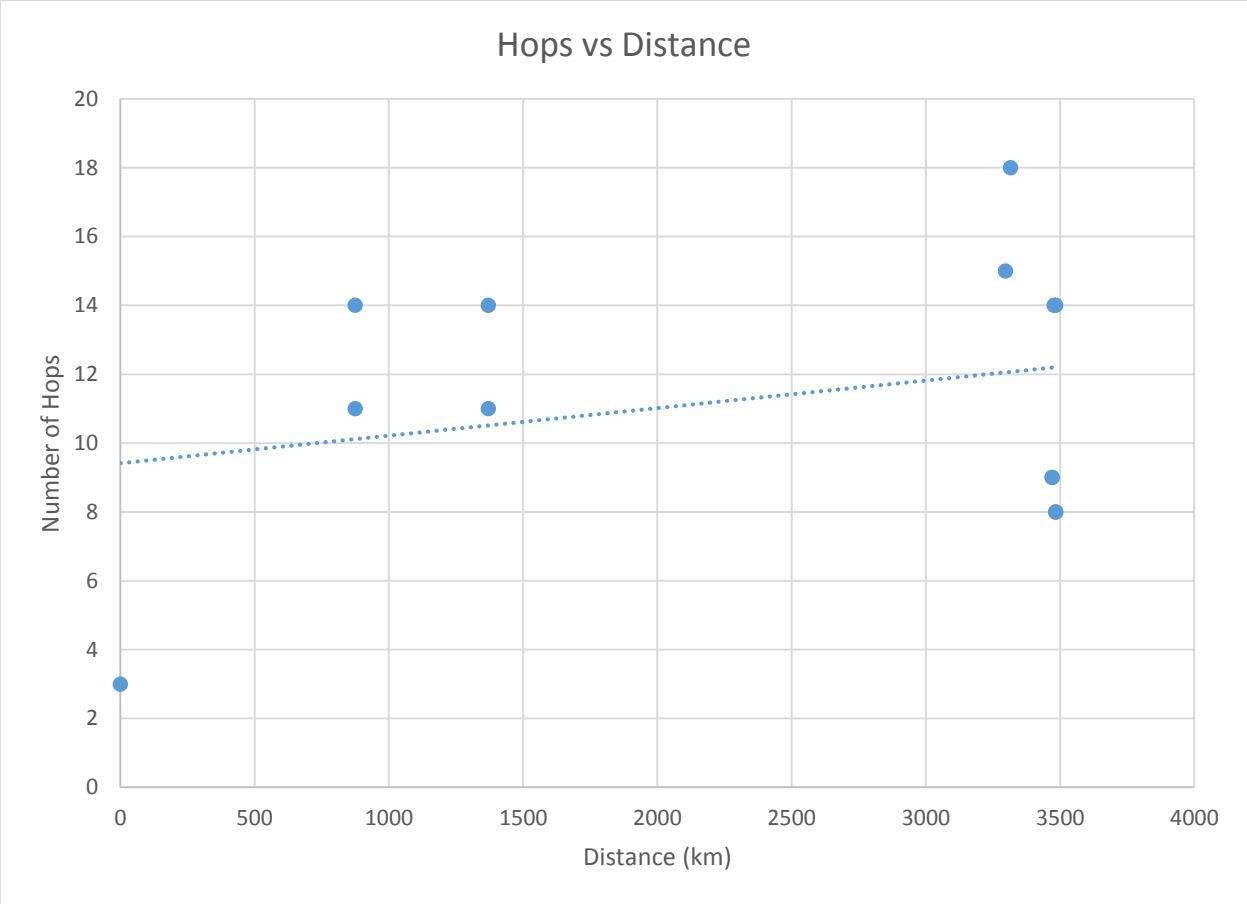
RTT and Hops: .57
RTT and Distance: .30
Hops and Distance: .27

Conclusions

There is some limited correlation but so many outliers that is difficult to tell with certainty that one metric will increase another. However given some thought about networks in general we should expect RTT to increase with distance and number of hops (more so with distance). Hops should probably also increase with distance as either the links must get longer to go farther or more routers must be added and as such more hops. However it is obviously not just these factors that contribute to RTT as there could be a much higher RTT time during heavy traffic loads but likely the same or similar (+/- 1 or 2 for normal routing decisions) hops and the distance would stay the same. Additionally for things like google.com clearly even though they are far away they have some of the lower RTT times this is an effect of their infrastructure using faster links and other optimizations that other smaller organizations don't or can't do. These optimizations lead to outliers in my data and lower correlation coefficients. Additionally there may be a single bad link or proxies/firewalls along a given route that alter the RTT but not the distance or hop count as if a proxy adds time to the trip it will only increase the RTT this is common for institutions. Traceroute would be able to better identify those types of delays (single hops causing issues) however given that this tool only sends one probe this is reasonable data. Graphs are attached.


RTT vs Hops
RTT (micro seconds)
Number of Hops
0
20
40
60
80
100
120
0
2
4
6
8
10
12
14
16
18
20


RTT vs Distance
RTT (microseconds)
120
100
80
60
40
20
0
0
500
1000
1500
2000
2500
3000
3500
4000
Distance (km)


Hops vs Distance
Number of Hops
Distance (km)
0
2
4
6
8
10
12
14
16
18
20
0
500
1000
1500
2000
2500
3000
3500
4000

Raw Program Output

distMeasurement.py

```
root@eecs-socket-104:/home/eecs-student/project2# python distMeasurement.py
Data for host: google.com.pe (216.58.216.195)
Hops: 9
RTT (milliseconds): 13.886929
Data for host: kohls.com (104.76.108.19)
Hops: 14
RTT (milliseconds): 22.560120
Data for host: lemonde.fr (93.184.220.20)
Hops: 11
RTT (milliseconds): 106.021166
Data for host: google.com (216.58.216.78)
Hops: 9
RTT (milliseconds): 14.213085
Data for host: case.edu (129.22.108.21)
Hops: 3
RTT (milliseconds): 0.247002
Data for host: news.ycombinator.com (198.41.191.47)
Hops: 8
RTT (milliseconds): 18.402100
Data for host: callmedrew.com (74.124.211.171)
Hops: 15
RTT (milliseconds): 66.869974
Data for host: facebook.com (69.171.230.68)
Hops: 14
RTT (milliseconds): 88.037968
Data for host: twitter.com (199.16.156.38)
Hops: 8
RTT (milliseconds): 41.708946
Data for host: espn.com (199.181.132.250)
Hops: 18
RTT (milliseconds): 70.734024
Data for host: dropbox.com (108.160.172.200)
Hops: 14
RTT (milliseconds): 99.351168
Data for host: target.com (72.247.9.136)
Hops: 11
RTT (milliseconds): 34.421921
Data for host: etsy.com (192.147.0.123)
Hops: 14
RTT (milliseconds): 18.975019
```

geoDistance.py

```
root@eecs-socket-104:/home/eecs-student/project2# python geoDistance.py
Distance for google.com.pe: 3470 km
```

Distance for kohls.com: 874 km
Distance for lemonde.fr: 1370 km
Distance for google.com: 3470 km
Distance for case.edu: 0 km
Distance for news.ycombinator.com: 3484 km
Distance for callmedrew.com: 3296 km
Distance for facebook.com: 1526 km
Distance for twitter.com: 3484 km
Distance for espn.com: 3315 km
Distance for dropbox.com: 3484 km
Distance for target.com: 874 km
Distance for etsy.com: 1370 km