Assignment 4
Due: May 10, 2015 (Sunday) at 12 noon
Submit in
Dropbox: https://adb.auckland.ac.nz/Submission/Submit/391

This assignment is worth 7% of the total marks. All questions have the same weightage. Please submit your answers in a PDF file on Dropbox before the deadline. This is an individual assignment. The answers in the assignment should be your own. Many questions require you to comment on the observations. Please keep your comments succinct. The comments should not exceed 150 words. It is important for your learning process to find a good explanation. Assignments after the deadline will be accepted until May 13, 2015 at a flat penalty of -25% of the marks attained. Late assignments must be submitted by email to Se-young Yu (se-young.yu@auckland.ac.nz) and CCed to Aniket Mahanti (a.mahanti@auckland.ac.nz). We are always available to help you with the assignment. Please do not hesitate to contact the lecturer or tutors. This assignment is based on assignments by Prof. Stoica and Prof. Williamson.

Part 1 (Traceroute and Ping)

Using your home computer, perform *ping* and *traceroute* to the following hosts around the world:

1. East coast USA (Boston): www.bu.edu
2. West coast USA (Riverside): www.ucr.edu
3. Southern USA (Austin): www.utexas.edu
4. Central India (Guna): www.juet.ac.in
5. Japan (Kyoto): www.kyoto-u.ac.jp
6. West Africa (Ilorin, Nigeria): www.unilorin.edu.ng
7. Central Europe (Lausanne, Switzerland): www.epfl.ch
8. Central Europe (Berlin, Germany): www.tu-berlin.de
9. Northern Europe (Linkoping, Sweden): www.liu.se
10. Eastern Europe (Novosibirsk, Russia): www.nsu.ru
11. Middle East (Tel Aviv, Israel): www.tau.ac.il
12. Australia (Sydney): www.unsw.edu.au
13. South East Asia (Bangkok, Thailand): www.ait.ac.th

a) Take seven ping measurements to each of the distant hosts over a period of seven different days. Use the default options when using ping. For each ping, four measurements are taken, and ping reports the maximum, minimum, and average RTT. If you take seven ping measurements, then you will have 7 *4 = 28 RTT values. Save these ping measurements so that you can analyze them later.

Simultaneously, perform seven traceroute measurements to each of the distant hosts over a period of 7 different days. Use the default options when using traceroute. Count the number of hops for each traceroute measurement. Save these traceroute measurements for analysis later.

Using these 28 RTT values (ping) and 7 traceroutes, complete the following table:

Host	Average RTT (ms)	RTT Variance (ms)	Average Hops	Hop Variance
Boston				
Riverside				
Austin				
Guna				
Kyoto				
Ilorin				
Lausanne				
Berlin				
Novosibirsk				
Linkoping				
Tel Aviv				
Sydney				
Bangkok				

Use the sample variance formula. Sort the table in ascending order of average RTT. The city with lowest average RTT will be at the top and the city with highest average RTT will be at the bottom. Briefly comment on why the average RTT and average hops are higher for some cities compared to other cities and why you observe variance in RTT values and hops. Do you notice the major undersea links in the traceroutes?

b) Calculate the distance between the above-mentioned cities and Auckland using the *great circle distance* formula (The shortest distance between two points on the surface of the Earth). You can use the following website to perform the calculation: http://www.infoplease.com/atlas/calculate-distance.html Let us call this value $\mathbf{D}$.

For each of these cities, calculate the time it takes for a beam of light to travel from Auckland and back (round trip). Use the speed of light in vacuum ($3 * 10^8$ m/s) for your calculations. Let us call this value **T**.

Therefore, **T** = $\frac{\mathbf{D}}{(\mathbf{3} * \mathbf{10^8})}$

Let us call the average RTT value from the table in (a) as $\mathbf{R}$.

For each city calculate $\frac{\mathbf{R}}{\mathbf{T}}$.

Complete the following table:

Host	R/T	D (km)
Boston		
Riverside		
Austin		
Guna		
Kyoto		
Ilorin		
Lausanne		
Berlin		
Novosibirsk		
Linkoping		
Tel Aviv		
Sydney		
Bangkok		

Sort the table in ascending order with the nearest city at the top and the farthest city at the bottom. Briefly comment on why R/T > 1 and whether you observe any correlation between distance and average RTT.

Part 2 (IP Geolocation)

You will utilize *traceroute* and *ping* to *geolocate* an IP address and compare your results with an online geolocation database.

You will need to identify where the host with **IP address 212.162.63.104** is located. To do this you will use the *shortest ping* method. The shortest ping method relies on pinging the host from multiple known landmarks around the world. The landmark with the lowest RTT is used to identify the location of the host. That is, we can assume that the host is located in the same region where the landmark with the lowest recorded RTT is located.

To improve accuracy of geolocation, you will also need to investigate traceroutes from the landmarks to the host. If you inspect the last few hops of the tracroutes to the host, it will give you an idea of the location of the host. The router names in the traceroute often have identifiers for cities (e.g., abbreviations for airport codes or city names). Thus, by analyzing the names of routers closest to the host will help in improving the geolocation.

You will perform ping and traceroutes using http://www.super-ping.com/

In which country is the host located? Compare you answer with a popular geolocation database (https://www.maxmind.com/en/home) and an Internet Registry (https://apps.db.ripe.net/search/query.html). Briefly explain your answer.

For additional details about IP geolocation and the shortest ping method, refer to the following article and the references therein: https://djw.cs.washington.edu/papers/imc2006-geoloc.pdf

Part 3 (Internet Traffic Characterization)

Please refer to the trace file called *trace*.txt, which depicts the traffic activity in a production network link. The trace was collected several years ago.

Each line in the plain text file shows one (TCP) packet. There are 6 columns in the trace file separated by a space:

Column 1 is timestamp (in seconds) at which the packet was observed by the packet sniffer

Column 2 is the source ID (IP address anonymized) for the packet

Column 3 is the destination ID (IP address anonymized) for the packet

Column 4 is the source port for the packet

Column 5 is the destination port for the packet

Column 6 is the packet size in bytes. Note the size only includes the size of the actual data and not packet headers. Thus, a packet of size 0 is a TCP acknowledgement.

Let us look at an example as to how to interpret the entries in the trace file:
0.186911 2 1 2436 23 1
This entry in the trace file shows that a packet was transferred from host 2 (using port 2436) to host 1 (using port 23) at time 0.186911. The packet transferred 1 byte of data.

Using the trace file, complete the following tasks:

a) Complete the given table:

Number of packets transferred	
Average packet arrival rate	
Average data rate (bits/second)	
Trace duration (seconds)	
Number of distinct sender hosts	
Number of distinct receiver hosts	
Number of distinct source ports	
Number of distinct destination ports	

b) For each consecutive one-second interval, calculate the number of packets transferred in the link. Plot this packet count versus interval value as a time series plot. Create a similar plot for data transferred per one-second interval. Comment on the packet and byte traffic characteristics. Do you observe bursty packet arrivals? Are there times when the link is busier than other times?

c) Using the packet size column in the trace plot a frequency histogram to show the packet size distribution. Use the following packet size intervals in bytes for the histogram: 0-200, 201-400, 401-1000, and 1001-1600. Do you see a bimodal distribution? Provide a brief summary of your observations.

d) Calculate the number of packets sent by each unique host. Now, sort the results in descending order based on the number of packets transmitted by each host. The host that transmits the highest number of packets has a rank 1, the next busiest host is ranked 2, and so on. Plot the number of packets sent by each host (y-axis) versus the host rank (x-axis). Use log scale on both the x-axis and y-axis. Do you observe a straight line with a sharp drop off towards the lower ranked hosts? The plot exemplifies power-law behaviour in Internet traffic. Provide a brief summary of your observations.

e) Calculate the number of packets sent from each distinct port in the trace. Similar to (d), plot the number of packets sent from each unique port (y-axis) versus the port rank (x-axis). Use log scale on both axes. Provide a brief summary of your observations. Comment on the usage of well-known ports (0-1023), registered ports (1024-49151), and dynamic ports (49151-65535). Also, comment on how applications can be identified using port numbers, and if this method is accurate. Additional details about port number assignments can be found here: http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.txt

Learning Objectives

After successfully completing this assignment, the student will be able to:

- Understand and characterize various delays involved packet transfers in the Internet, which includes transmission, propagation, queuing, and processing delays.
- Understand and explain the structure of today's Internet, which includes major top-tier ISPs, regional ISPs, access ISPs and content providers.
- Understand Internet resource (e.g., IP address) registration and management system through network of regional Internet registries, which ultimately facilitates communication of hosts on the Internet.
- Perform geolocation of IP addresses, which is used for better delivery of services on the Internet.
- Analyze trace files to understand and characterize Internet traffic behaviour, which can be used for traffic modeling and performance improvement of Internet protocols.
- Understand usage of Internet applications, which can be used for better traffic management and capacity planning in networks.

The skills developed after completing the assignment will be helpful in both industry (e.g., network administrators) and academia (e.g., graduate studies).