Assignment 1: Getting To Know Network Traffic

COL 334/672, Diwali'24

August 5, 2024
Deadline: August 19, 2024

Goal: The goal of this assignment is to familiarize you with network data collection, traffic analysis, and basic network measurement tools. Appropriate hints have been provided throughout the assignment. If you still have questions, you are encouraged to start a discussion on Piazza.

1 Measurement Tools

The first part of the assignment will involve using `ping` and `traceroute`. Yo ucan read about these tools from:

- Ping
- Traceroute

1.1 Ping

Ping the following two websites: `google.com` and `sigcomm.org` (FYI, Sigcomm is the top conference in computer networking). You should ping these websites 10 times and attach screenshots for each case. Also, perform the ping from two different networks: first from within the IITD network and second from a mobile network by connecting your laptop to a cellular network hotspot.

A. Compare the average ping latencies for the two websites in the same network. What could be the reason for differences in average ping latency? Now, compare the difference between latencies of the two networks for the same website. Explain the potential reasons for the difference. You can refer to traceroute results in the next part to answer this question.

B. Explain the protocol being used by the `ping` tool. What is the theoretical upper limit of packet size for the ping protocol? Are you able to ping the websites with the theoretical maximum? Explain why or why not.

C. Now try to force both networks to ping using IPv6. Explain how you did it and whether you were successful (attach the relevant screenshots). If not, what is the reason in each case?

1.2 Traceroute

Log the server IP addresses for the two websites in above case. Use traceroute to find the path taken by the packets in each of the four cases (<website, network>pair) and attach the screenshot.

A. Mention the number of IP hops as well as the list of autonomous systems observed in each case. Note there are online tools that can map an IP address to its autonomous system.

B. Did you observe "*" in your output? If yes, explain the reason.

C. Did you observe multiple IP addresses for the same hop count? If yes, explain the reason.

D. What is the IP address of the first hop router when tracerouting google.com using IITD WiFi. Ping this address from mobile data. Should you receive a response? Justify your answer.

E. Do you observe a 3-tiered (or a 2-tiered) Internet architecture in any traceroute? What is happening in the case where you don't observe such an architecture?

F. Try to geolocate the IP addresses. You can use two different methods: First, try doing the reverse DNS lookup on the IP address and see if you can infer the location from the DNS address. If the reverse DNS lookup fails, use the Maxmind database for IP geolocation. Note the IP geolocation can sometimes be wrong, especially if you are using the Maxmind database. In fact, accurate geolocation of IP addresses is still an active area of research. Now compare the geographical path with the observed RTTs. Do these intuitively make sense? Explain why.

2 Network Data Collection and Header Analysis

For this part, you need to first collect network traffic for a 2-person, 1-minute long Microsoft Teams call. Keep the video and microphone on during the call. You should do this part in a pair. You can use Wireshark or CLI tools such as tcpdump to collect the network data. Answer the following questions:

B. What are the network, transport, and application-layer protocols used by the Teams call? Log the number of packets for each protocol as a percentage of total packets. Try to identify as many application-layer protocols in the traffic. *[Hint: You can use Wireshark filters for this analysis.]*

C. Do you observe a direct connection between the two hosts? If not, what is the endpoint for each host (both IP and the network)? Is it the same endpoint or not? Explain what could be happening if it is not a direct connection.

D. Identify the audio and video packets from the traffic capture and report their number. Explain the logic that you used. Plot a time-series diagram showing the bandwidth utilization by the two media types. You can use either Wireshark display filters or write a script for this analysis.

Note, you need to upload the PCAP along with the submission.

3 Traffic Analysis and Network Performance

In this part, you are given a network traffic trace corresponding to a speed test using the M-Lab NDT7 tool. The NDT7 speed test tool works by flooding the network path between the client and the server for a pre-decided duration and logs the observed throughput. This is done for both downlink (server to client) and uplink (client to server) direction in a sequential manner. Note that there might have been some background traffic while the test was running, which also gets logged in the traffic trace. You need to achieve the following objectives:

A. Isolate the traffic corresponding to the speed test from the background traffic. You will need to read about how the NDT7 speed test works. What percentage of traffic is the speed test?

B. Plot a time-series of observed throughput over time in each direction. You can plot average throughput per second.

C. Find the average download and upload speeds.

You need to write a Python script to answer the above questions. There are Python libraries like scapy and dpkt to parse a PCAP file. Your script should be named `speedtest_analysis.py`. It should take the PCAP as input as well as the following command-line arguments, each corresponding to one question:

- `--plot` should output a single time-series plot corresponding to part B.

- `--throughput` should output the average download and upload speeds [in Mbps] as comma-separated values.

For example, `python speedtest_analysis.py speed.pcap --plot` should answer part B.

Submission Instructions

Your submission should contain a single PDF (other formats will not be graded) called `report.pdf`.

1. For part 1, you should attach the screenshots in the PDF itself. All questions in the first part should be answered in the report.
2. For part 2, you should submit the collected PCAP. Name it `<entry_no_1>_<entry_no_2>.pcap` with your and your partner's entry number. The answers to the questions should be in the PDF. If you used Wireshark (highly encouraged) for the analysis, please mention the filters used as well as the steps (briefly) to get the plot. If you use a script to analyze traffic, please submit the script and name it `vca.py`.
3. For part 3, you should include the logic, plot, and observed throughput values in the PDF. In addition, submit the `speedtest_analysis.py` file.

Please submit a single zipped file containing all the above files.