[CS-335a] ASSIGNMENT 1

The Hidden Life of Networks: Network Measurements

Deadline: 29/10/2021
Professor: Maria Papadopouli
TA: Eva Perontsi
TAs mailing list: hy335a@csd.uoc.gr TA: evaperon@csd.uoc.gr

---

1. Network Measurements (50 pts)

A. Passive measurements (30 pts)

The objective of this part is to get familiar with *passive network measurements,* packets and protocols. You will have to use **Wireshark**, an open-source network analysis tool. Please follow the setup instructions in the Wireshark-setup.pdf file.

1. Run Wireshark and start capturing packets. Visit SkylineWebcams and select a video stream of a destination you like. Play the live stream of your choice for 1 minute. Then, stop the streaming video and the packet capturing via Wireshark. Estimate the following metrics, and report them:
    a. How many packets have you captured? (10 pts)
    b. What was the bitrate of the video? (10 pts)
    c. How many different protocols have you captured? (10 pts)

    Include screenshots in your review.

B. Active measurements (20 pts)

The main objective here is to perform simple *active measurements*. In contrast to the previous task where you *passively* observed whatever went through your network, now you are the one who initiates the analysis. You will use **ping** to measure packet loss and delays, and **traceroute** to discover the topology of the network. Your target website will be www.paper-io.com.

1. Run **traceroute** and count the hops. If the router at some hop is not responding, try sending a different type of packet. (10 pts)
2. **Ping** the target with 10 ICMP echo requests and write down how much time (in milliseconds) it took for these 10 packets, the packet loss percentage and the statistical metrics (min, max, avg, mdev). (10 pts)

In your report you must include both the answers to the questions and screenshots of your experiments.

2. Network Analysis (50 pts + bonus 20 pts)

In this section, you will see, process and illustrate real network data. It is recommended to use python3 for your scripts, but you can write them in the language of your preference.

A. Datasets

You will use the packet-delivery dataset from CRAWDAD (Data license here). CRAWDAD is a community resource for archiving wireless data at Dartmouth. Even though registration is mandatory in order to download data from CRAWDAD, we have asked their permission to share a specific dataset with you for this assignment. **Please do not further share this dataset, anyone outside this course who wants to use it should register.**

The two datasets you will be using are in this drive folder: CS335a .

DATASET1.csv consists of traces from Android apps (primarily video) collected under different values of parameters, such as video length, connection strength and device mobility, for the purpose of mobile video app traffic pattern identification.

DATASET2.csv is the Dataset of BitTorrent traffic from Korea Telecom's mobile WiMAX network, collected in March 2010.

The fields of the datasets are:

1. No.: the serial number of the packet

2. Time: the time of the transmission/receiving of the packet (starts from 0, the moment that the capturing started) in seconds
3. Source: the source IP address
4. Destination: the destination IP address
5. Protocol: the protocol used
6. Length: the length of the packets in bytes
7. Info: extra information about the packet (header fields, flags etc)

B. Simple Measurement Analysis

1. For the first question you will use **DATASET1**, which is in .csv format. You have to write two scripts:
    a. A parser that will read the .csv file and store its data (10 pts)
    b. A script that creates the empirical cumulative distribution function (ECDF) of the length field of (only!!) the **UDP** packets. (10 pts)
    c. A barplot of the total amount of traffic sent and received for the top-5 protocols that you detect in this file, e.g., UDP, HTTP, TCP. The top-5 is determined based on the occurrence in the recorded packets of that file. In the barplot the x-axis indicates the protocol, and the y-axis the total amount of traffic sent and received for the corresponding protocol. (15 pts)

    Include the ECDF plot and barplot in your report, and don't forget to name the axes and make the figure presentable (experiment with colors, type of lines etc)

2. For this question you will use **DATASET2**, which is also in .csv format.
    a. Write a parser that will read the .csv file and store its data (feel free to reuse the one from the previous question if it works properly :)) (5 pts)
    b. Compute the average length of all TCP packets (10 pts)
    c. **BONUS**: A script that creates a scatter plot: (20 pts)
        i. Get 100 **random** samples of data that use the TCP protocol
        ii. Plot a time series with bin size of 1minute (x-axis) and y-axis the amount of *total traffic* (in bytes) (i.e., sent and received) from that network *during that time bin*

iii. Estimate the mean, median, standard deviation, and 95% percentile of the total traffic (i.e., sent and received) in one minute;

iv. Plot the ECDF of the total traffic sent and received in one minute;

Do *not* forget to include your plots in your report.

Enjoy!