ECS 152A: Computer Networks

Fall 2022

Project 1

(100 points)

Due Date: Friday October 21, 2022 by 5:00 PM

Team: The project is to be done in a team of at most 2 students. You *cannot* discuss your code/data with other classmates (*except* your project partner).

All submissions (including your code) will be checked for ***plagiarism*** against other submissions as well as the public Internet. Plagiarized submissions will be entitled to ***zero*** points.

Project 1 consists of two parts:

1. Monitoring network traffic of a website
2. Analyzing network traffic in a Pcap file

Part 1: Monitoring Network Traffic from a Website (*50 points*)

In this part, you will visit and analyze a given set of websites in your web browser with Wireshark running in the background.

Upon visiting each website you will perform the following actions:

1. Click on at least 5 different links on the website
2. Click on at least 5 different images on the website
3. Play at least one video on the website (for sites which include videos)

Note that Wireshark captures all the traffic to and from your chosen network interface, which can include irrelevant packets generated from applications running in the background, network activity caused by the operating system, etc. These irrelevant packets will be mixed with packets exchanged with the website you are visiting. You should try to minimize it (it is very difficult to completely stop all other network activity) by ensuring that no other application is running in the background and only one browser tab is open while visiting the websites. Also make sure you are not using any ad blocker or any other content blocking extension while visiting these websites.

After performing these actions, you will save the Wireshark capture for the site in a Pcap file (do not use PcapNg). You will analyze this Pcap file with the help of dpkt library (https://dpkt.readthedocs.io/en/latest/index.html). **You will generate a separate Pcap file for each site you visit.**

In addition to monitoring network traffic through Wireshark, you will also make use of DevTools (e.g., https://developer.chrome.com/docs/devtools/open/ in Chrome, https://firefox-source-docs.mozilla.org/devtools-user/network_monitor/ in Firefox). These

DevTools tools allow you to monitor the network traffic between the browser and the websites' server. While Wireshark will capture raw information about the packets going to and from the website, DevTools will provide relevant information about the nature of the web traffic. You will make use of both Wireshark and DevTools to better understand the differences between types of traffic present on these sites. Record your observations with DevTools and export it into a HAR (HTTP Archive) file (https://browserhow.com/how-to-record-and-generate-har-file-in-chrome-browser/). Remember to check the *preserve log* option in DevTools otherwise the HAR file will be unable to capture data across page visits of the same website. **You will generate a separate HAR file for each visited website.**

Along with the report, you will also need to submit the Python code used for analysis, Pcap files generated for each visited website, and the HAR files generated for each visited website (20 points).

You will visit the following websites:

1. https://youtube.com
2. http://www.videolan.org/
3. https://joinpeertube.org/
4. https://www.etrigannews.com/NS51au6RhG//news_site
5. https://www.tmz.com

Report:

wireshark_[name1]_[student_id1]_[name2]_[student_id2].pdf (30 points)

At the beginning of the page, specify the following:

1. Full Name of student 1 (Student ID) (Discussion Group)
2. Full name of student 2 (Student ID) (Discussion Group)
3. Name of the code and Pcap files submitted

Answer the following questions in your report:

1. How many UDP and TCP packets did you observe for each website? (3 points)
2. How much network traffic (number of packets sent) is secure (HTTPS) vs vulnerable (HTTP) on each site? (https://www.cloudflare.com/learning/ssl/why-is-http-not-secure) (3 points)
3. What is the distribution of different types of packets that you observed for each site? Calculate and report the percentage of packets observed for HTTP, HTTPS, DNS, FTP, SSH, DHCP, TELNET, SMTP, POP3, and NTP. (**hint:** look at port numbers) (6 points)
4. Report the number of unique destination IP addresses per site. Is there any discernible difference between each site based on the number of destination IP addresses? Do you see any direct relationship between number of destination IP addresses and load time of the site?

5. List the top 5 destination IP addresses based on the number of packets sent. Can you identify who owns these IP addresses? (**hint**: making use of Dev Tools and the HAR files generated to determine hostnames of some of the IP addresses can make this easier). (12 points)
6. Is it possible that different IP addresses are mapped to the same hostname? Can you find an example of this from the sites that you visited and explain why this might be happening. (6 points)

Part 2: Analyzing Pcap file (*50 points*)

In this part, you will be given a Pcap file (https://ucdavis.box.com/s/8t2018rvcbikvaa74sj5oibx070v71pz). The Pcap file was generated while Wireshark was listening to a "hotspot" on a machine. The hotspot provides Internet connectivity to wireless devices using WiFi. Devices such as mobile phones, smart watches, laptops, smart speakers, etc that have a WiFi Network Interface Card (NIC) are able to connect to the Internet via this hotspot. Each device in this Pcap file has a distinct IP address which can be used to identify the packets coming from/to that device. An IP address is a unique number which is separated by periods and identifies a device on the internet or a local network. It consists of 4-bytes of data— an example address might be 192.158.1.38. Each number in the set can range from 0 to 255. So, the full IP addressing range goes from 0.0.0.0 to 255.255.255.255.

Whenever a device is connecting to a network, a unique IP is assigned to the device by Dynamic Host Configuration Protocol (DHCP). The overall goal of DHCP is to simplify the configuration and management of IP addresses allocated to devices. When a device is configured to dynamically learn its IP address from a DHCP server, three pieces of information are necessary: (a) IP address, (b) subnet mask, and (c) default gateway.

Configuring a DHCP server to hand out IP addresses on a subnet is known as a DHCP pool. This pool of addresses is usually a range of consecutive numbers within a single IP subnet. The subnet mask tells devices how large the subnetwork is that they are connected to. Finally, the default gateway is the IP address that signifies "the way out" of the subnetwork to which the device belongs. Connecting to a hotspot is as similar as connecting to a new network. Here the hotspot becomes the default gateway.

In the provided Pcap file, the default gateway is **10.42.0.1** and the subnet mask is 255.255.255.0 which means that IPs assigns to the connected devices can range from **10.42.0**.2 - **10.42.0**.255

Your task is to write a Python script which makes use of the dpkt library to calculate some basic statistics and analyze the given Pcap file. Along with the report, you will also need to submit the Python script you used to generate these statistics (20 points).

Report:

pcap_[name1]_[student_id1]_[name2]_[student_id2].pdf (30 points)

At the beginning of the page, specify the following:

1. Full Name of student 1 (Student ID) (Discussion Group)
2. Full name of student 2 (Student ID) (Discussion Group)
3. Name of the code and Pcap files submitted

Answer the following questions:

1. How many devices are connected to this hotspot? (4 points)
2. Which device sends out the most number of packets? (3 points)
3. Which device receives the most number of packets? (3 points)
4. Is there any endpoint where more than one device sends out a network packet to it? List the IP addresses of these endpoint(s)? (4 points)
5. Which application layer protocol has been used the most by the devices? (4 points)
6. Identify how much time did it take for us to capture this Pcap file (in minutes). (4 points)
7. Can you tell whether the devices send packets concurrently or sequentially[1]? Explain your approach to figure the sets of devices with concurrent/sequential network traffic. List the sets of devices with concurrent traffic. (4 points)
8. Can you figure out at approximately what point of time the devices were disconnected from the hotspot? (**hint**: Look at the final packets using **Wireshark**) (4 points)

Testing Environment:

All submissions will be tested on Python 3+.

Late Submission Policy:

No late submissions are allowed. However, if you barely miss the deadline, you can get partial points upto 24 hours. The percentage of points you will lose is given by the equation below. This will give you partial points up to 24 hours after the due date and penalizes you less if you narrowly miss the deadline.

$$Total\ Marks\ you\ get\ =\ (Actual\ Marks\ you\ would\ get\ if\ NOT\ late)\ \times \left[1\ -\ \frac{hours\ late}{24}\right]$$

Late Submissions (later than 24 hours from the due date) will result in zero points, *unless you have our prior permission or documented accommodation*.

[1] In concurrent browsing, devices use the network at the same time; however, in sequential browsing, the devices use the network one after the other.

Best of luck

Include this signed page along with your submission

Submission Page

I certify that all submitted work is my own work. I have completed all of the assignments on my own without assistance from others except as indicated by appropriate citation. I have read and understand the university policy on plagiarism and academic dishonesty. I further understand that official sanctions will be imposed if there is any evidence of academic dishonesty in this work. I certify that the above statements are true.

Team Member 1:

_____	_____	_____
Full Name (Printed)	Signature	Date

Team Member 2:

_____	_____	_____
Full Name (Printed)	Signature	Date