Assignment 1

Part 1- Wireshark

Wireshark is a free and open-source packet analyzer. It is used for network troubleshooting analysis, software and communication protocol development and education. Wireshark can capture packets in various modes. In promiscuous mode all packets within the same network are captured not just the ones destined to your computer. In this assignment we will use this tool to capture and analyze packets.

Getting Started:

Go to https://www.wireshark.org/download.html and download and install the relevant version of wireshark on your machine. After installation, first ensure there is no network activity on your machine. Then, go to Wireshark interface and start a packet capture.

- Set the wireshark capture filter to list HTTP protocol packet.
- Start a new live capture
- Open your browser and go to www.facebook.com
- Stop the live packet capture

Using the captured packets information, answer the following questions:

A) What location did you run this experiment from? (e.g. Home or some building on campus).

B) Using the captured information only; what is the IP address of your machine?

C) What is the IP address of the Sejrver hosting www.facebook.com? How did you find it from the captured data? A very brief answer is needed.

D) Do you also see other source and destination IP addresses in the packet trace? What do you think are these?

E) Will the captured packet trace show a different IP address for your machine if you change your location (to another building on campus or a different home location)?

F) Will the captured packet trace show a different IP address for www.facebook.com if you change your location (to another building on campus or a different home location)?

G) What are the source and destination hardware addresses (or physical addresses or MAC addresses) in the http GET request? Which one belongs to your machine? To which machine/device does the other belong?

H) What are the source and destination hardware addresses in the http response? Which one belongs to your machine?

I) What are the source and destination IP addresses in the http response? Which one belongs to your machine? To which machine/device does the other belong?

Hint: When you open a webpage from your browser, an HTTP GET request is sent from your machine to the server hosting that webpage. Upon receiving this request, the server sends a response (for example, an html file) back to you.

Part 2- Web Crawler

A web crawler is a program or automated script which browses the Web in a methodical, automated manner. Many legitimate sites, in particular search engines, use crawling as a means of providing up-to-date data. Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine, that will index the downloaded pages to provide fast searches. Crawlers can also be used for automating maintenance tasks on a Web site, such as checking links or validating HTML code.

In this part of the assignment you will be required to build a simple web crawler that recursively downloads the complete website when given a starting URL. To simplify things, you will only be required to crawl a static website.

To start off, you can generate a **GET** request using python, save the html file, extract the objects (You can look into the library *beautifulsoup* or any other library for this purpose). Repeat the same procedure on the saved objects. The goal is to save the entire website using these recursive steps.

Note:

1. This assignment is due at **11:59 pm on 4th February 2019** and has to be done individually.
2. All assignment submissions will be checked for plagiarism.
3. No deadline extensions will be given so please start early.
4. Good Luck, have fun and feel free to ask the TAs for any help.