Homework 2: HTTP, TCP, and Wireshark
CSE 534, Spring 2016
Instructor: Aruna Balasubramanian
Due date: 3/2/2016, 9.00pm

The last few weeks we have been talking about HTTP, TCP, and Congestion Control. The goal of this assignment is to dissect the HTTP and TCP protocols using the Wireshark tool.

To do this, you should be familiar with the HTTP header, the TCP header, and the HTTP/TCP packet format. It will also be useful to know the difference between HTTP 1.1 (that uses parallel and persistent HTTP) and HTTP 2 (that uses multiplexing).

Part A Wireshark Programming Task (40 points)

Your task is to write a program that can analyze a Wireshark trace to characterize the TCP flows in the trace. A TCP flow starts with a TCP "SYN" and ends at a TCP "FIN".

You need to use the pcap library to analyze the traces. Importantly, you need to write code to analyze the Wireshark trace in binary format. You cannot convert it into text and perform the analysis. This is important because the main goal of this homework is to learn how to parse network packets.

Use your program to analyze the file "http_first_sample.pcap".

(1) Draw the HTTP sequence between the client and the server for the program. Your HTTP sequence should look like those in Pages 12, 13, and 14 (whichever is appropriate) in lect-http-http2.pdf. I am looking for the values of Sequence number, Acknowledgment number, and the Window size for each transaction. You can either draw this by hand, or output this as part of your program (the drawing does not have to be pretty or exactly as I have presented in the class. It just needs to have all the necessary information).

(2) For the first 3 transactions after the TCP connection is set up (in both directions combined), explain the values of the Sequence number, Ack number, and Window size.

PART B (30 points)

Your next task is to use your program from Part A to parse the files HTTP_SampleA.pcap and HTTP_SampleB.pcap. Ignore any non-TCP packets. Answer the following questions

1. What kind of HTTP protocol are the two files using (HTTP 1.0, HTTP persistent connection, HTTP parallelization, HTTP pipelining, HTTP 2.0)? Explain your answer.

2. In each file, for each TCP connection in the file, estimate the:

- throughput
- goodput
- average round trip time
- the initial congestion window size.

All estimations need to be done programmatically. Show your work.

Part C (30 points) Next, use your program from Part A to parse the files HTTP_Sample_Big_Packet.pcap .

(1) Compute the congestion window size for each RTT for the first 20 RTTs after the TCP handshake. Visualize the congestion window size across time. The x axis is the # of RTTs, the y axis is the congestion window size.

(2) After the TCP connection has been established, compute the first 3 retransmission timeout values estimated at the HTTP client. You need to follow RFC 6298 to estimate the retransmission timeout (https://tools.ietf.org/html/rfc6298). Assume G = 1 second.

As before, you may write your programs in the following languages: Python, Ruby, Java, C/C++, or Perl. If you use any other language, please talk to me.

Note that viewing these traces on Wireshark is helpful, but may not always be completely accurate. This is because Wireshark may sometimes parse HTTP/2 packets incorrectly.

Submission instruction

You need to submit your homework in a single zip file as follows:

- The zip file and (the root folder inside) should be named using your last name, first name, and the homework name, all separated by a dash ('-')
  e.g. lastname-firstname-HW2.zip
- The zip file should contain your code for Part A, Part B, and Part C.
- Include the expected output file for each part.
- Include the text file containing the answers for Part A, Part B, and Part C if applicable.
- You should provide a README.txt file describing:
  - the high level view of your design.
  - how to run your programs.