Ryan Slyter
CS455
Written Assignment #3

All of my scripts included with this write-up are labeled 'prob1.sh', 'prob2.sh', etc...

1. Shell output:
614867.801699899253 average packet size for trace.

I computed this by simply adding up the total bytes (bytes*packets per occurrence) and total packets and modifying the example script that came with the assignment prompt so that the final awk part printed total_bytes / total_packets. Note that you have to set a delimiter in awk now to get the two separate fields.

2. I output my calculations to a text file and then plotted the 2 columns (reversed) from it in Gnuplot. Results 1 and 2 ("image1.png"..) are the flow duration graphs (linear, log scales repectively) and results 3 and 4 are for flow sizes. Cumulative probability distribution plots tell you for the trace, the probability of a given flow duration for a line and flow size, and of those flow durations and flow sizes smaller than it. I don't know why I have a vertical asymptote in my flow sizes part where we were just calculations the occurrences of a given flow size. The log-log scales give you a better idea of how the probability of a given flow size, for example, trends up or down, because it takes into account vast magnitudes between points.

3. These are the output tables I copied from my terminal:
srcport percOfBytes
80 3.76168e-08
443 3.7232e-08
53 4.84782e-09
0 3.49269e-08
25 1.76559e-09
22 3.71226e-09
1935 6.01467e-06
3074 3.75753e-09
3389 3.45002e-08
2128 2.58047e-08

dstport percOfBytes
80 2.34385e-07
443 3.73112e-09
53 8.49594e-09
445 8.04826e-11
25 5.71664e-06
123 2.58047e-09
1935 2.80419e-08
3074 3.75753e-09
2048 1.40341e-09
0 3.49269e-08

NOTE: By 'traffic' I thought you meant occurrences (lines) of the file containing a specific port

number or value we are targeting. So by 'total traffic' I summed all occurrences of all ports. Port 80 and and port 0 make sense to be in the top 10 since they represent common server ports are with zero you are stating “any port” when you try to bind a socket. Another port I looked at was port 443, which is supposed to be a port to transfer very sensitive data. This makes sense since WSU would need to safeguard many of its data transfers for being a public entity. Port 25 looks to be for mail transfers, which again make sense with respect to a larger university. Finally, port 53 is used commonly because of DNS lookup.

4. These are the results I got running my script:
873733 <-**I just printed total traffic first**
3375 <-**I also printed the total number of different prefixes**
0.92903210 of total traffic comes from top 10 percent of prefixes.
0.71629205 of total traffic comes from top 1 percent of prefixes.
0.33480365 of total traffic comes from top 0.1 percent of prefixes.

Note that again I took “traffic” to mean occurrences of that prefix with respect to all occurrences in the trace.

5. This is what I get for calculating WSU's prefix (34425) for both percentage of total packets in the trace and percentage of total bytes and then running my script to find % of outgoing and incoming packets and total bytes involving WSU:

0.000037630059 percent of packets in the trace are outgoing from WSU.
0.000000050351 percent of total bytes in the trace are outgoing from WSU.
0.000205934363 percent of packets in the trace are incoming from WSU.
0.000015134273 percent of total bytes in the trace are incoming from WSU.