High Level View

I have used the dpkt library to parse the tcp data, which gives us a timestamp and a buffer.

There is a Packet class in the code which is used for packet parsing from binary to a human readable format using from_bytes. The object created using this class has the information about the packet. I have also used struct to extract the request and response for the packet.

I have also used a class named HTTP_Reassembler. This class is used to encapsulate a HTTP request response and all the TCP segments in between if they exist

There are various functions used in the code.

1. PACP_File_Reader is used for extracting a list of packets from a parsed pcap file
2. Flow_Initializer is used to segregate all the packets into different flows as well as counting the number of tcp connections, packet count and total payload
3. http_reassemble is used to reassemble the packet list into GET request packets and HTTP response packets

This code also later calculates the number of TCP flows for all pcap files. These pcap files were generated using the following commands

sudo tcpdump -i en0 -n port 1080 -w http_1080.pcap
Enter this in command line and open the website - http://www.sbunetsyslabs.com:1080 in browser

sudo tcpdump -i en0 -n port 1081 -w http_1081.pcap
Enter this in command line and open the website - https://www.sbunetsyslabs.com:1081 in browser

sudo tcpdump -i en0 -n port 1082 -w http_1082.pcap
Enter this in command line and open the website - https://www.sbunetsyslabs.com:1082 in browser

Moreover, this code also calculates the total time taken by each website (each port), total packets transferred and the raw data size for the sender to receiver port.