FTP Commands:

1. DATA PORT (PORT):

The argument is a HOST-PORT specification for the data port to be used in data connection.

This address information is broken into 8-bit fields and the value of each field is transmitted as a decimal number (in character string representation). The fields are separated by commas.

Example: PORT h1,h2,h3,h4,p1,p2 | where h1 is the high order 8 bits of the internet host address.

2.SITE PARAMETERS (SITE)

This command is used by the server to provide services

specific to his system that are essential to file transfer

but not sufficiently universal to be included as commands in

the protocol.

3. HELP

4. CWD Command: Changing the directory

5. PASS: Request for Password

6. Response Codes:

150 File status okay; about to open data connection.

250 Requested file action okay, completed.

257 "PATHNAME" created.

350 Requested file action pending further information.

450 Requested file action not taken.

File unavailable (e.g., file busy).

550 Requested action not taken.

File unavailable (e.g., file not found, no access).

451 Requested action aborted. Local error in processing.

551 Requested action aborted. Page type unknown.

452 Requested action not taken.

Insufficient storage space in system.

552 Requested file action aborted.

Exceeded storage allocation (for current directory or

dataset).

553 Requested action not taken.

File name not allowed.

Questions

1. How many unique server IPs do you see? How many unique client IPs?
   1. Done By checking all SYN packets, put them in a file and then checking unique server and client from that.

Day 1 – Client IPs – 522

Server IPs – 45

Day 2 – Client IPs – 939

Server IPs – 50

Day 3 – Client IPs – 510

Server IPs – 89

1. How many unique TCP flows do you see? Hint: A TCP flow is
   1. Got all lines having TCP ACK, and then identified flow uniquely by its source IP, destination IP, source port, and destination port.

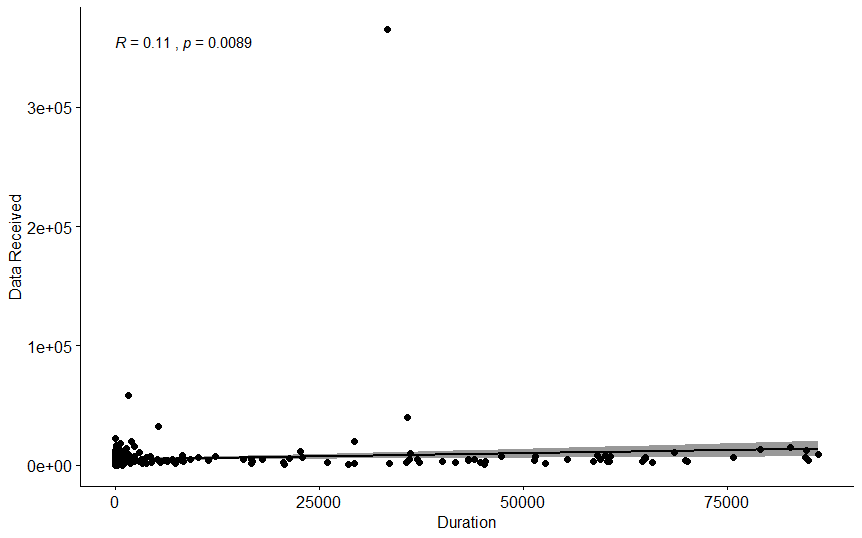
|  |  |  |  |
| --- | --- | --- | --- |
|  | lbnl.anon-ftp.03-01-11.csv | lbnl.anon-ftp.03-01-14.csv | lbnl.anon-ftp.03-01-18.csv |
| No. of clients | 522 | 939 | 510 |
| No. of servers | 45 | 50 | 89 |

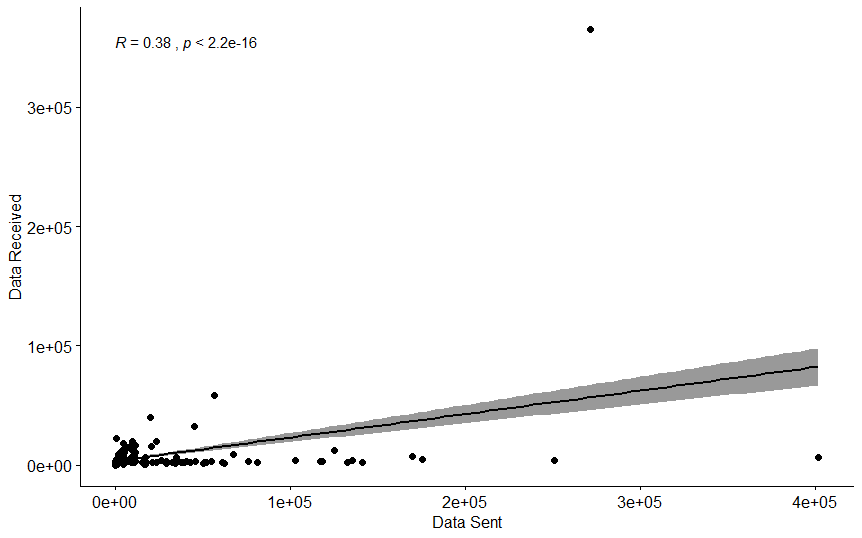
1. Draw a plot of the number of connections (TCP flows) opened to any FTP server within 60min windows over the 24-hour duration, ie. within hours 00:00-01:00, 01:00-02:00, 02:00-03:00, etc. How can you use these traffic profiles to detect if the system is under a DoS attack?

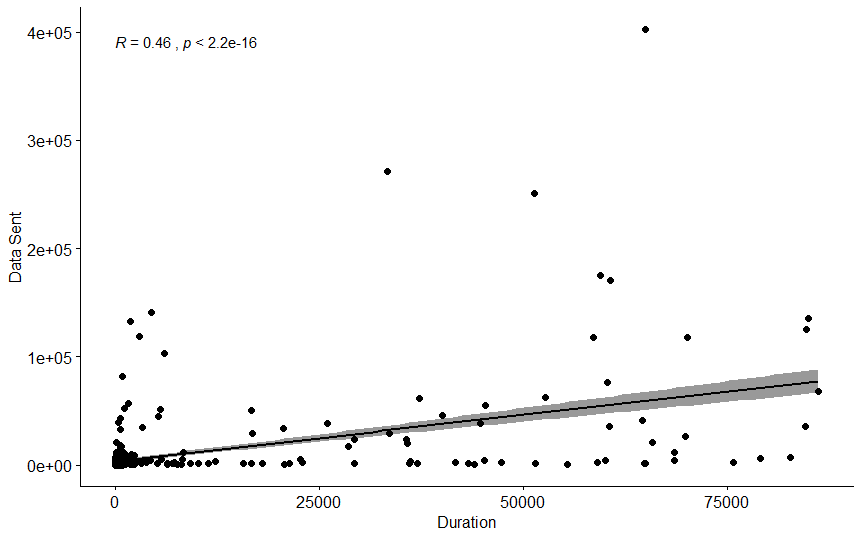
DdoS attack can be detected by learning the fluctuations over no. Of connections in the time intervals, using the these profile histogram data. Since many-a-times, a fluctuation is reasonable for example, in the night. Thus we can use machine learning to model our average profiling every day, and if the error in predicated and actual traffic is above a certain threshold, we declare a DdoS attack.

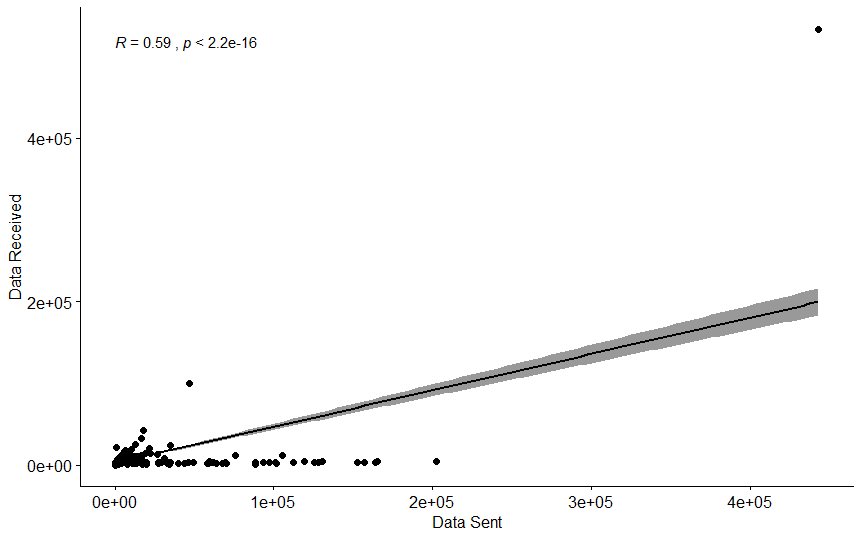
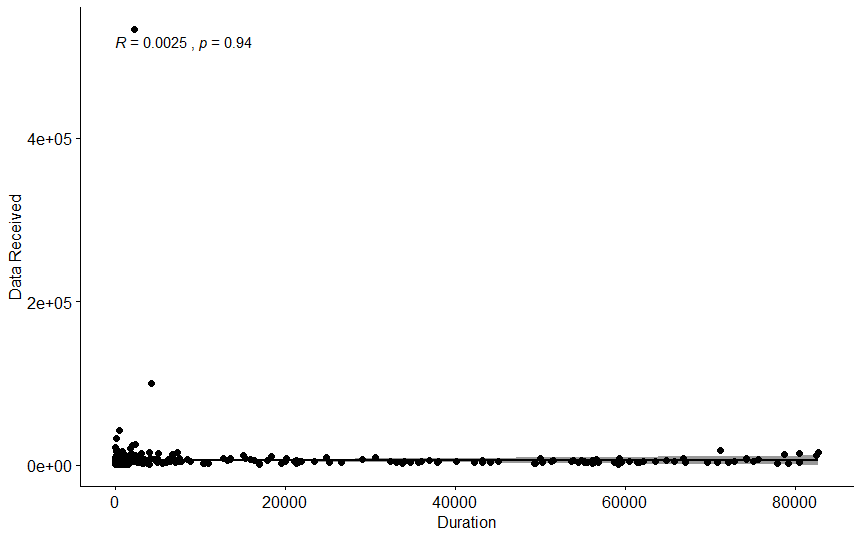
1. For all the connections, find the duration over which a connection was kept open, and plot the CDF (cumulative distribution function) of these connection durations.
   1. Notice how most connections are of a short duration. Why do you think so?
      1. The control connection in FTP is persistent but the data connection is created and disposed after each data transfer (aka. Non-persistent). So that end of file equals end of stream, so that there doesn't have to be a superimposed protocol on the data connection.
2. On the same lines as above, for all the connections find the number of bytes sent and received over the connection, and check if there is a correlation between the connection duration (as computed in question 4 above, ie. the duration over which the connection was kept open) and the number of bytes sent to the servers. Similarly check if there is a correlation between the number of bytes sent to the servers and number of bytes received from the servers.
   1. Do you find that there is a correlation? Does the scatterplot help you identify any outliers? If you eliminate the outliers, does your correlation improve?

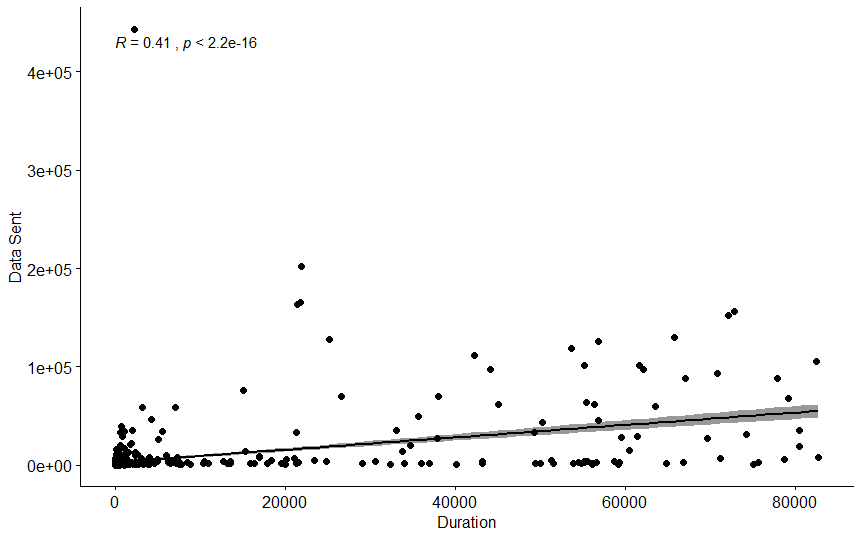
Yes, correlation was found. More the data to be recieved, more the connection period. The scatterplot helps to identify the outliers. Removing outliers will be much of an improvement here. Thus square fitting will give much of a different answer.

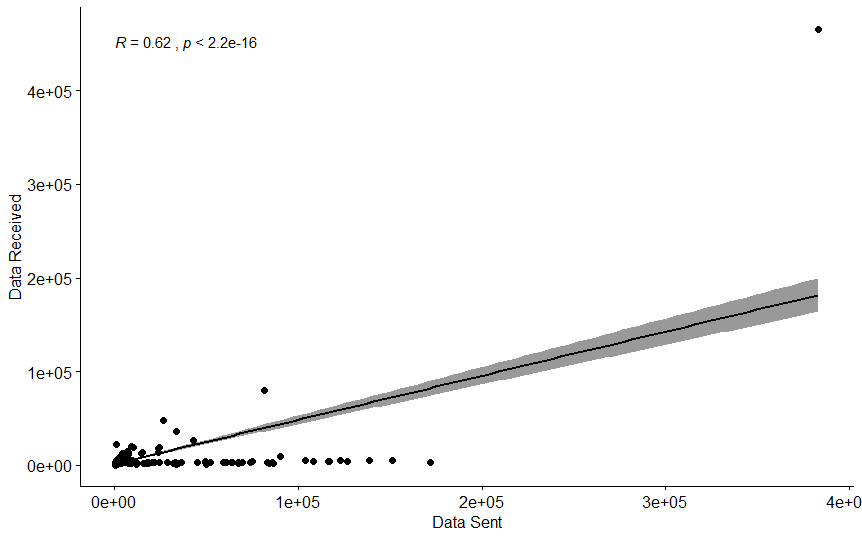
Day1: Plots

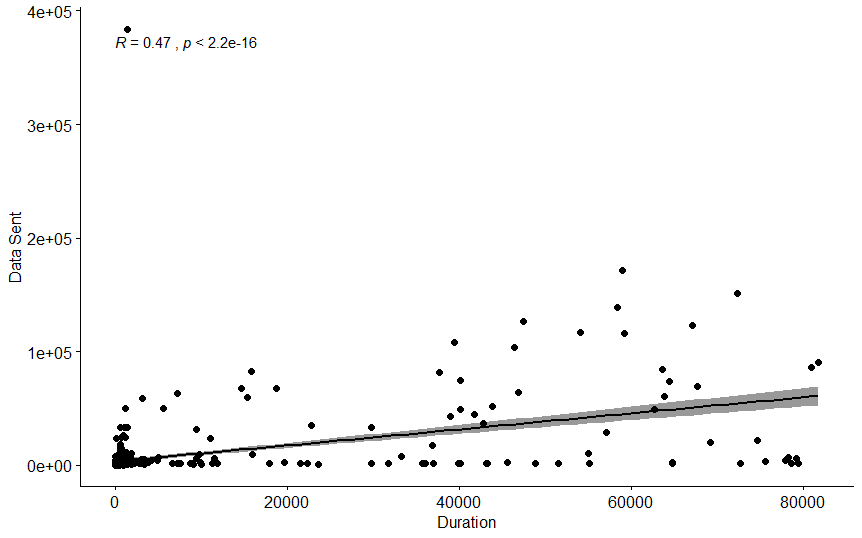
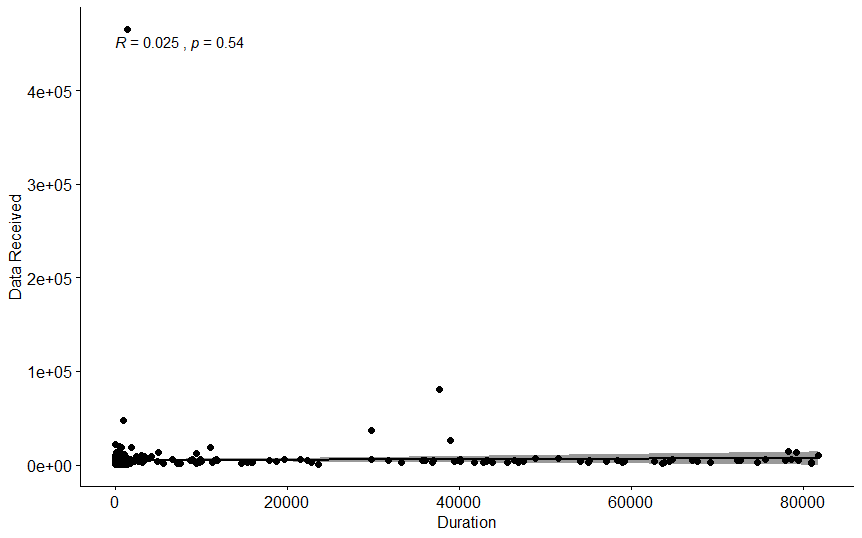




Day2: Plots



Day3: Plots



7. Most interarrival times are short, but the range is quite wide between the maximum and minimum interarrival times. Why do you think this is the case?

ANS: This is because there might be many clients requesting to server at same time and their packets are queued. Thus interarrival time is very low. To resolve this packet size is reduced by the client(as observed).

8. Does the CDF appear to be clustered around a few values? Why do you think this is the case?

ANS: Evidently, CDF appears to be clustered around a few values. This is because all the clients get the info that queueing delay is rising or decreasing. Thus every client decrease or increase the packet size accordingly.