Packet Analysis

Sukriti Gupta(2016CS50084) Chinmay Rai(2016CS50615)

1 FTP Codes and Response Codes

We observed the following FTP response and command codes:

1.1 Commands

- USER : Authentication username.
- PASS : Authentication password.
- OPTS : Select options for a feature
- SYST : Return system type.
- PWD : Print working directory. Returns the current directory of the host.
- NOOP : No operation (dummy packet; used mostly on keepalives).
- CWD : Change working directory.
- LIST : Returns information of a file or directory.
- PORT : Specifies an address and port to which the server should connect.
- QUIT : Disconnect.
- FEAT : Get the feature list implemented by the server.
- TYPE : Sets the transfer mode (ASCII/Binary).
- PASV : Enter passive mode.
- SIZE : Return the size of a file.
- RETR : Retrieve a copy of the file
- STAT : Returns information on the server status, including the status of the current connection
- REST : Restart transfer from the specified point.

- MDTM : Return the last-modified time of a specified file.
- ABOR : Abort an active file transfer.
- ALLO : Allocate sufficient disk space to receive a file.
- CDUP : Change to Parent Directory.
- EPRT : Specifies an extended address and port to which the server should connect.
- EPSV : Enter extended passive mode.
- HELP : Returns usage documentation on a command if specified, else a general help document is returned.
- SITE : Sends site specific commands to remote server
- MODE B,S : Sets the transfer mode (Stream, Block, or Compressed).
- NLST : Returns a list of file names in a specified directory.
- STOR : Accept the data and to store the data as a file at the server site
- STRU : Store file uniquely.

1.2 Responses

- 125 : Data connection already open; transfer starting.
- 150 : File status okay; about to open data connection
- 200 : The requested action has been successfully completed.
- 202 : Command not implemented, superfluous at this site.
- 211 : System status, or system help reply.
- 213 : File status.
- 214 : Help message.
- 215 : NAME system type. Where NAME is an official system name from the registry kept by IANA.
- 220 : Service ready for new user.
- 221 : Service closing control connection.
- 225 : Data connection open; no transfer in progress.
- 226 : Closing data connection. Requested file action successful

- 227 : Entering Passive Mode
- 230 : User logged in, proceed. Logged out if appropriate.
- 250 : Requested file action okay, completed.
- 257 : "PATHNAME" created.
- 331 : User name okay, need password.
- 350 : Requested file action pending further information
- 400 : The command was not accepted and the requested action did not take place, but the error condition is temporary and the action may be requested again.
- 421 : Service not available, closing control connection.
- 425 : Can't open data connection.
- 426 : Connection closed; transfer aborted.
- 451 : Requested action aborted. Local error in processing.
- 500 : Syntax error, command unrecognized and the requested action did not take place.
- 501 : Syntax error in parameters or arguments.
- 502 : Command not implemented.
- 504 : Command not implemented for that parameter.
- 530 : Not logged in.
- 550 : Requested action not taken. File unavailable
- 553 : Requested action not taken. File name not allowed.

2 Packet Trace Analysis

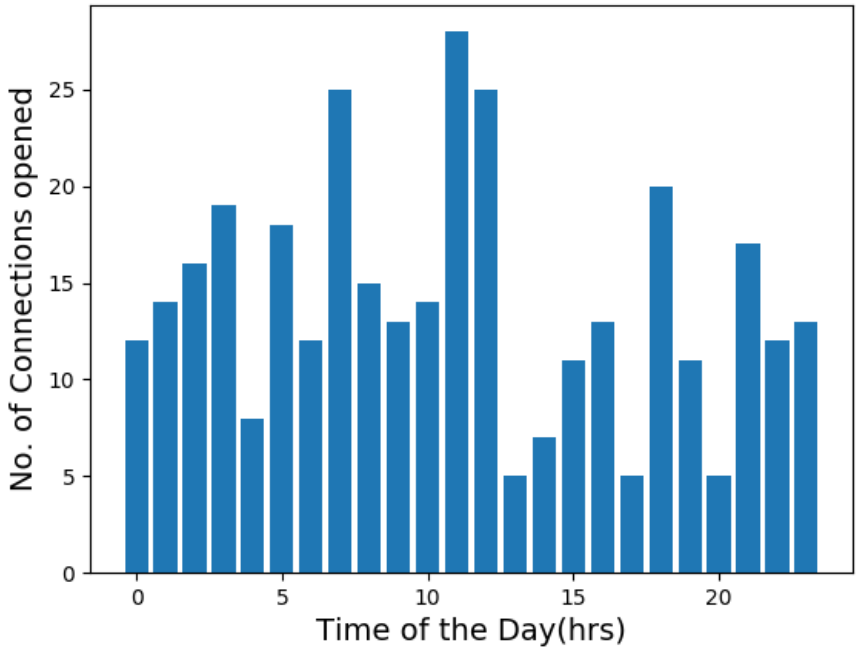
2.1 P1 & P2

In order to determine the No. of unique server and client IP's we have the counted looked into the unique addresses in the set if SYN Packets. Uniqueness of TCP flow is defined as uniqueness of the 4 tuple.

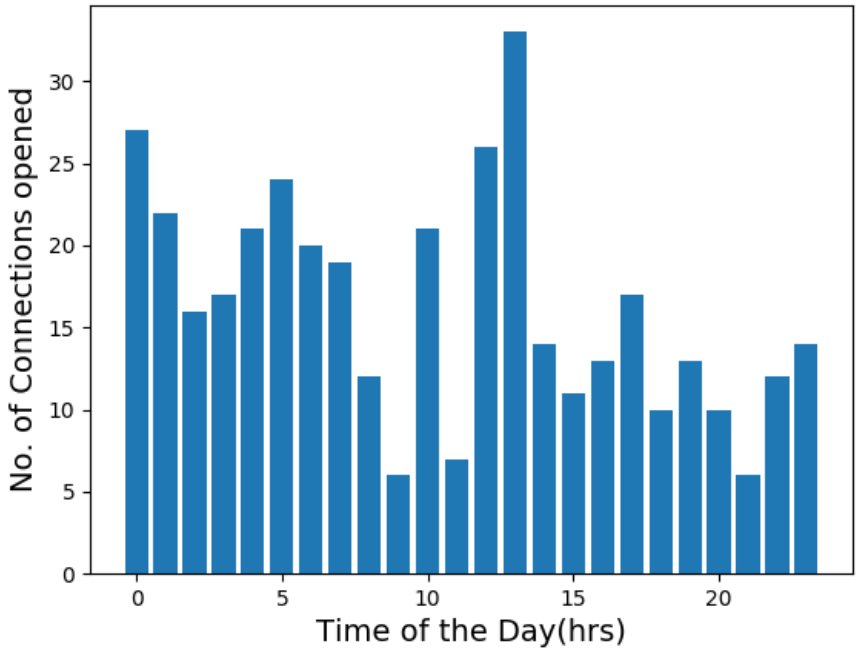
Question :	PART 1		PART 2
Day 1	#Servers: 45	#Clients: 526	#Flows: 3256
Day 2	#Servers: 50	#Clients: 945	#Flows: 5422
Day 3	#Servers: 89	#Clients: 519	#Flows: 3280

2.2 Daily Profile: P3

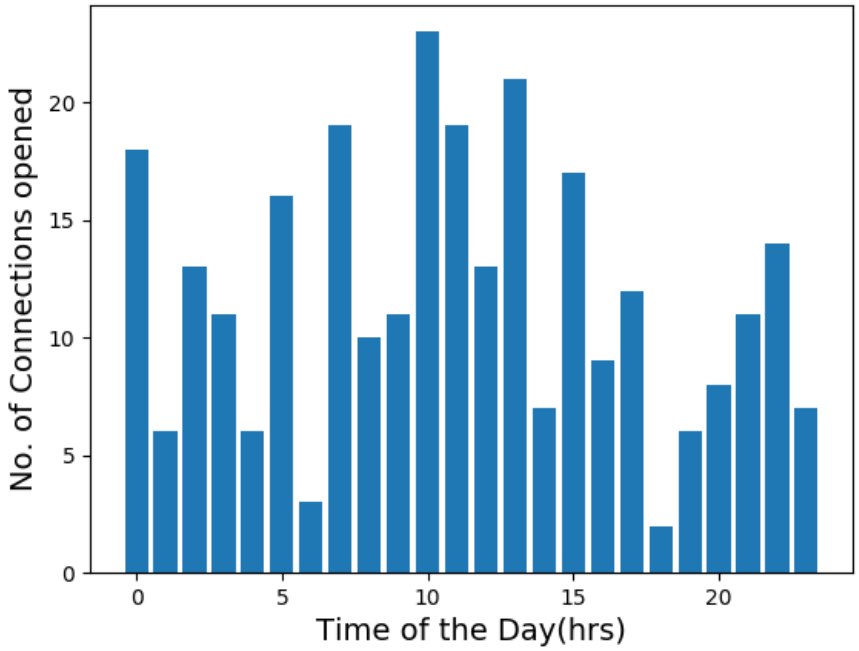
Done for server:131.243.2.12



Day 1



Day 2

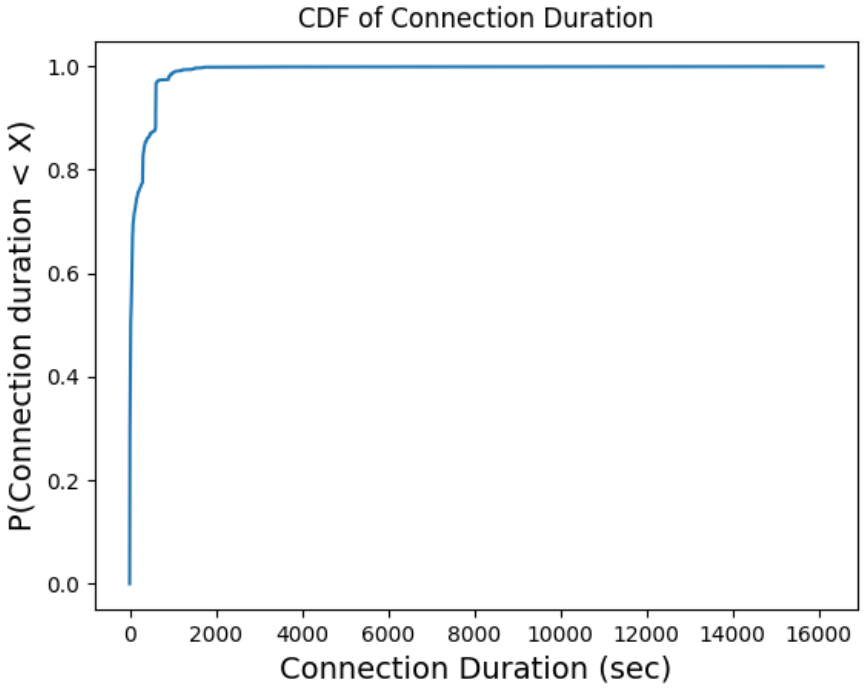


Day 3

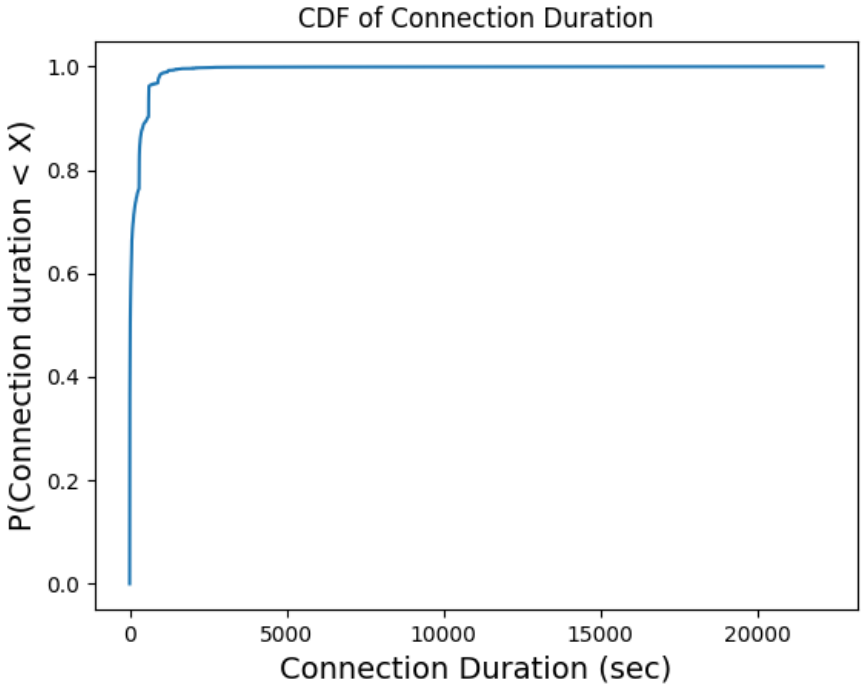
Anomaly detection on this Daily profile can help us identify the DOS Attack. The Characteristic of DoS attack is expected to be a very sharp peak in the Profile.

2.3 Connection Duration: P4

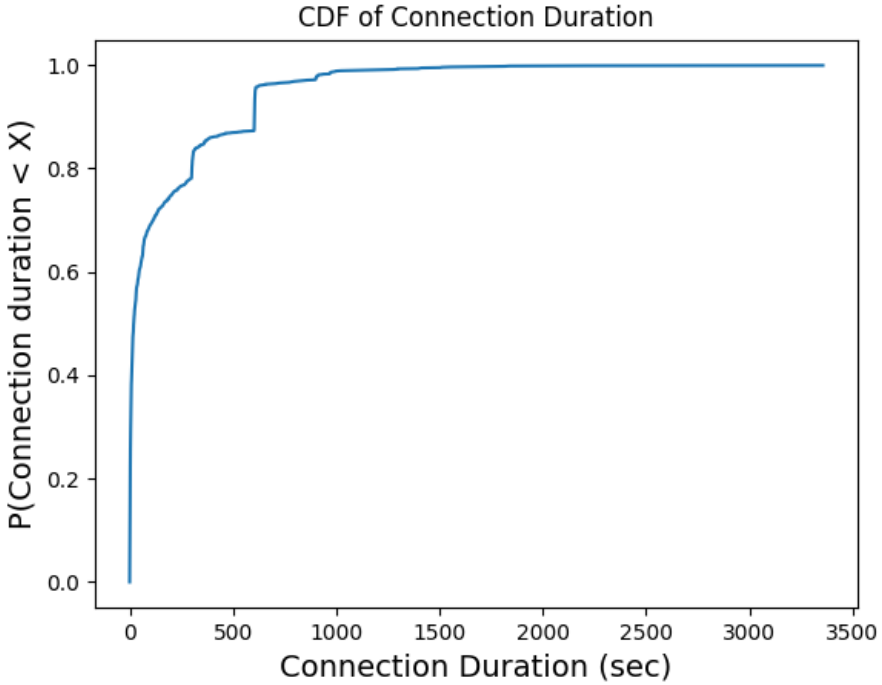
For Day 1, Mean = 160.02790 sec Median = 24.12750 sec



For Day 2, Mean = 166.68434 sec Median = 18.58785 sec



For Day 3, Mean = 153.98576 sec Median = 19.50087 sec



We see that most of the connection are of short duration because most of them are control instructions, which are probably established over a non-persistent HTTP setting, where the connection is closed after one transmission and a new one is opened for the next transmission.

2.4 Data transfer Vs. Connection Duration: P5

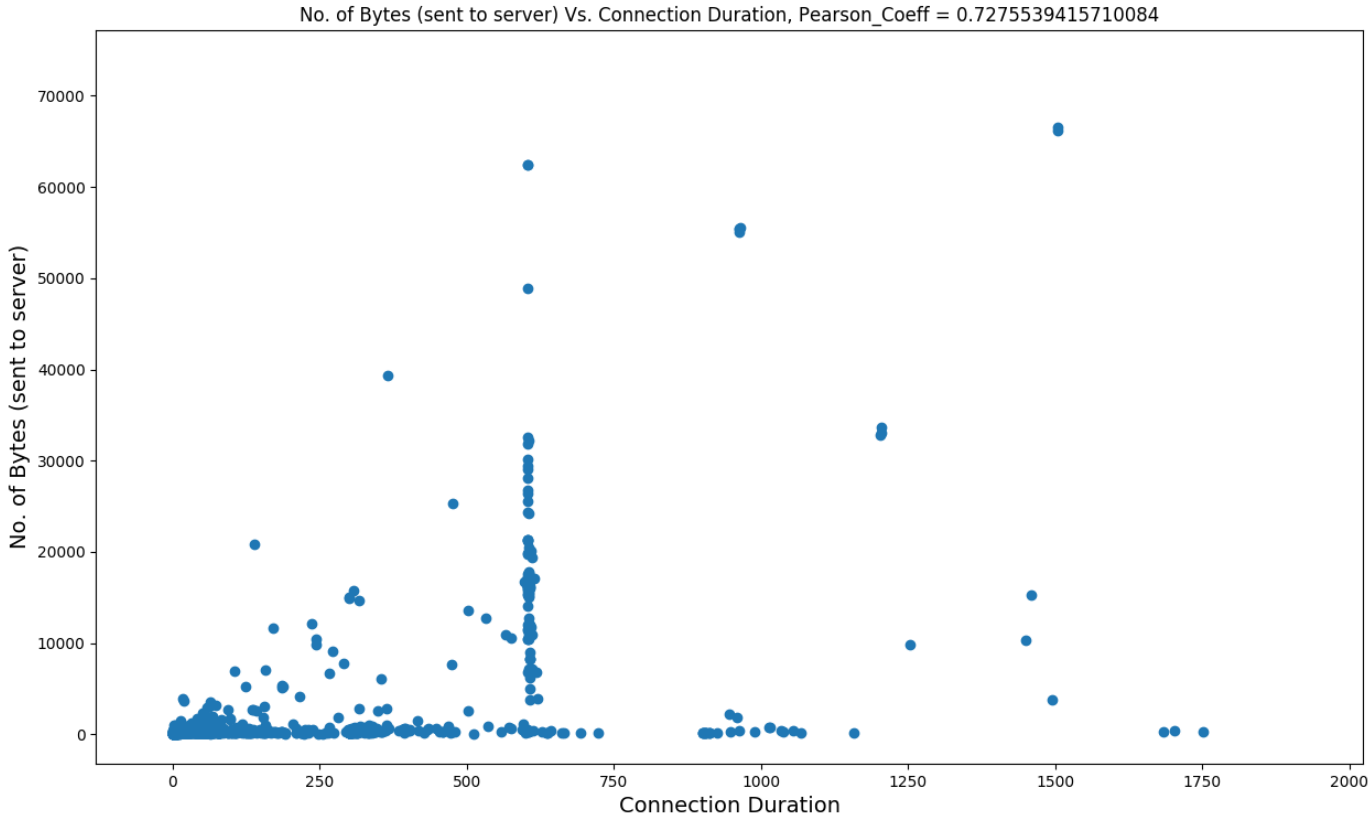
We have tried to correlate the :

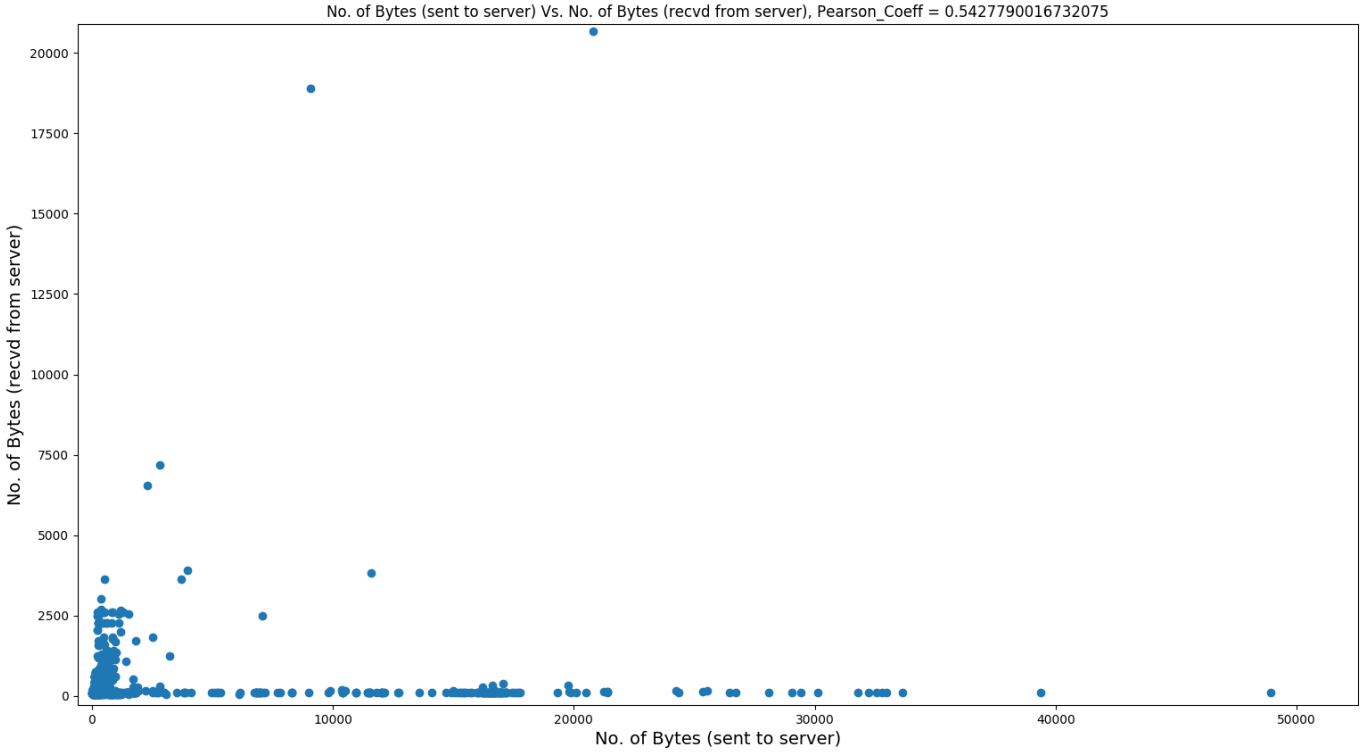
- No. of Bytes(Sent to server) Vs. Connection Duration
- No. of Bytes(Sent to server) Vs. No. of Bytes(Received from server)

for the 3 days of data that we were provided.
They correlation that we observe is rather weak which is indicated by the Pearson correlation coefficient. The Scatter plot also helps us identify the data points which don't conform to the liner trend. For eg: There is host of connections which terminate at approx. 625 sec irrespective of the data exchanged in those connections

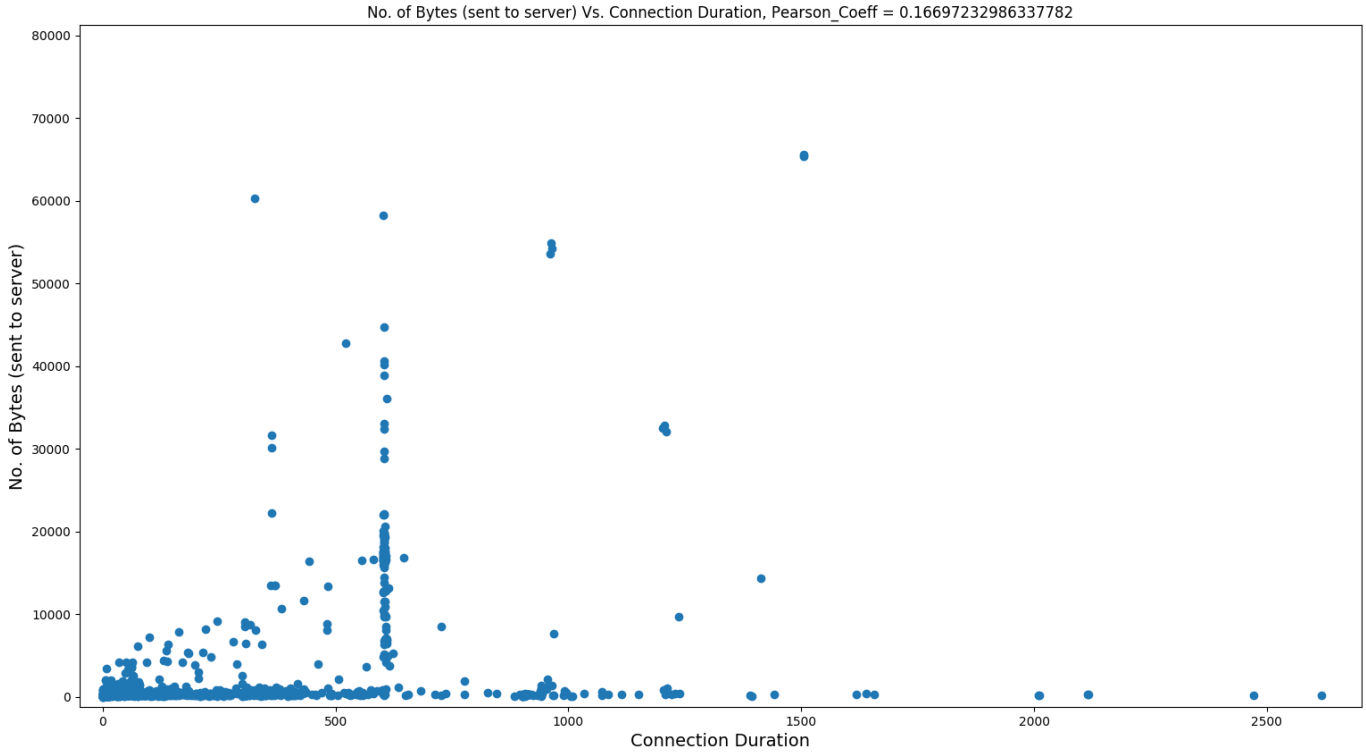
The results(in order of date on file) are:

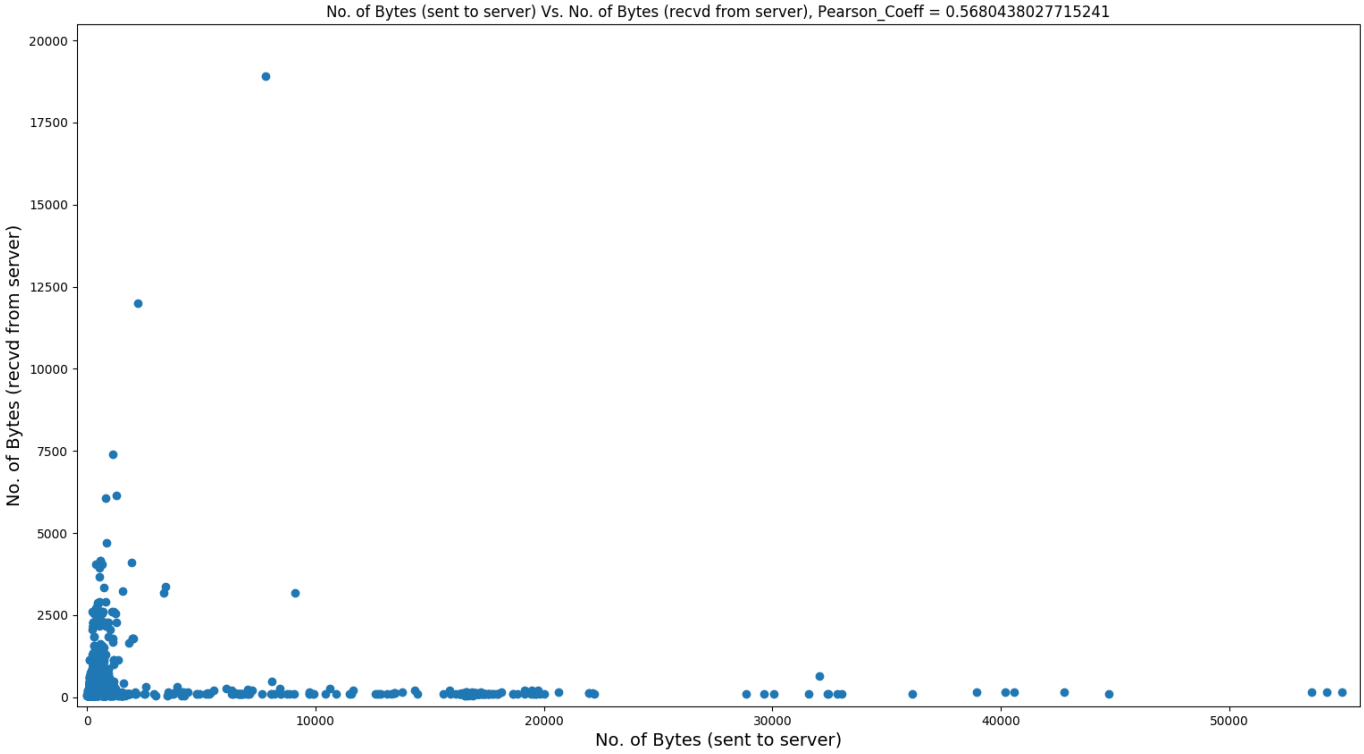
Day 1:


No. of Bytes (sent to server) Vs. Connection Duration, Pearson_Coeff = 0.7275539415710084
No. of Bytes (sent to server)
Connection Duration
0
10000
20000
30000
40000
50000
60000
70000
0
250
500
750
1000
1250
1500
1750
2000

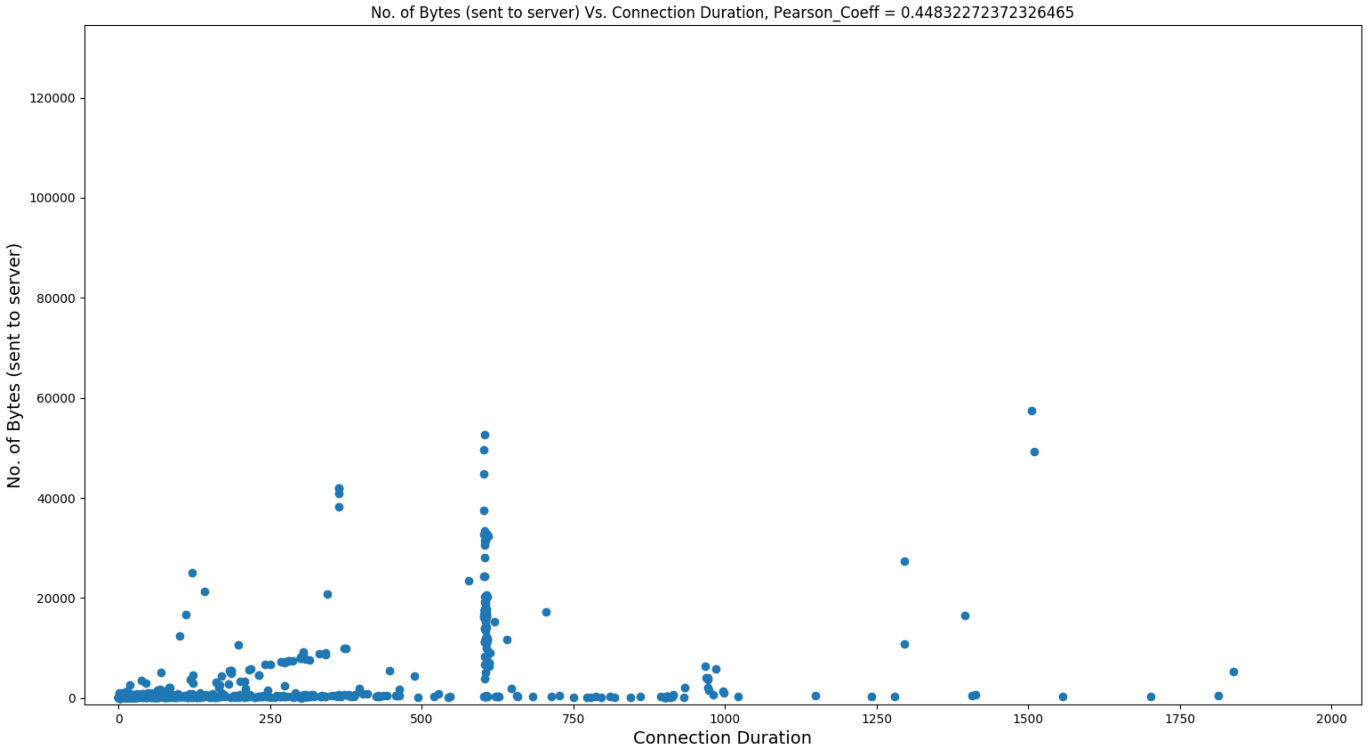

No. of Bytes (sent to server) Vs. No. of Bytes (recvd from server), Pearson_Coeff = 0.5427790016732075
No. of Bytes (recvd from server)
No. of Bytes (sent to server)
0
2500
5000
7500
10000
12500
15000
17500
20000
0
10000
20000
30000
40000
50000

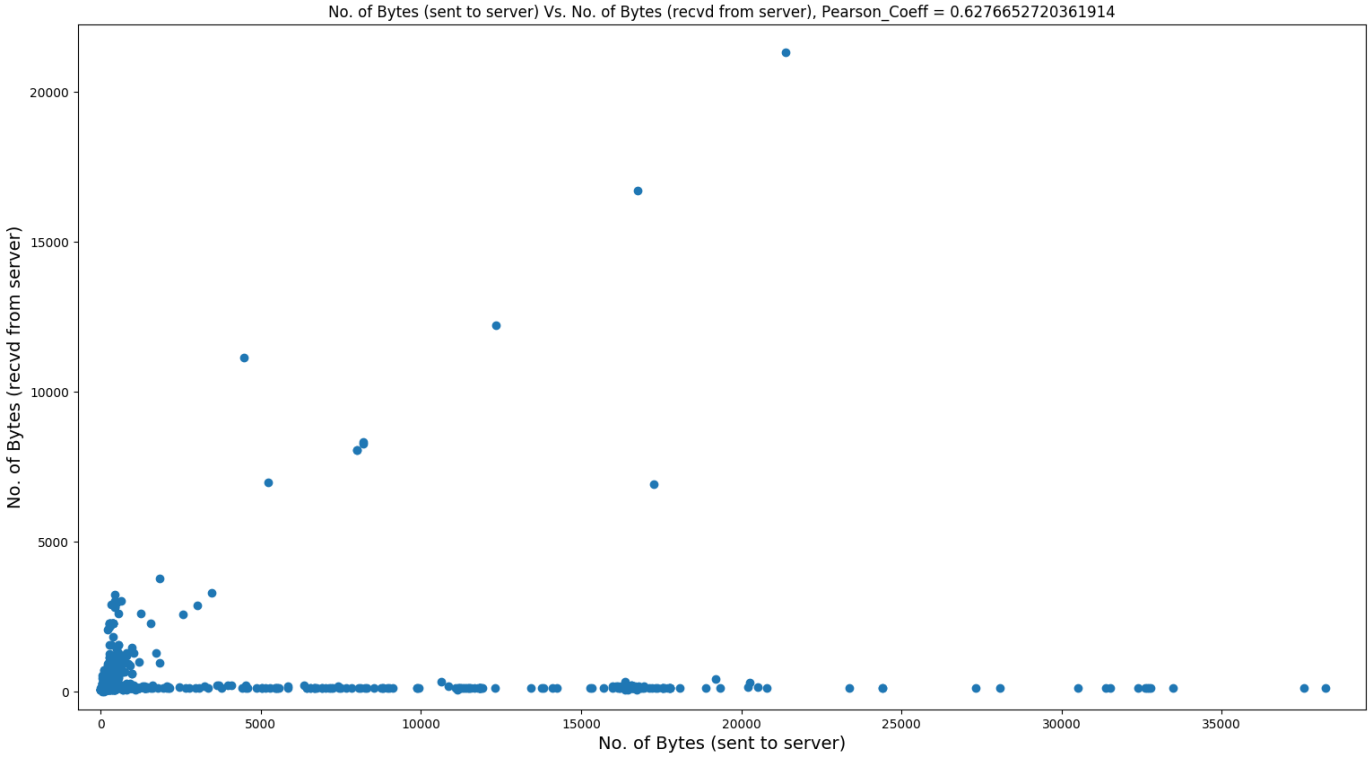
Day 2:


No. of Bytes (sent to server) Vs. Connection Duration, Pearson_Coeff = 0.16697232986337782
No. of Bytes (sent to server)
Connection Duration
0
10000
20000
30000
40000
50000
60000
70000
80000
0
500
1000
1500
2000
2500

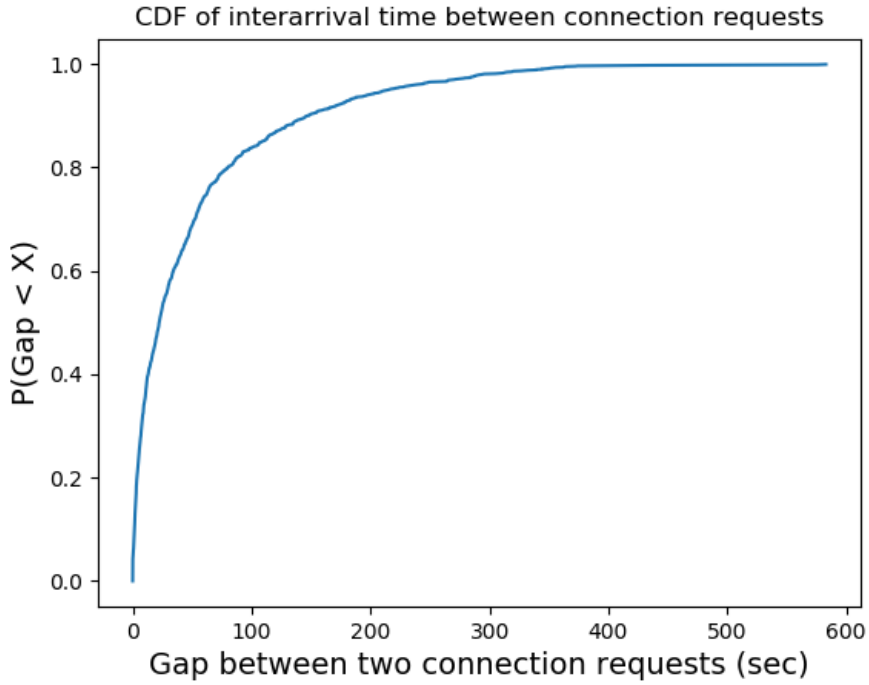

No. of Bytes (sent to server) Vs. No. of Bytes (recvd from server), Pearson_Coeff = 0.5680438027715241
No. of Bytes (recvd from server)
No. of Bytes (sent to server)
0
2500
5000
7500
10000
12500
15000
17500
20000
0
10000
20000
30000
40000
50000

Day 3:


No. of Bytes (sent to server) Vs. Connection Duration, Pearson_Coeff = 0.44832272372326465
No. of Bytes (sent to server)
Connection Duration
0
20000
40000
60000
80000
100000
120000
0
250
500
750
1000
1250
1500
1750
2000

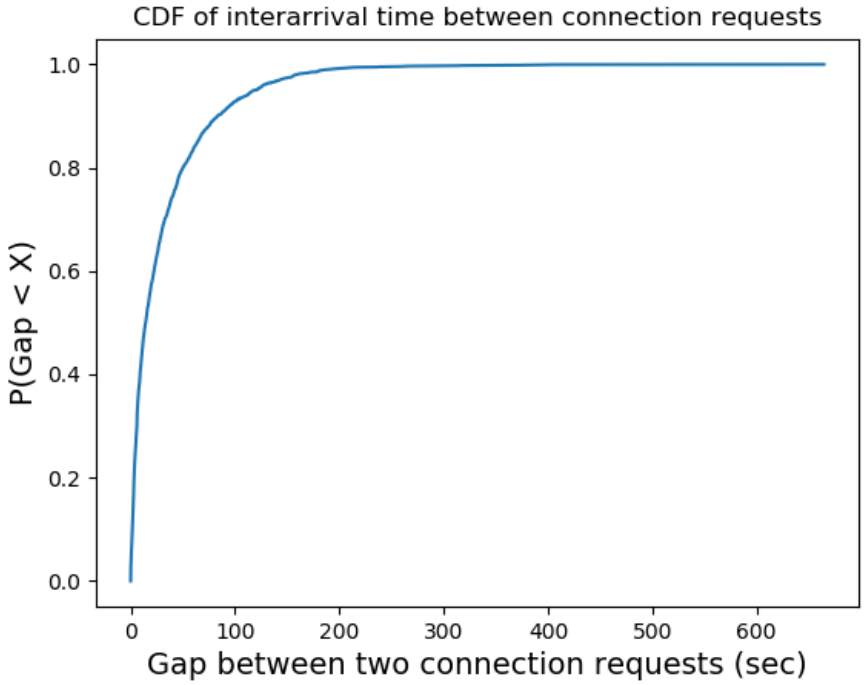

No. of Bytes (sent to server) Vs. No. of Bytes (recvd from server), Pearson_Coeff = 0.6276652720361914
No. of Bytes (recvd from server)
No. of Bytes (sent to server)
0
5000
10000
15000
20000
0
5000
10000
15000
20000
25000
30000
35000

2.5 CDF of Inter-Arrival times b/w opening up connections: P6



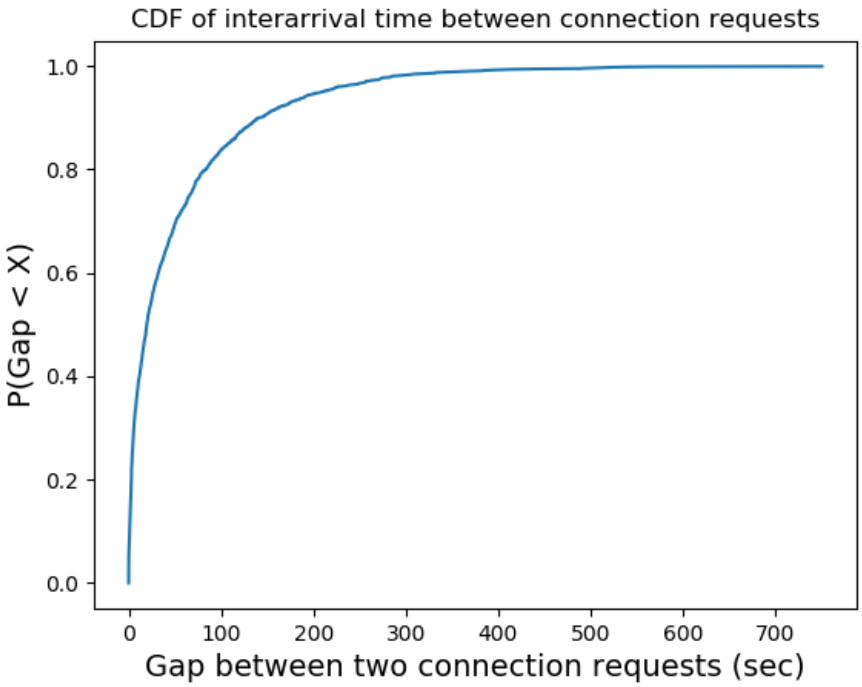
Day 1

Mean of Inter-arrival Time: 51.6563137835 Median of Inter-arrival Time: 22.2736705



Day 2

Mean of Inter-arrival Time: 31.3369680945 Median of Inter-arrival Time: 14.351235

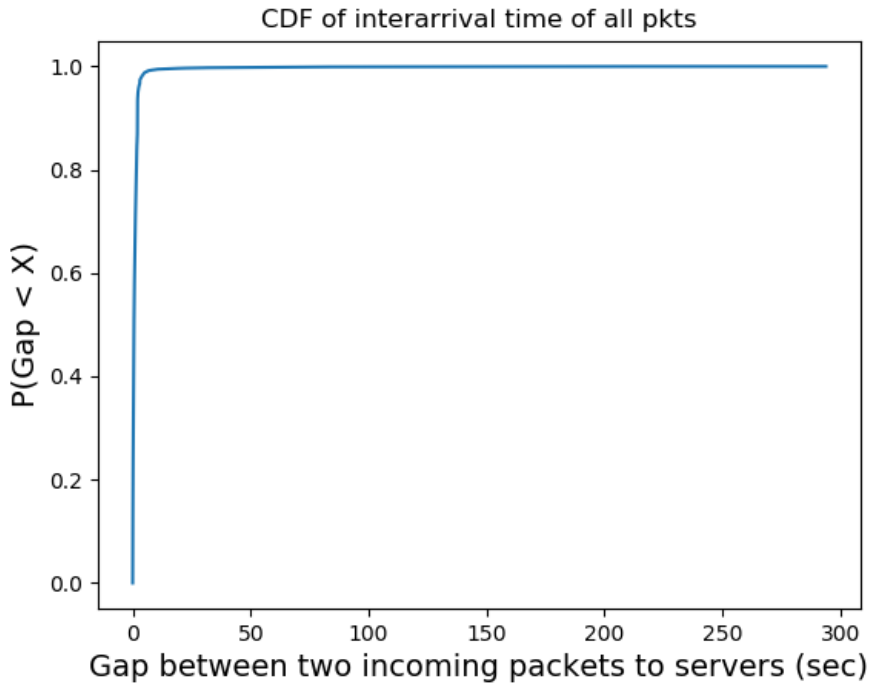


Day 3

Mean of Inter-arrival Time: 50.9182253282 Median of Inter-arrival Time: 19.582489

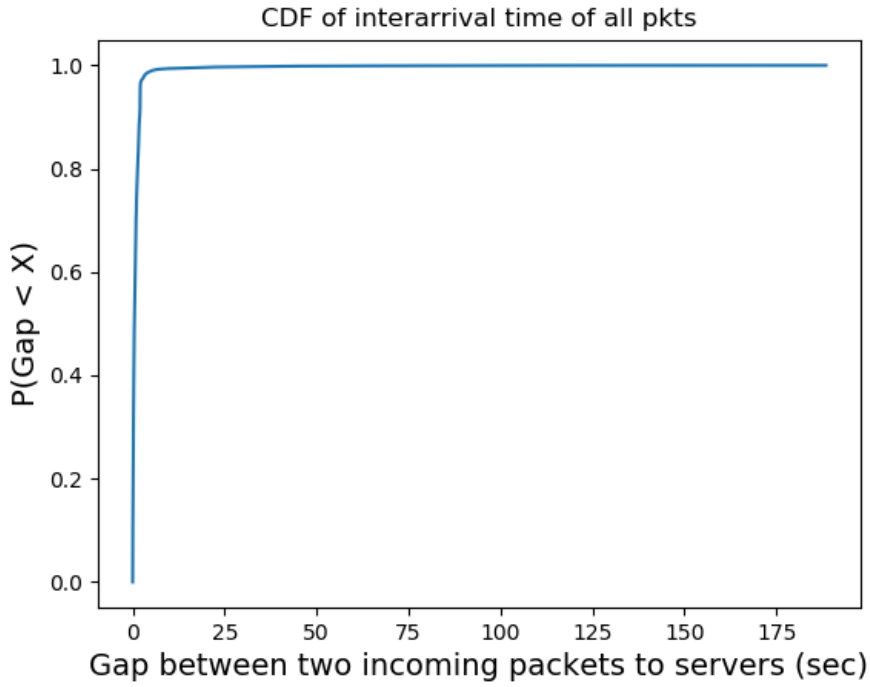
2.6 CDF of Inter-arrival time of Incoming packets: P7

The system is memory-less . Arrival of one connection request is independent of the other and most likely follows exponential distribution. In general, servers would be receiving requests without too much gap but some times there would be large gaps also. This would not happen a lot though. Such a distribution leads to comparatively larger means but very small medians.



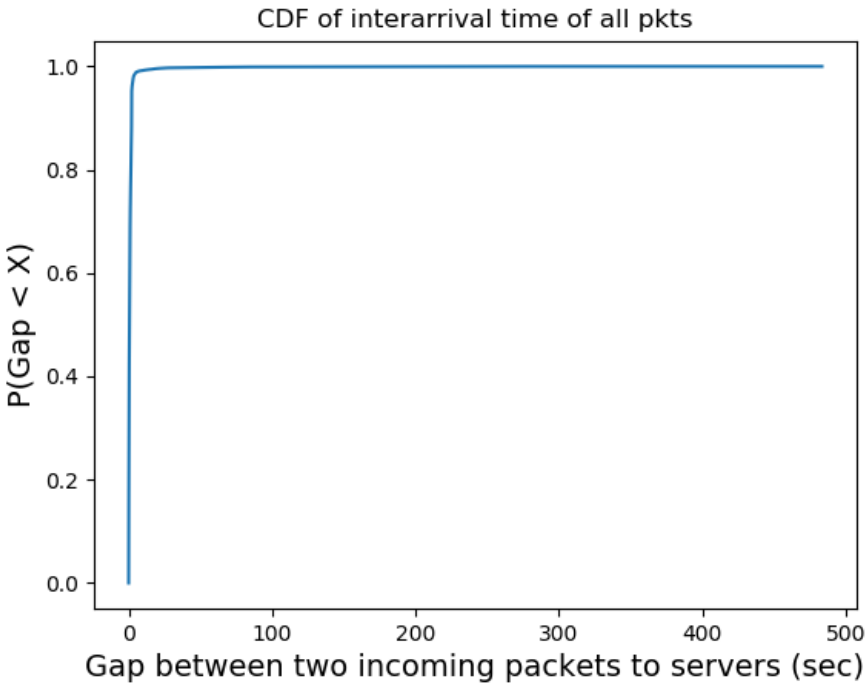
Day 1

Mean of Inter-arrival Time incoming packets to server: 1.132866375 Median of Inter-arrival Time incoming packets to server: 0.550200000



Day2

Mean of Inter-arrival Time incoming packets to server: 0.947197669 Median of Inter-arrival Time incoming packets to server: 0.4576430000



Day 3

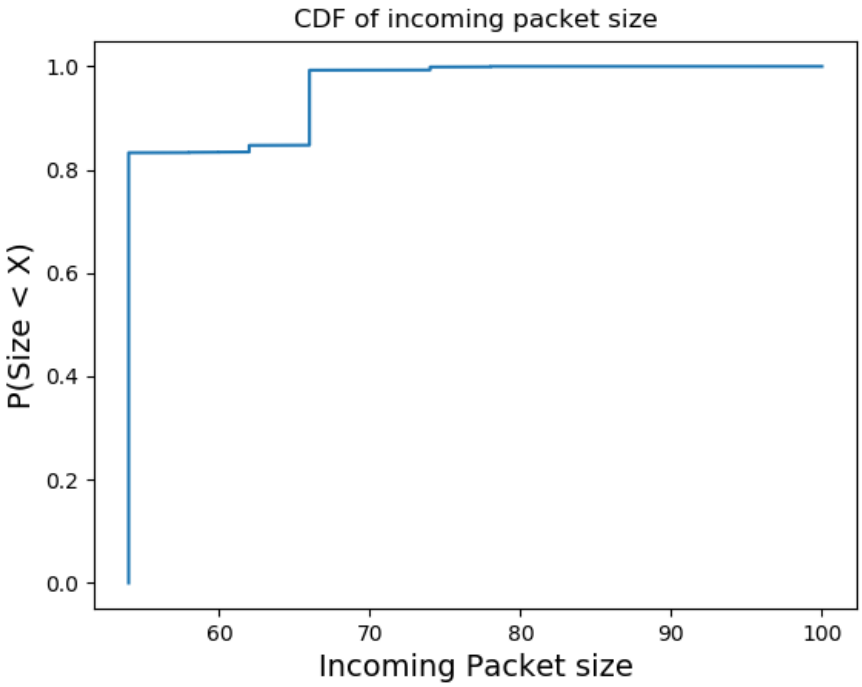
Mean of Inter-arrival Time incoming packets to server: 1.1746484 Median of Inter-arrival Time incoming packets to server: 0.55356800

2.7 CDF of the packet length: P8

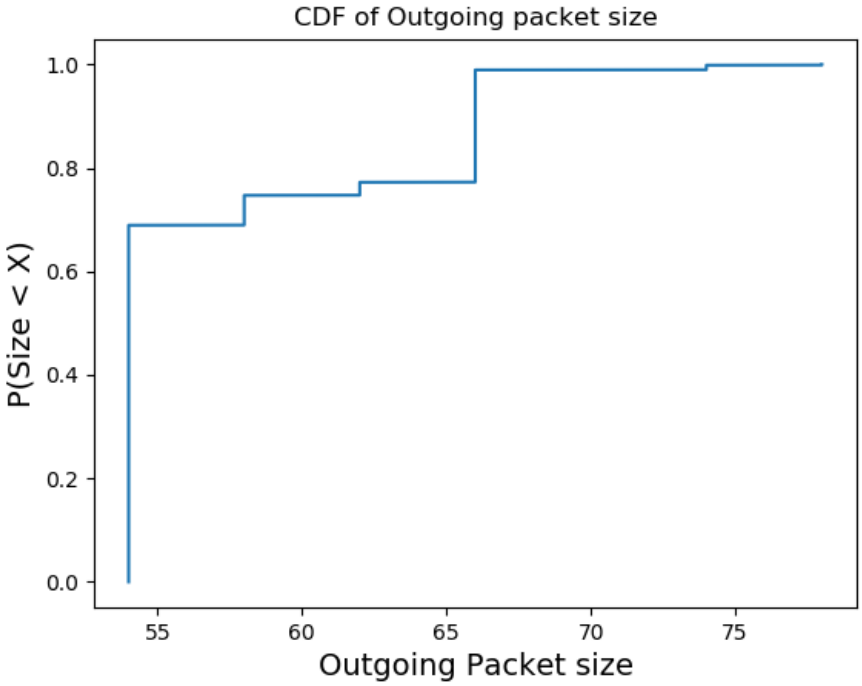
The CDF is probably clustered because the packets sizes are probably fixed to be some discrete values and not the entire range of continuous integers.
The packet size does tend to be concentrated into few clusters of values like 54, 58, 62, 67. This is probably the approximate size of the standard control package that is control package that is being transferred.

2.7.1 Day 1

Incoming Packet Size CDF

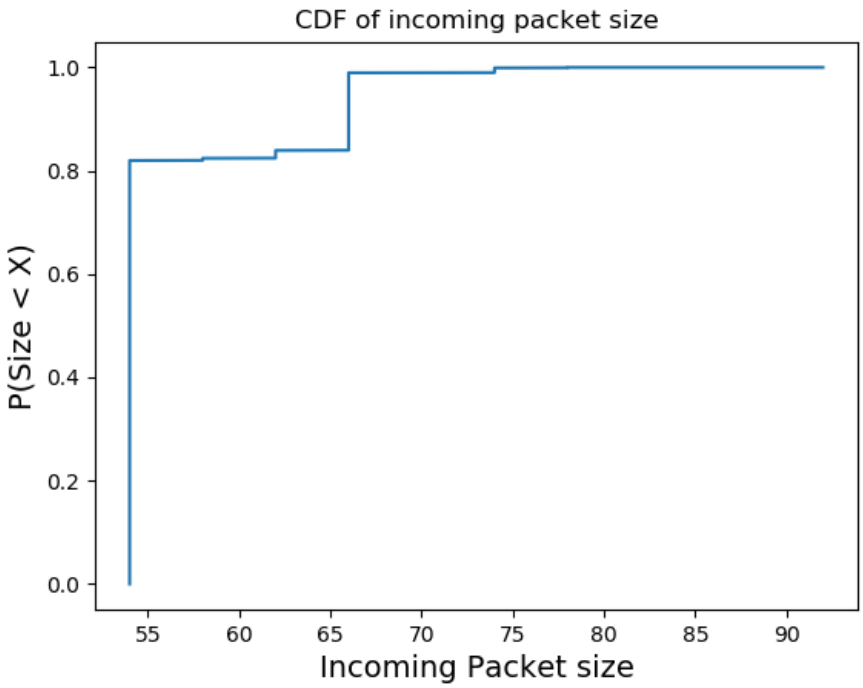


Outgoing Packet Size CDF

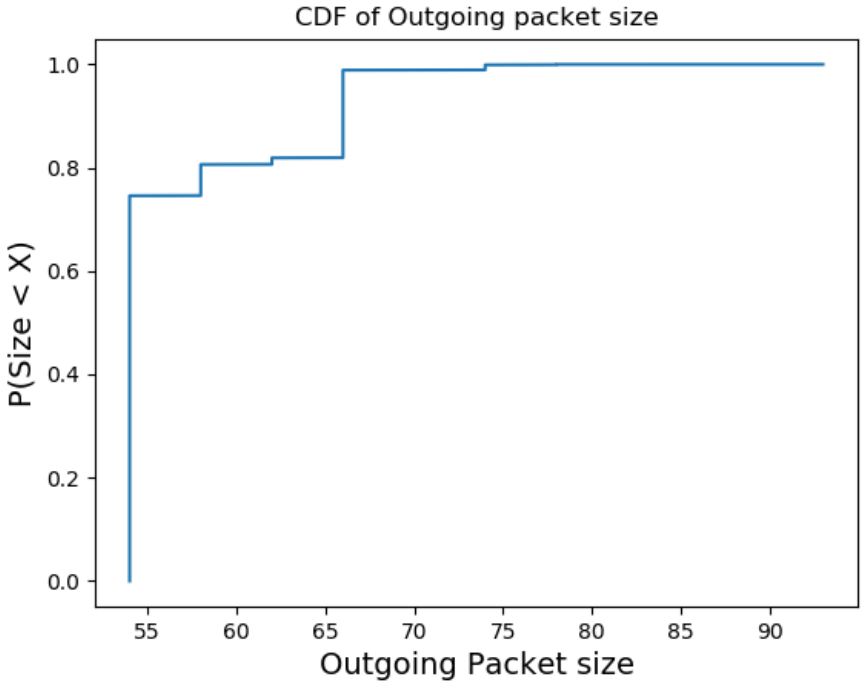


2.7.2 Day 2

Incoming Packet Size CDF

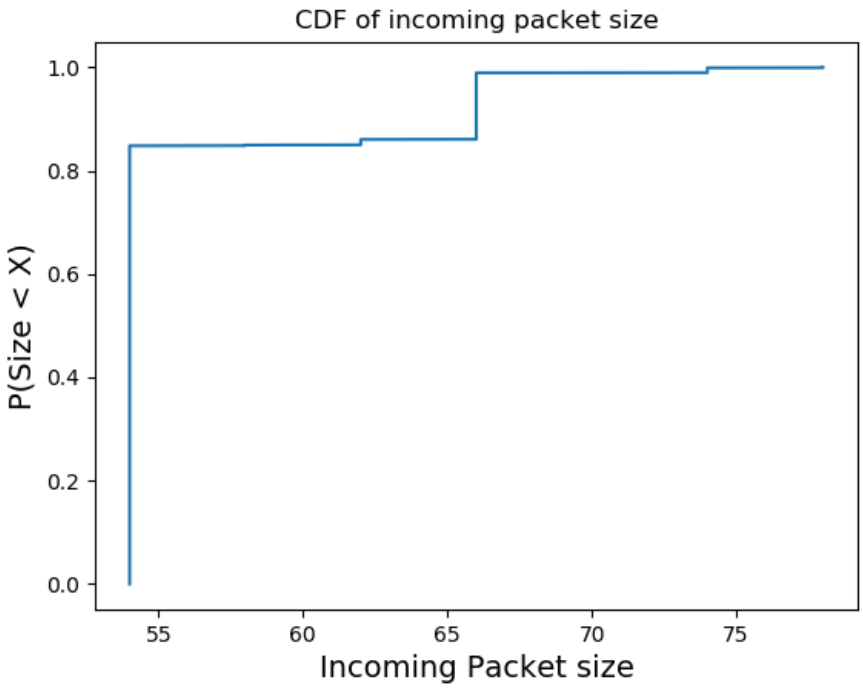


Outgoing Packet Size CDF

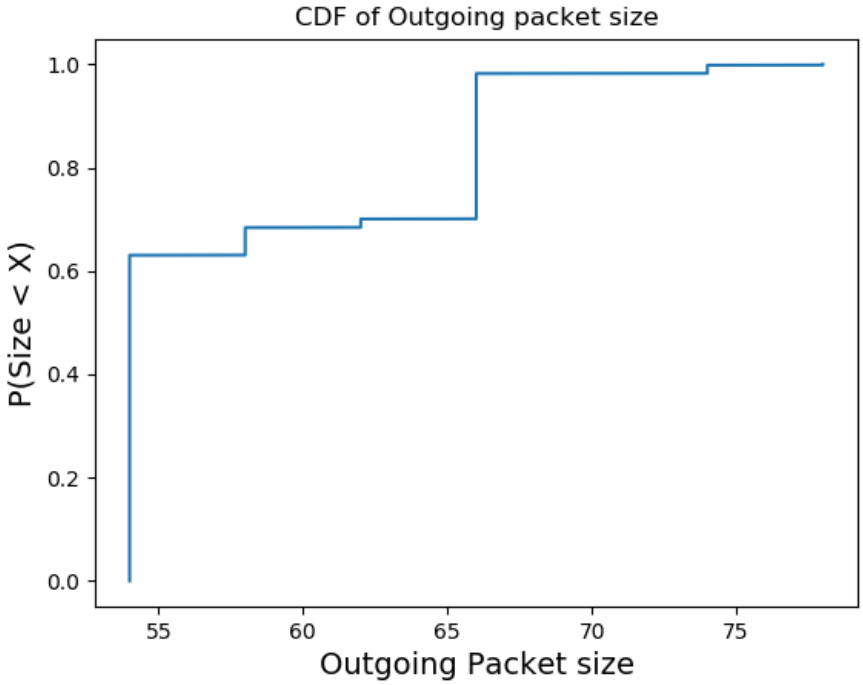


2.7.3 Day 3

Incoming Packet Size CDF

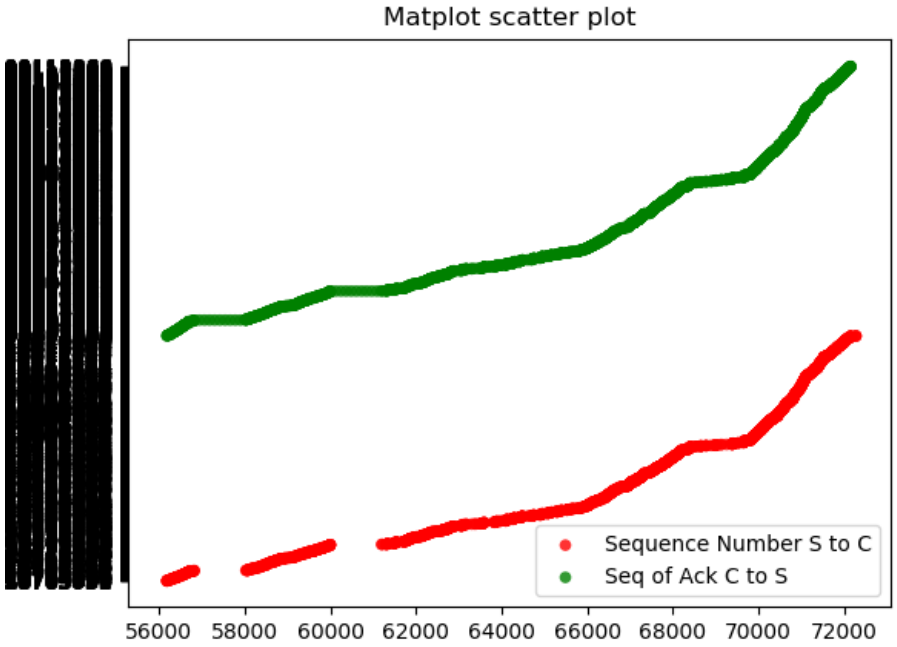


Outgoing Packet Size CDF

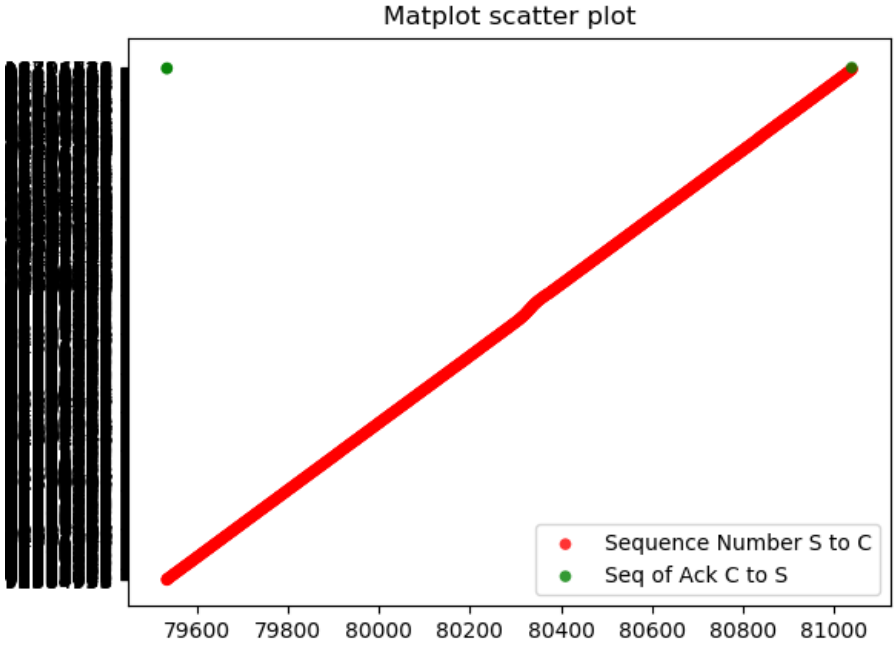


2.8 Flow Analysis: P9

2.8.1 Data Intensive Connections

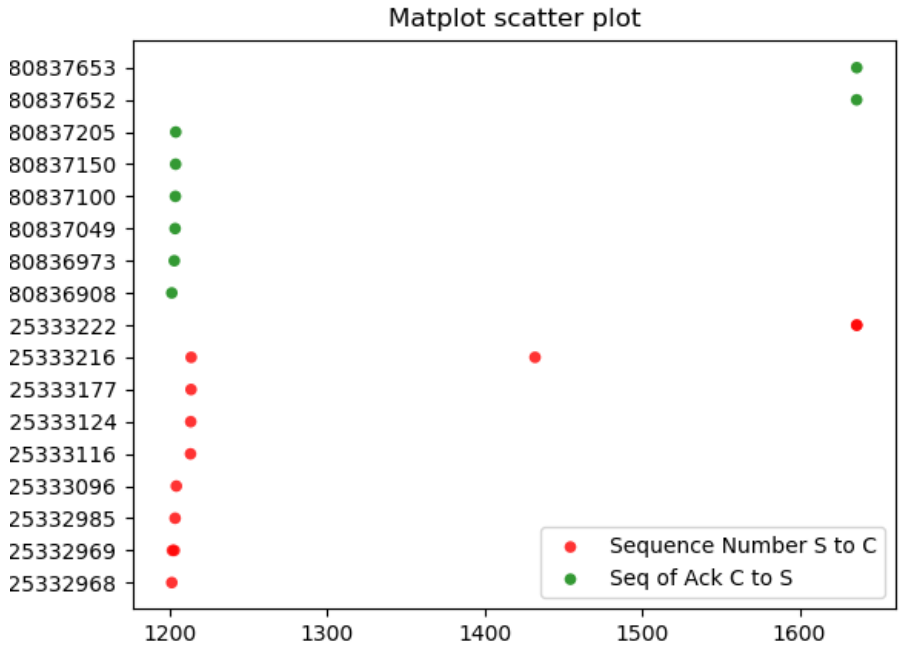


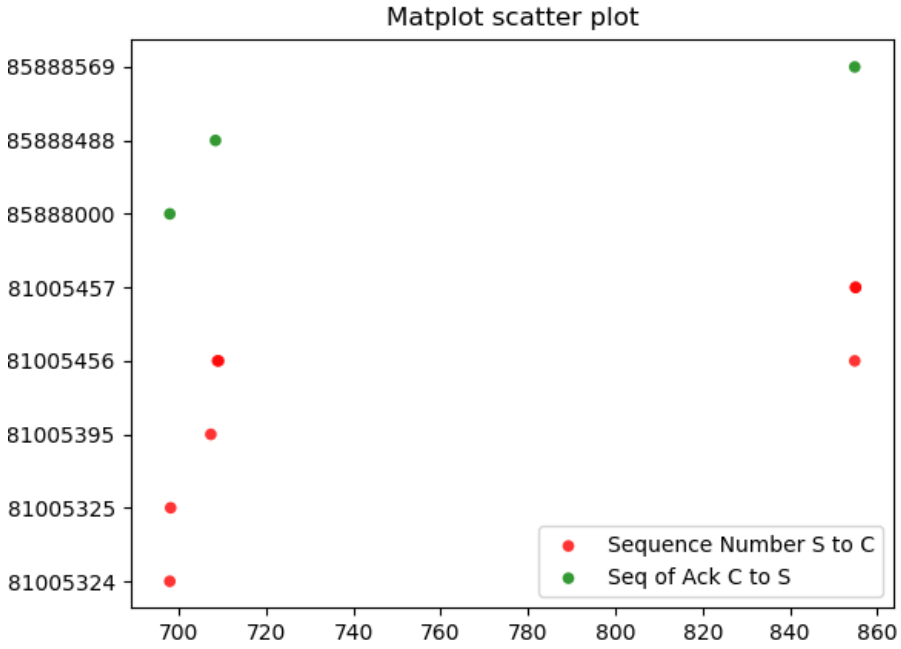
Probably a large file was requested.



2.8.2 Re-transmission

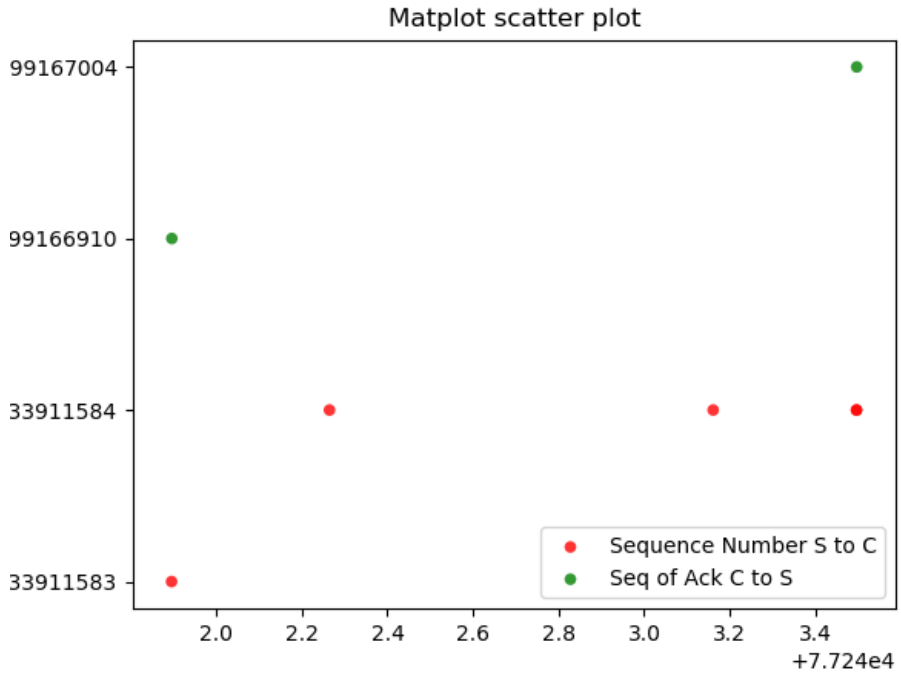
Same sequence number is seen multiple times: two data points at the same y



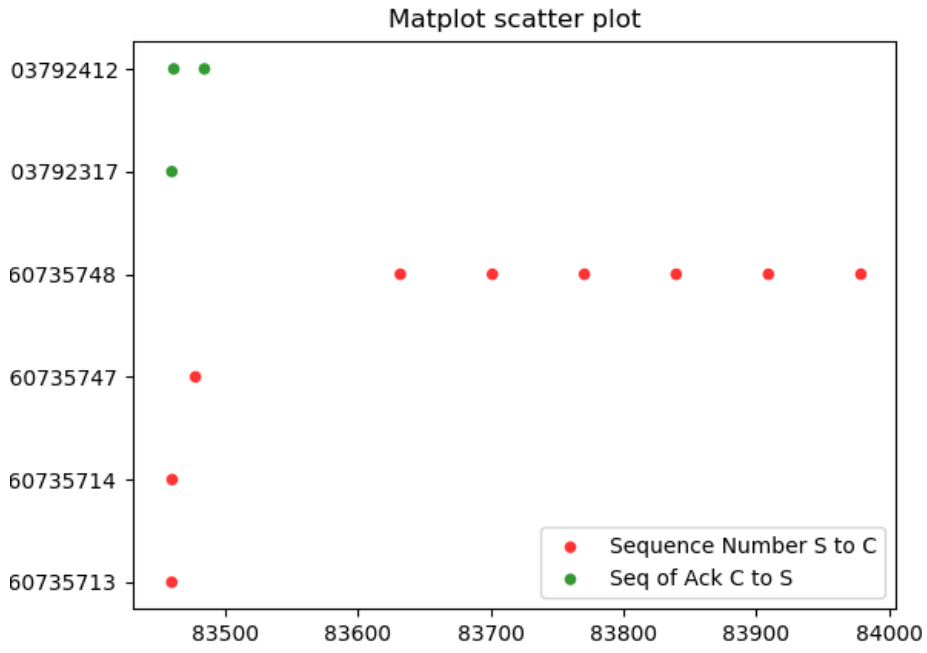


2.8.3 Spurious Re-transmission

Probably the Ack was received almost at the same time as the timeout.

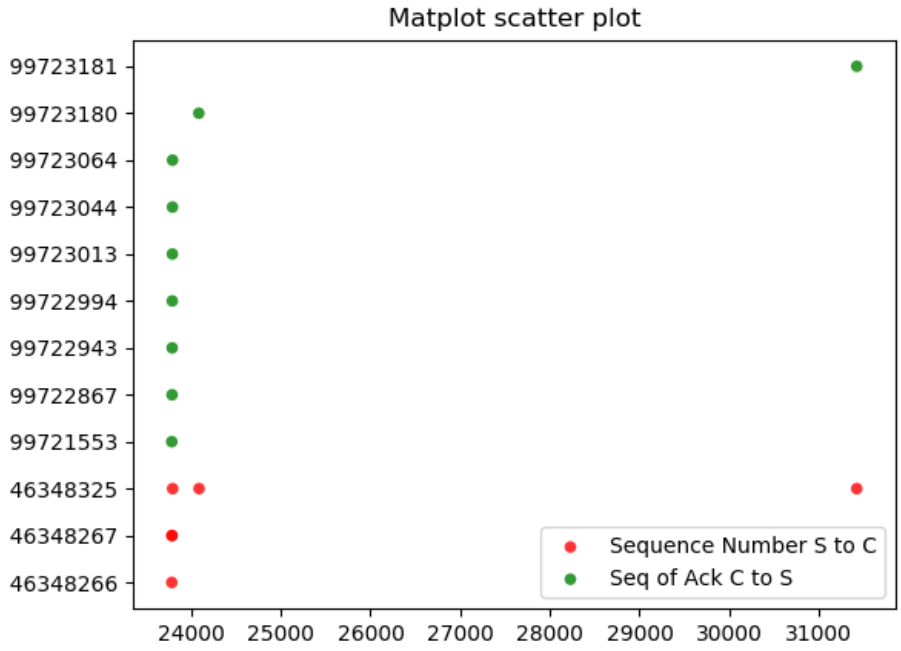


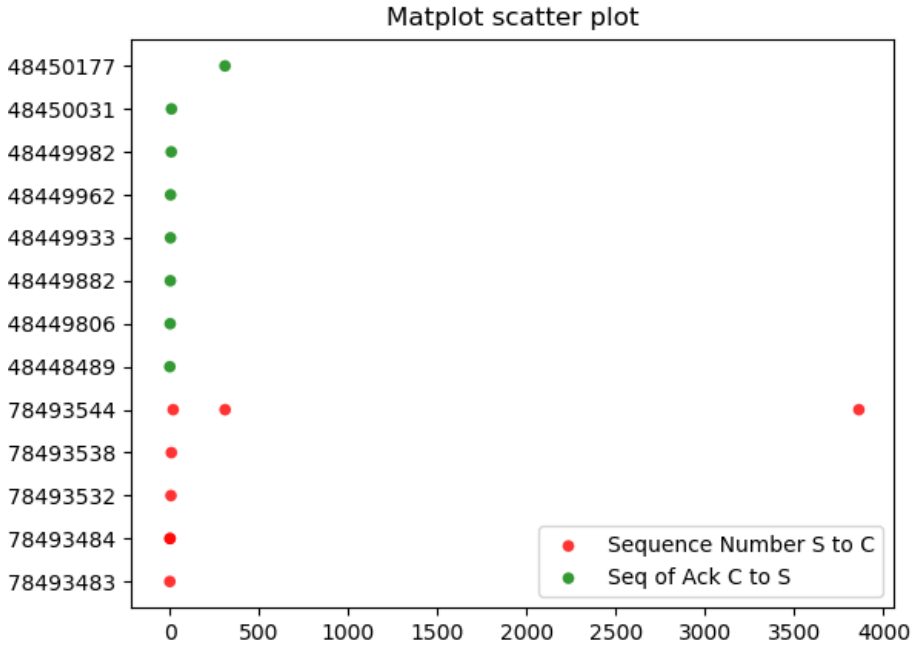
Probably the Ack was corrupted and a data point is missing.



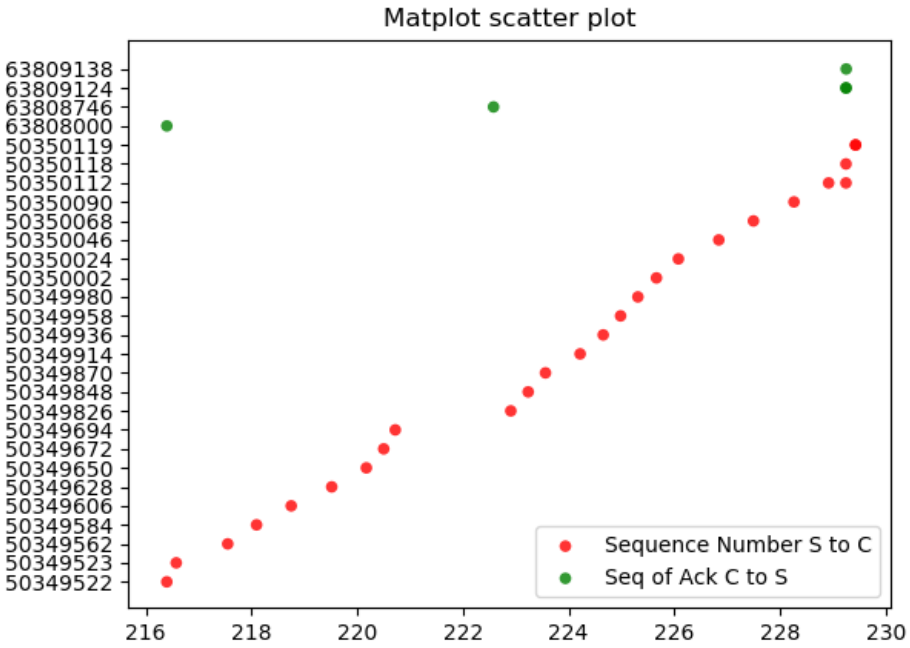
2.8.4 Duplicate Acks

Probably client wants to send more data packets to the server





2.8.5 Out of Order Delivery



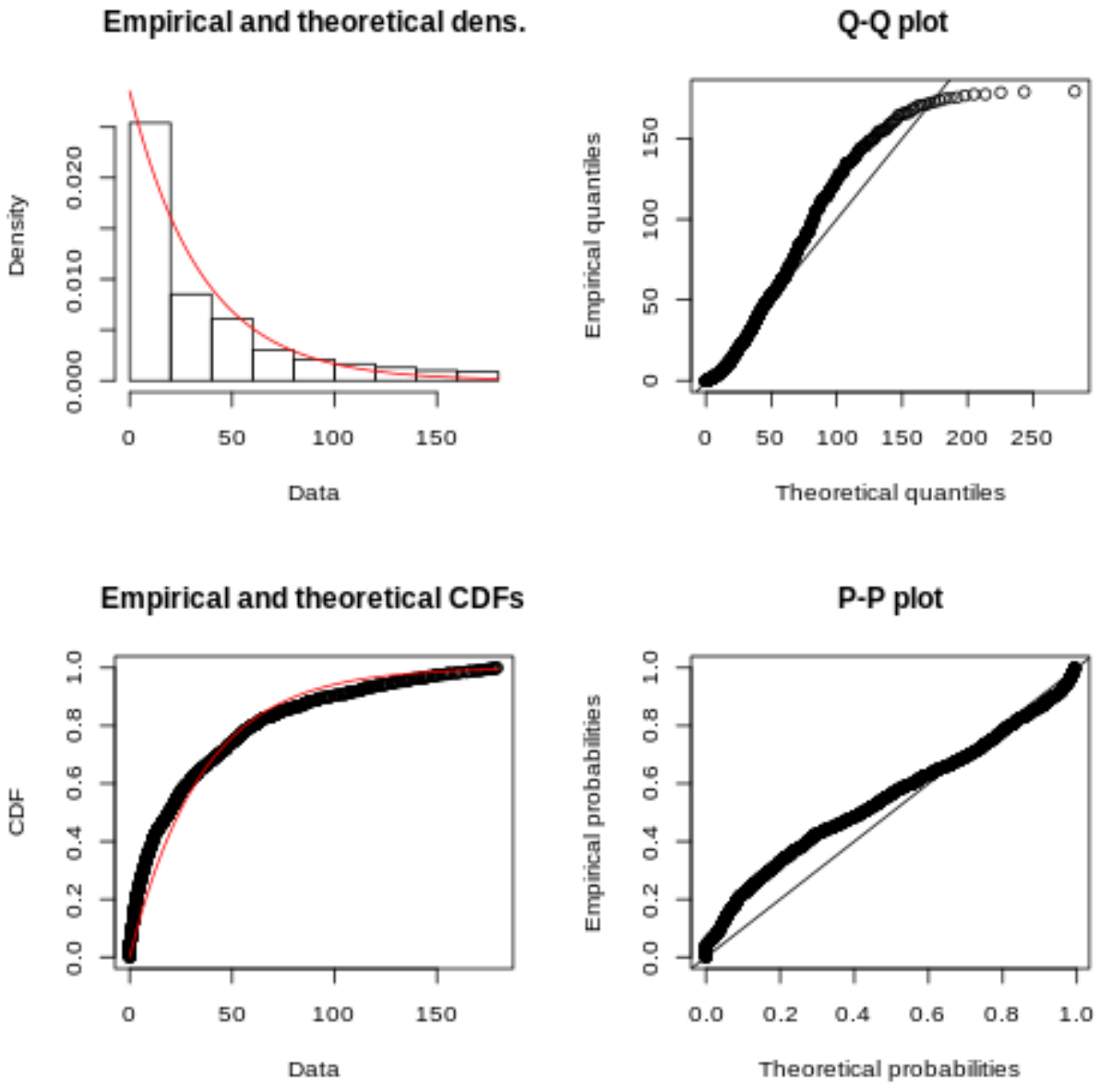
We could only find one such flow (on day 2). The out of order delivery seems to be caused due to re-transmission of data from the server to the client

2.9 Distribution of Inter-Arrival Times: P10

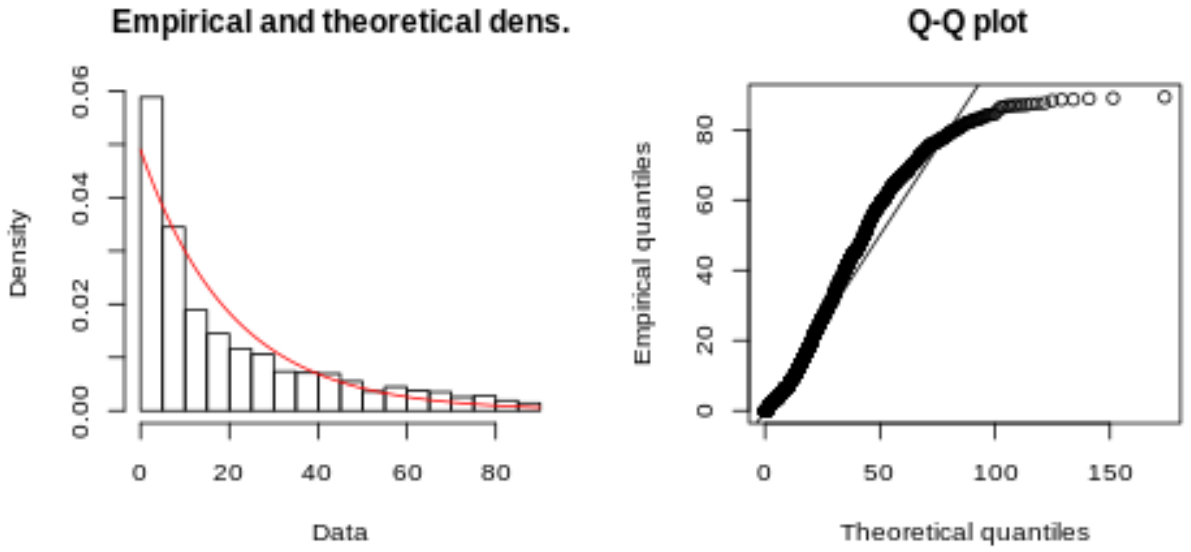
We tried to fit the exponential, log-normal, Normal, weibull distributions to the data from P6 and P7, but it turns out that exponential distribution fits them the best. The absolute decision of which distribution fits the data well was made by looking at the AIC(Akaike information criterion) score obtained from the distfit() function of R.
we also noticed that the data fitted the exponential distribution better after removal of some outlier points. In fact, the fittings shown below are those after removal of outliers.

2.9.1 Two consecutive connections being opened

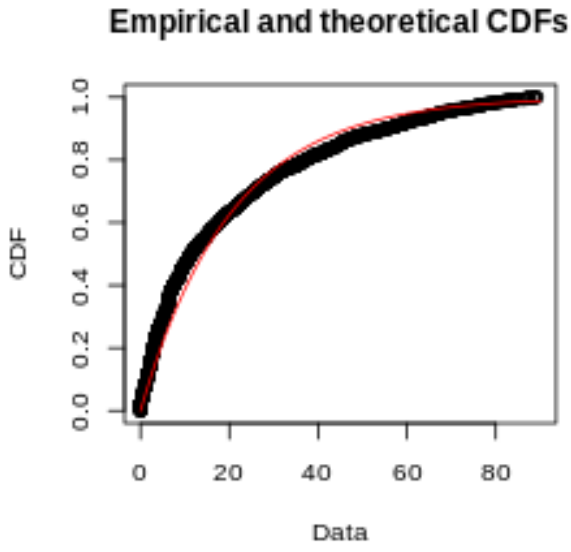
Day1:

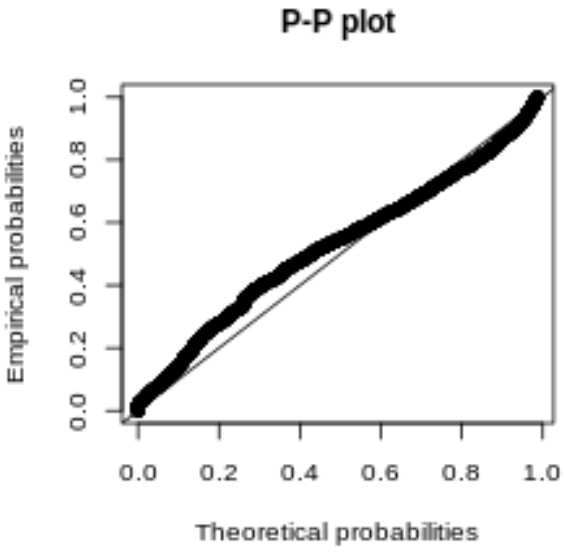


Day2:

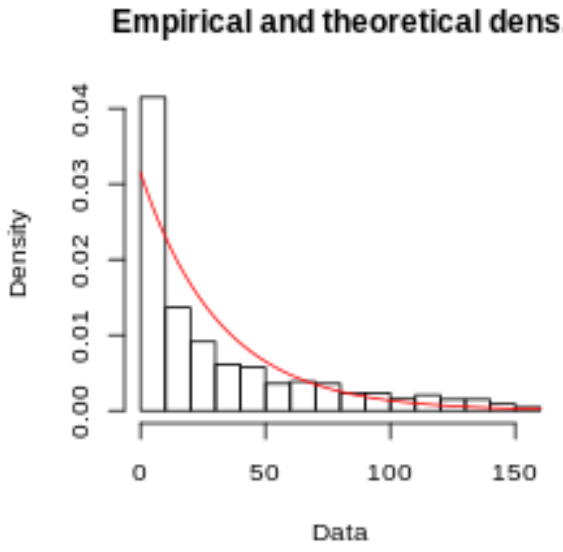

Empirical and theoretical dens.
Density
0.00
0.02
0.04
0.06
0
20
40
60
80
Data

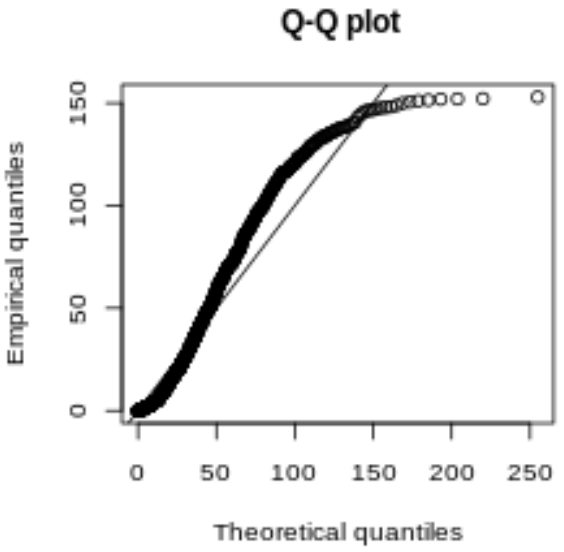
Q-Q plot
Empirical quantiles
0
20
40
60
80
0
50
100
150
Theoretical quantiles

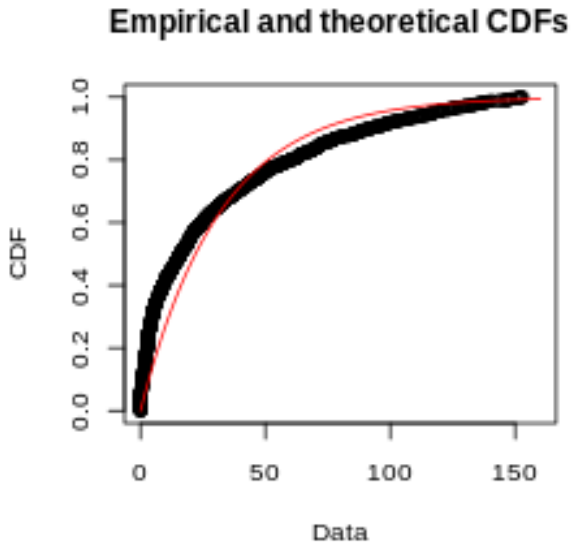

Empirical and theoretical CDFs
CDF
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
Data

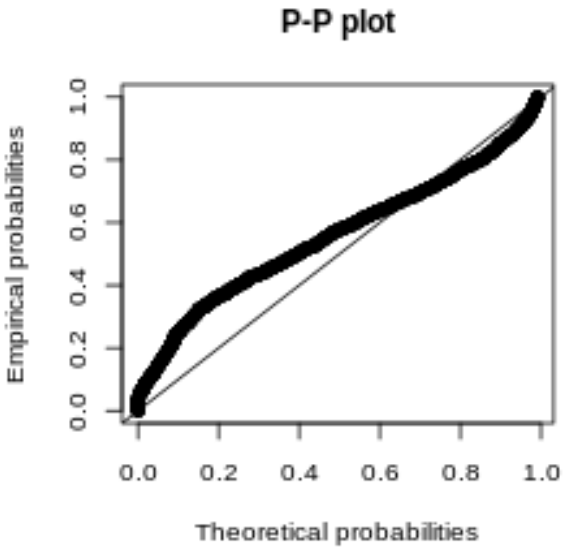

P-P plot
Empirical probabilities
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Theoretical probabilities

Day3:


Empirical and theoretical dens.
Density
0.00
0.01
0.02
0.03
0.04
0
50
100
150
Data

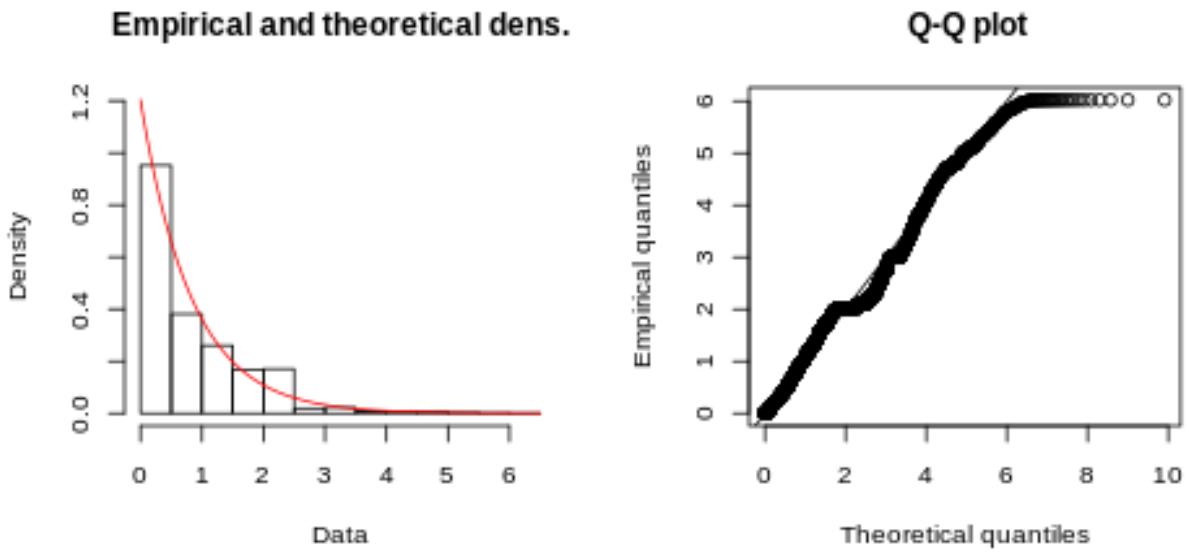

Q-Q plot
Empirical quantiles
0
50
100
150
0
50
100
150
200
250
Theoretical quantiles

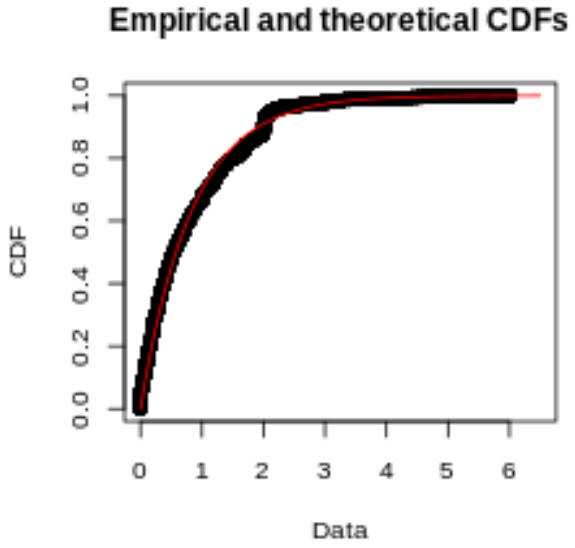

Empirical and theoretical CDFs
CDF
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
Data

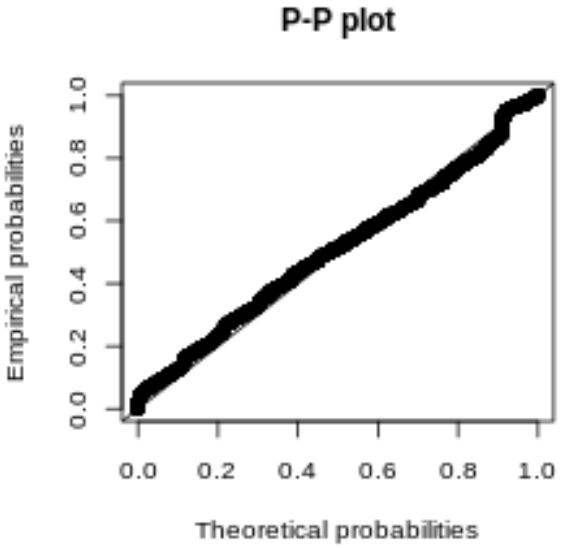

P-P plot
Empirical probabilities
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Theoretical probabilities

2.9.2 Two consecutive incoming packets to the servers

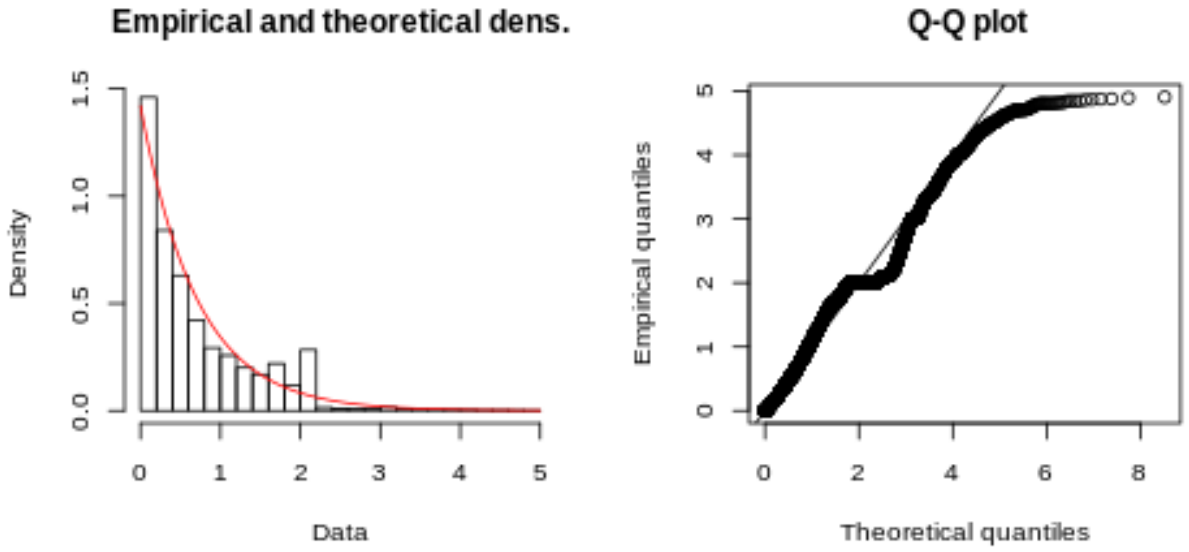
Day1:

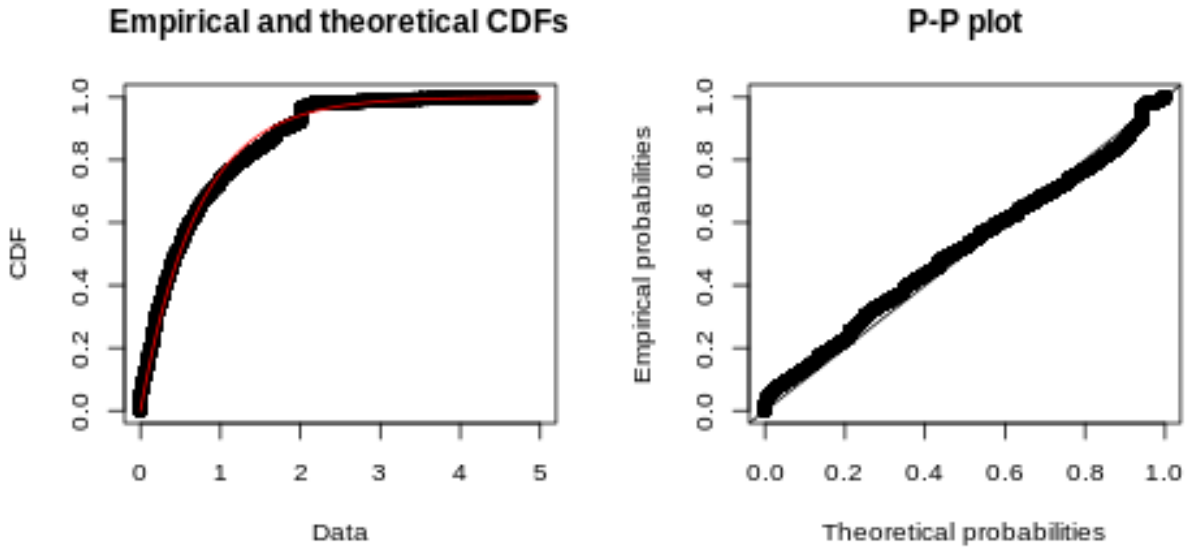




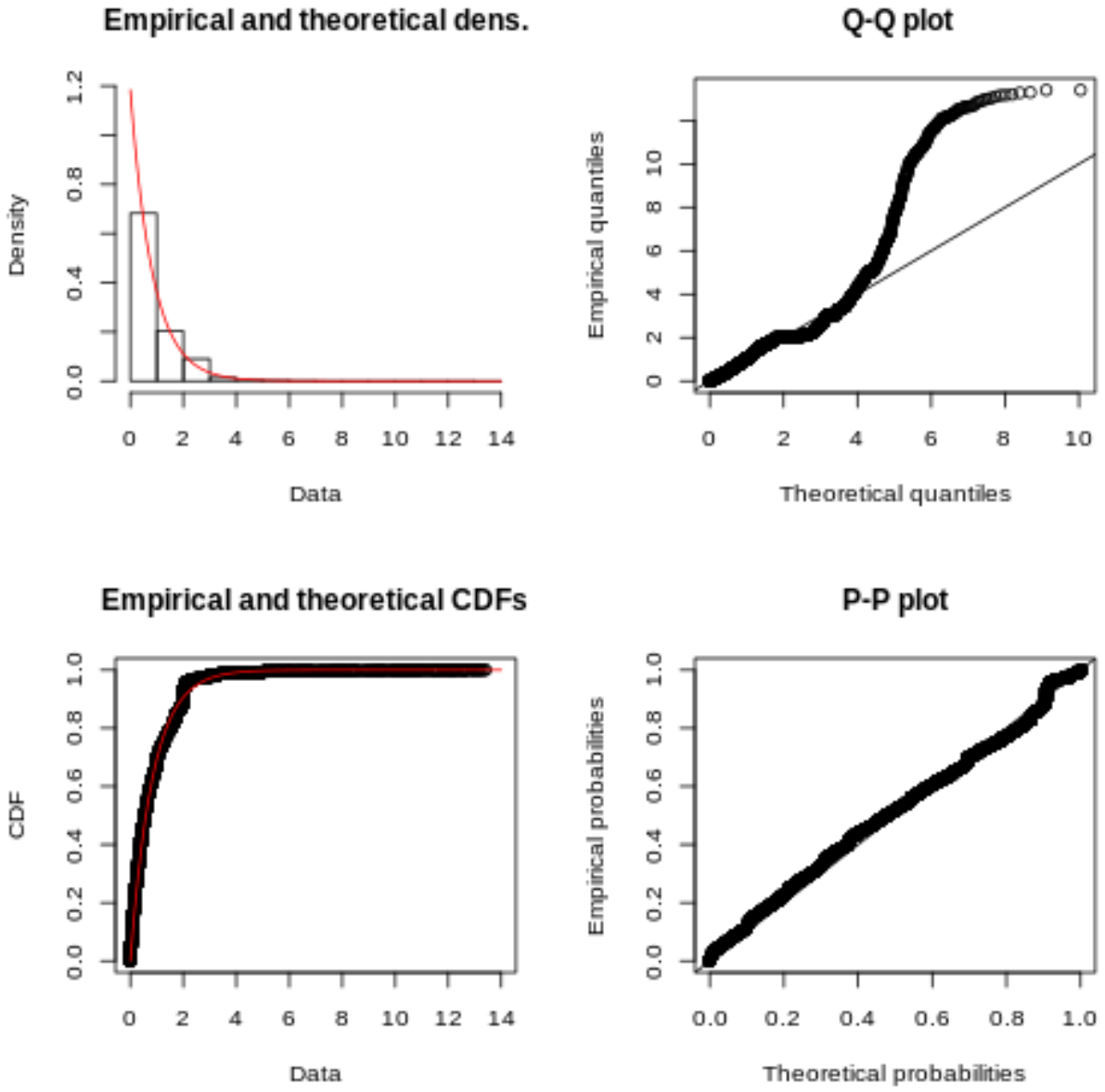


Day2:


Empirical and theoretical dens.
Density
Data
Q-Q plot
Empirical quantiles
Theoretical quantiles


Empirical and theoretical CDFs
CDF
Data
P-P plot
Empirical probabilities
Theoretical probabilities

Day3:



Rate (Expected No. of packets per second) Parameter of the Exponential Function Fitting:

Rate :	PART 6	PART 7
Day 1	0.03404845	1.118508
Day 2	0.04576802	1.337741
Day 3	0.03390657	1.183119

2.10 P 11

Inter-arrival Time between Packets(sec):

Rate :	PART 6	PART 7
Day 1	29.37	0.894
Day 2	21.85	0.747
Day 3	29.50	0.845

Mean packet size = 56 bytes

Value of mu = 285 packets per second

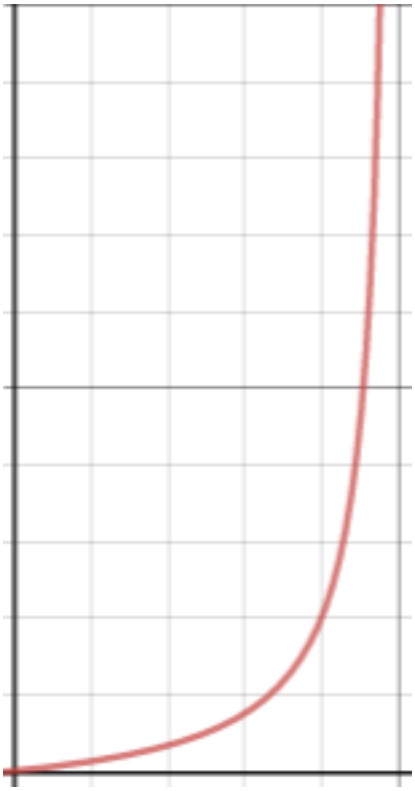
Utilization factor rho on the 3 days:
D1=0.0039
D2=0.0047
D3=0.0041

Average Queue Size:
D1=0.0039
D2=0.0047
D3=0.0041

Average Waiting Time:
D1=13 microseconds
D2=16 microseconds
D3=14 microseconds

Waiting time vs lambda (for constant mu)

As can be seen the waiting time and the queue shoot up exponentially as lambda (arrival rate of packets) gets closer to mu



Average Queue Size vs Lambda