SELF SIMILARITY AND QUEUING ANALYSIS OF LTE SYSTEMS

Prem Sujan(170123027)
&
Garvit Mehta(170123018)

Project Guide: Prof. N. Selvaraju

Objective

The main objective of this project is to analyse self-similar property of LTE network traffic to understand burstiness of the network. We also provide a Queuing Model in the context of LTE Networks. We then calculate the hurst parameter for traffic accumulated at several base stations, and simulate M/M/1 queuing system using NS3 to derive its performance metrics experimentally.

M/M/1 Model Performance Metrics

The quality of the service as seen from the customer perspective can be determined by the following metrics

L_s is the average number of customers in the system

L_q is the average number of customers in the queue

W_s is the average time spent by a customer in the system

W_q is the average time spent by a customer in the queue

Results

From Little's law,

$W_s = L_s/\lambda = (1/\mu-\lambda)$

$W_q = L_q/\lambda = (\lambda/\mu(\mu-\lambda))$

It can be observed that

$L_q = L_s\rho$ $\quad$ $W_q = W_s\rho$,where $\rho = \lambda/\mu$

Nburst/M/1 model

Need For a New Model

- There is an incredible growth in the number of wireless devices such as smart-phones, tablets and Internet of Thing (IoT) .
- In addition to the fast development of media streaming applications, IPTV, telemedicine and Internet gaming have led to a significant challenge to the design and deployment of cellular technology.
- Investigating and analyzing the distribution of data generated by each device in LTE network should be the most important factor to estimate the quality requirements and capabilities of LTE networks.

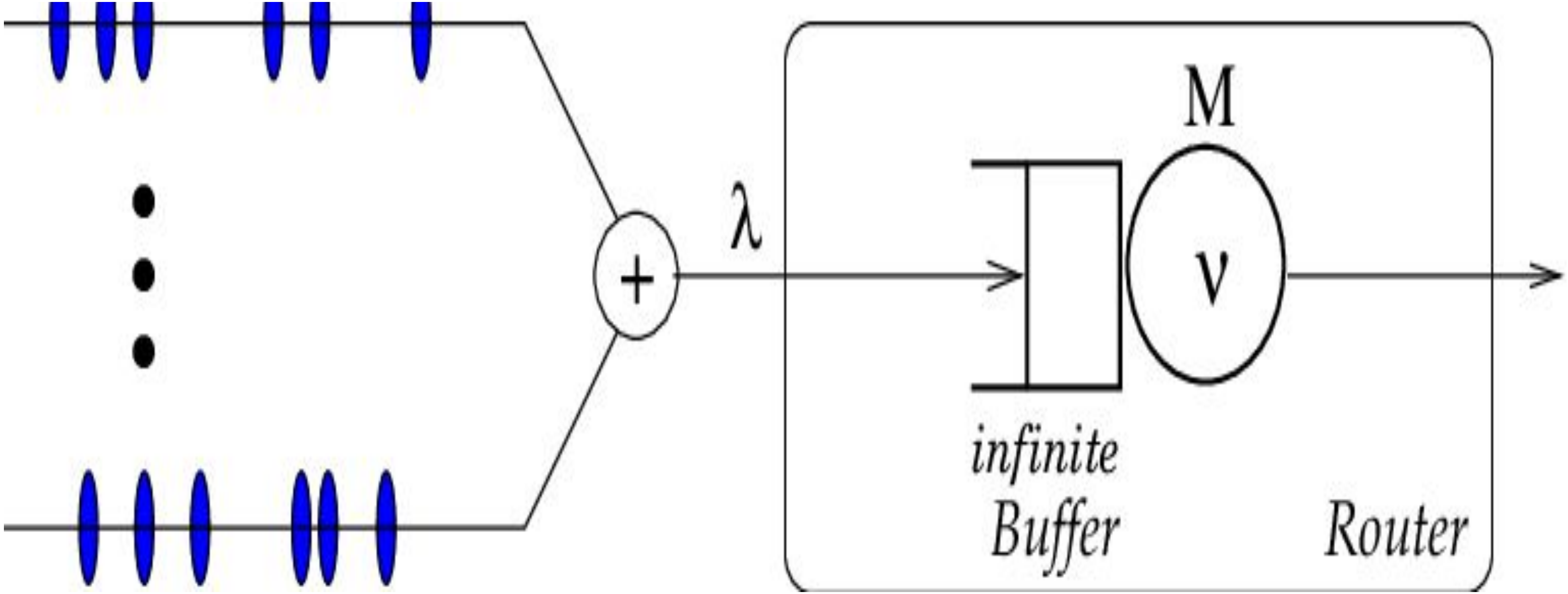
Required Properties for an appropriate model

The main difference from the standard models is the requests in telecommunication networks don't follow a distribution rather come in a continuous bursts.

The unpredictable traffic must be both bursty and self-similar.

Many models like M/G/1 queue with changing parameters, batch-arrival model, continuous burst flow models failed to mimic the traffic in telecommunication networks.

The packet generated by each user arrive at a single queue which is maintained by a Router.



Describing Network Queue

The number of packets requested by a mobile device is considered as a random variable.

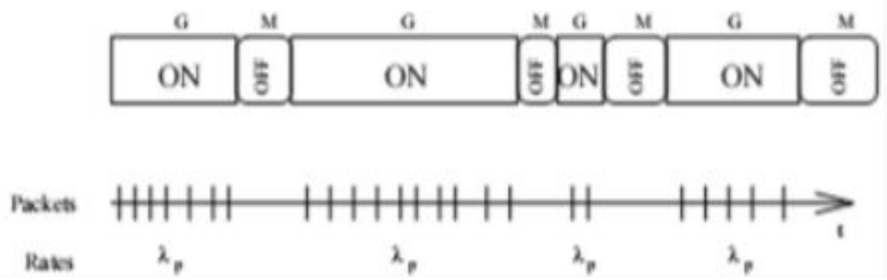
The properties and distribution of this random variable is purely based on the mobile application used to request the data.

This is the reason standard queuing models are not appropriate to understand telecommunication systems.

Describing Network Queue

N Burst Model is one of the variants of ON-OFF models

The arrival process of N-Burst is the superposition of N ON-OFF type identical and independently distributed source's traffic streams.



Essential parameters of Model

κ : average rate of packet arrival for each source.

ʌ: The overall arrival rate that is produced by N sources i.e. $\Lambda = \kappa N$

ʌp: the peak transmission rate during a burst i.e. at ON time.

ʌb: Mean burst arrival rate =ʌ/Nb.

Nb: Average number of packets during a burst (at ON period)

<u>ON</u> : the average time during which the node is active

<u>OFF</u> : the average time during which the node transmission is OFF.

SELF SIMILARITY

Self Similarity

Self-Similarity is the property we can associate with an object whose appearance remains same irrespective of the scale at which it is viewed.

Self similarity is used in the distributional sense in the case of stochastic objects like time series: the object's correlational structure remains unchanged when presented at different scales. As a consequence, at a variety of time scales, bursts in a time series.

Self Similarity

Since LTE technology is rapidly expanding in terms of coverage and user base, it is essential to investigate its network traffic

A bursty traffic can be described statistically using self-similarity. Since bursts are observed on all time scales, traffic at certain time are generally correlated with traffic at a future time.

Self Similarity

For a time series $X = (X_t : t = 0, 1, 2, ...)$, the m aggregated series is given by $X^{(m)} = (X^{(m)}_k : k = 0, 1, 2, ...)$ by summing the original series X over non-overlapping blocks of size m.

If X is self similar, it has the same auto-correlation function $r(k) = E[(X_t - \mu)(X_{t+k} - \mu)]$ as $X^{(m)}$ for all m.

Self Similarity

Self-similar time series shows long-range dependence, $r(k) \sim k^{-\beta}$ as $k \rightarrow \infty$ where $0 < \beta < 1$.

Self similarity is expressed using a single variable, representing the speed of decay of autocorrelation function known as Hurst parameter $H = 1 - \beta/2$. For self similar series $1/2 < H < 1$, when $H \rightarrow 1$, degree of self-similarity increase.

Existence of Self Similarity

There are several arguments made by researchers about why the Internet traffic is self-similar ranging from file size distribution on web servers, ON/OFF models of heavily tailed distribution, user behavior, network protocols, buffer in routers and the TCP congestion avoidance algorithms.

Extensive statistical analysis shows that the data at the level of user equipment or source-destination pairs are self-similar and exhibit high variability.

Effects of Self Similarity

When traffic increases in a self-similar network, the bandwidth and buffer sizes can't handle the bursts, resulting in packet loss, implying financial loss for network operator.

Packets are sent again which again leads to congestion and wastage of resources.

Effects of Self Similarity

Predictive feedback control method uses dynamic traffic flow control, by adjusting congestion based on either nodes have on-set of concentrated periods of high or low activity.

Error correction method uses re-transmission of non viable data like streaming audio or video. The level of redundancy is adjusted according to the congestion level. This method has the risk of damaging the congestion level due to high traffic from these nodes

Hurst Parameter

The Hurst parameter, also known as the Self-Similarity parameter, is a measure of time series long-term memory.

A time series with a value H in the range of 0.5-1 has long-term positive autocorrelation.

A value in the range of 0-0.5 means the time series has long-term swapping between high and low values in adjacent pairs.

Calculating Hurst Parameter

There are a number of methods to calculate hurst parameter like, Variance time plot, R/S plot.

We used Variance-time plot to estimate H, For a self-similar process, the variance of the aggregated time series follows ,

$$Var(X^{(m)}) \approx Var(X)/\, m^{\beta}$$

Calculating Hurst Parameter

Taking logarithm on both sides give us,

$$\log[\mathrm{Var}(X^{(m)})] \approx \log[\mathrm{Var}(X)] - \beta\log(m)$$

Since Var(X) is constant, if we plot $\mathrm{Var}(X^{(m)})$ and m on log-log plot, we should get a straight line with slope $-\beta$

Plugging this value in $H=1-\beta/2$ give the hurst parameter

Traffic analysis

Data collection

We used traffic of 4G cell towers traffic data to study self-similarity of the network.These cell towers serve user equipment in their vicinity. When a user makes a data service request, that device will be served by a 4G cell closest to the user. The traffic of a cell within an hour is given by the data capacity of all devices served by the station.

Example: Cell X is serving 30 subscribers, assuming if a customer on average uses 20Mb per hour. Traffic of cell X that hour = 30 * 20 = 600Mb.

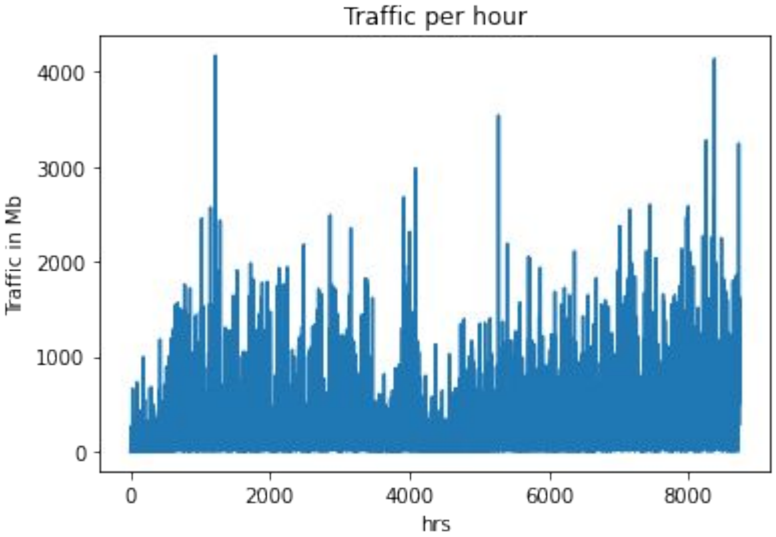
We will see that the nature of traffic varies from time to time, there is peak traffic around 10-12 AM and 11-12PM, and low traffic is observed during early hours of the day

Traffic analysis

Data consists of 50 cells collected over 1 year x 24 hours x 50 cells.

We will see that the nature of traffic varies from time to time

The data of one of the cells is represented here. The y-axis represents traffic accumulated at a eNode base station in Megabytes per hour over 1 year time frame.

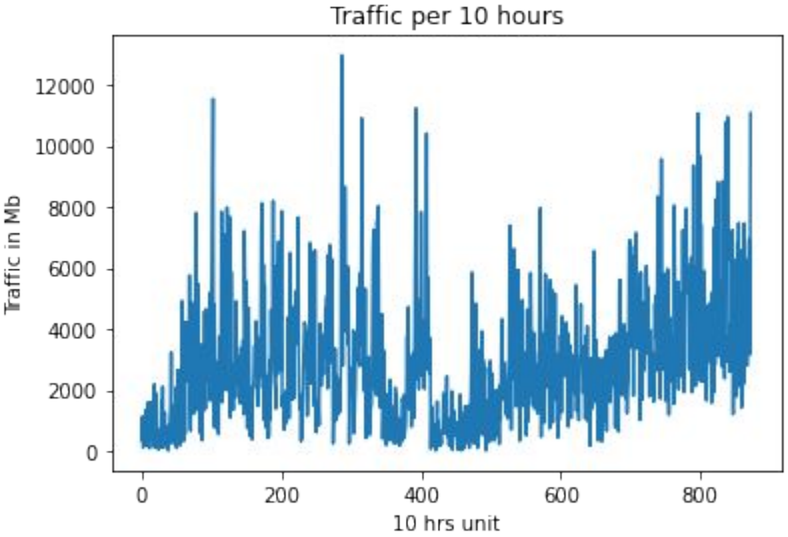


Traffic analysis

Next, we aggregate the traffic accumulated per 10 hours by summing the 10 non-overlapping consecutive traffic per hour.

This is represented with $X^{(10)}$

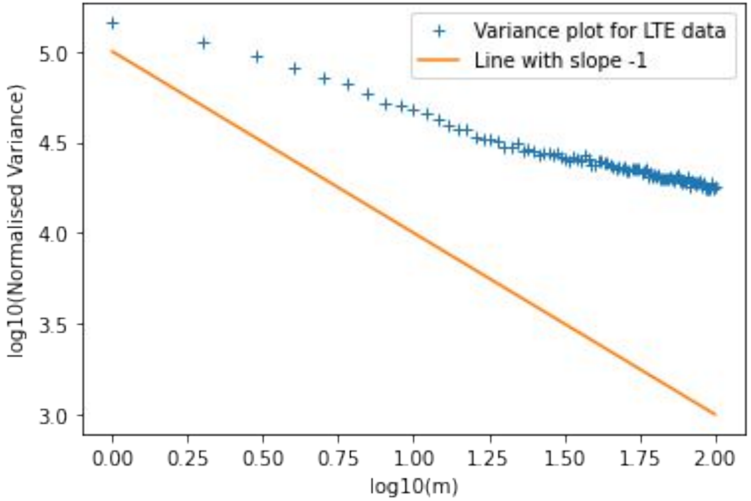
We aggregate for different values of m, to calculate H.



Variance time plot

We use the data presented above to calculate Variance of $X^{(m)}$ for all m ranging from 1 to 100 and plot $\log(Var(X^{(m)}))$ vs $\log(m)$

The blue cross points represent the log of variance of aggregated time series and the log of aggregation size, the red line represents a line with slope -1

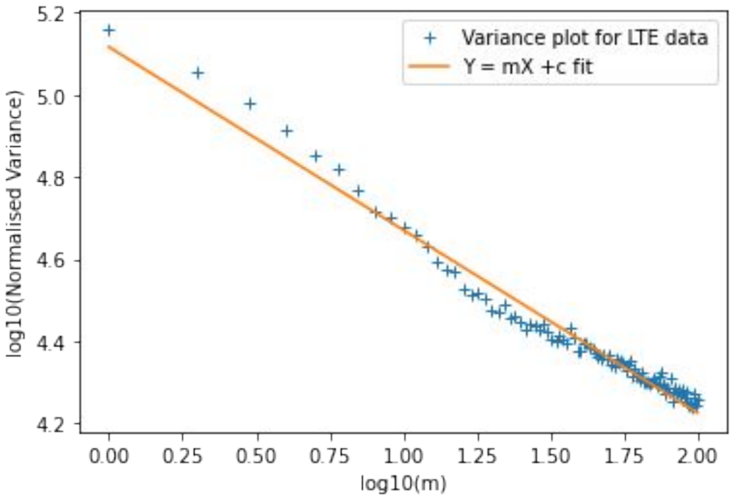


Variance time plot

To calculate the slope we use linear regression to fit the data points along a line,

The plotted curve is fitted with a Y=mX+C curve with best fit slope of -0.44535 and y-intercept of 5.11624

$\Rightarrow -\beta = -0.44535 \Rightarrow H = 1 - \beta/2 = 1-(0.44535)/2 = 0.777325$



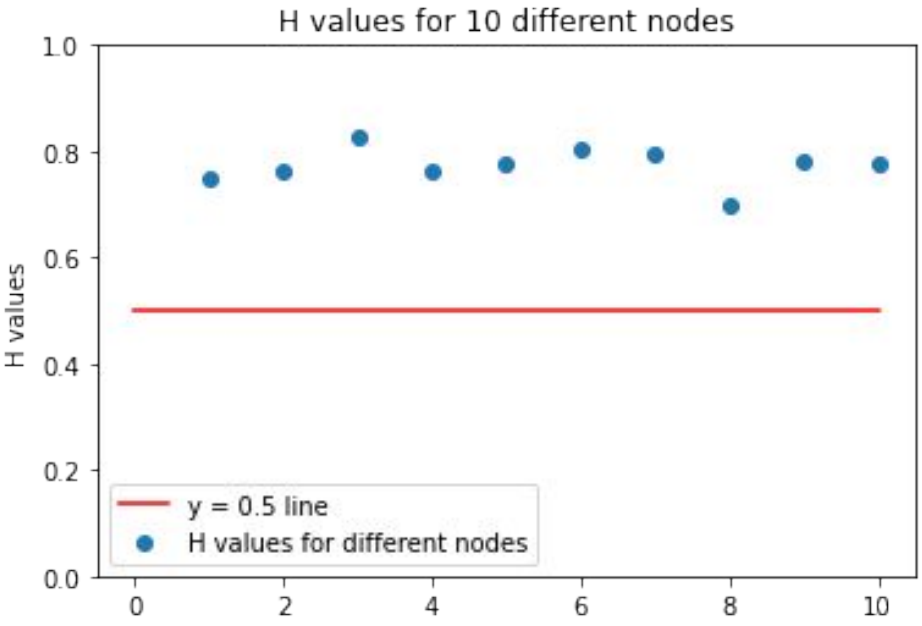
Hurst parameter for different base stations

We take data from more eNode base stations and verify the self-similarity of the traffic using the same method as above, the results of the fit and hurst parameter are given below

Cell Id	A*x + B	beta	Hurst parameter
Cell_000111	-0.5048886657050763*x + 5.881310697038597	0.504889	0.747556
Cell_000112	-0.47420761455870564*x + 5.3577196417669635	0.474208	0.762896
Cell_000113	-0.34961805216086006*x + 5.453306975104272	0.349618	0.825191
Cell_000231	-0.4732641697597102*x + 5.559633607880568	0.473264	0.763368
Cell_000232	-0.4518799424290425*x + 5.462535811927402	0.45188	0.77406
Cell_000233	-0.3898929645820931*x + 5.66215612546308	0.389893	0.805054
Cell_000461	-0.41319447360042366*x + 5.575030923528314	0.413194	0.793403
Cell_000462	-0.6007279557227103*x + 5.026070152619324	0.600728	0.699636
Cell_000463	-0.44186054206982117*x + 5.606661506167345	0.441861	0.77907
Cell_001912	-0.4453502764645639*x + 5.116624398168283	0.44535	0.777325

Hurst parameter for different base stations

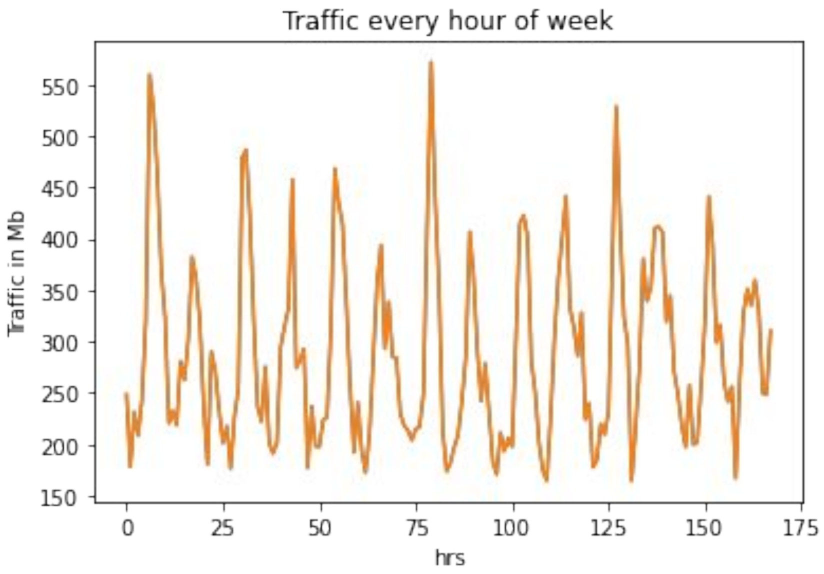
We can see that the Hurst parameter is well above 0.5 indicating the traffic in LTE network is self similar, implying the burstiness in the data requests



Hourly traffic in a week

Traffic of all hours in a week are averaged in the data and is plotted below.

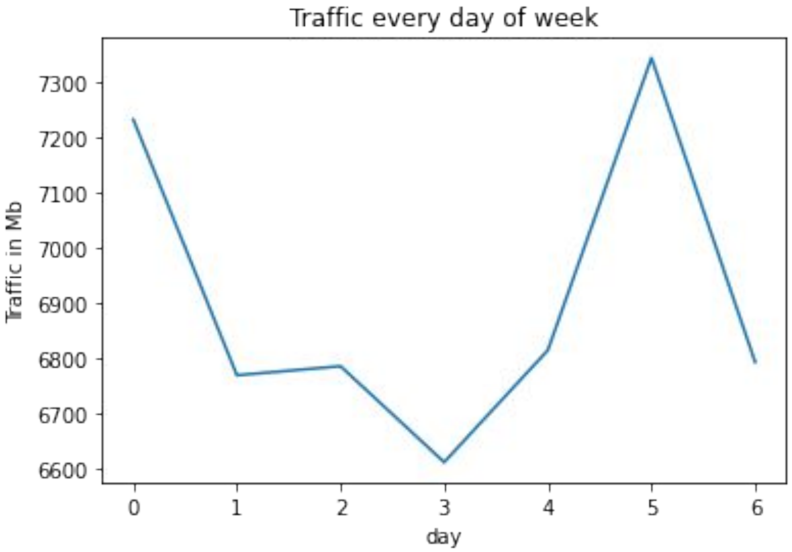
The data points start from Wednesday hour 0, and include traffic for each hour upto an year. We can observe that the traffic in the early hours of the day is much less compared to the peak traffic observed within a day.



Daily traffic in a week

Traffic of all days in a week are averaged in the data and is plotted below.

We can observe that the traffic is more during Wednesdays and Mondays, which is generally true for internet traffic as well.



M/M/1 Simulation

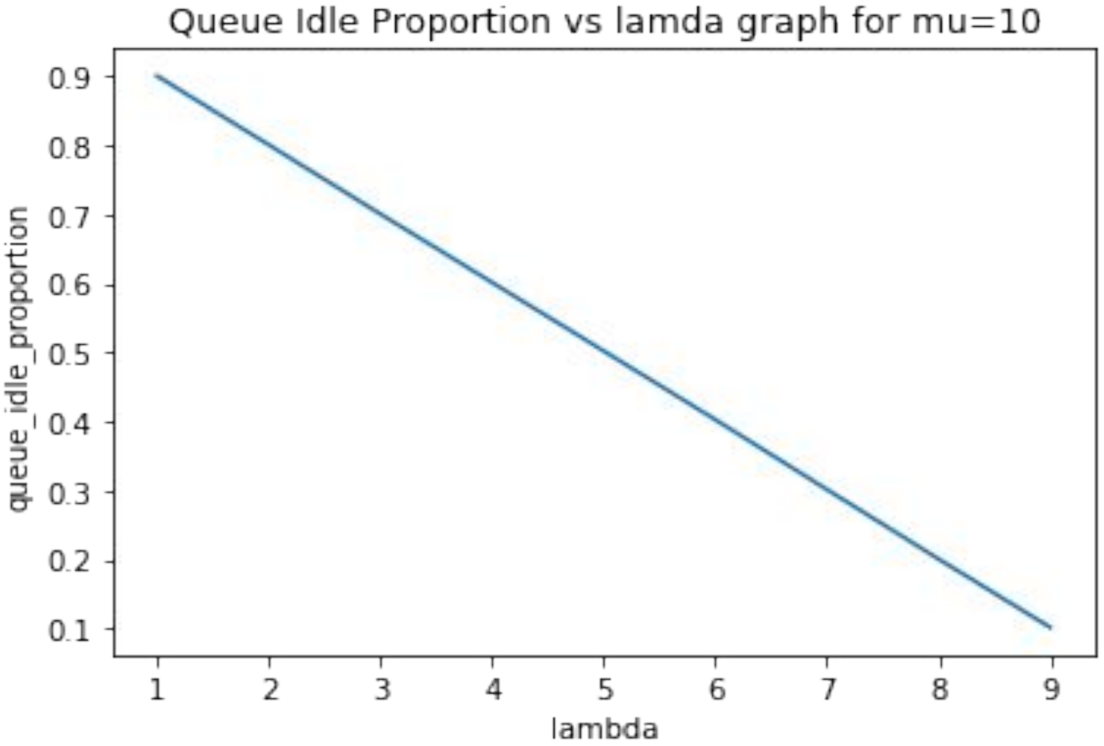
NS3 Simulator

- NS-3 is a discrete-event network simulator, targeted primarily for research and educational use.

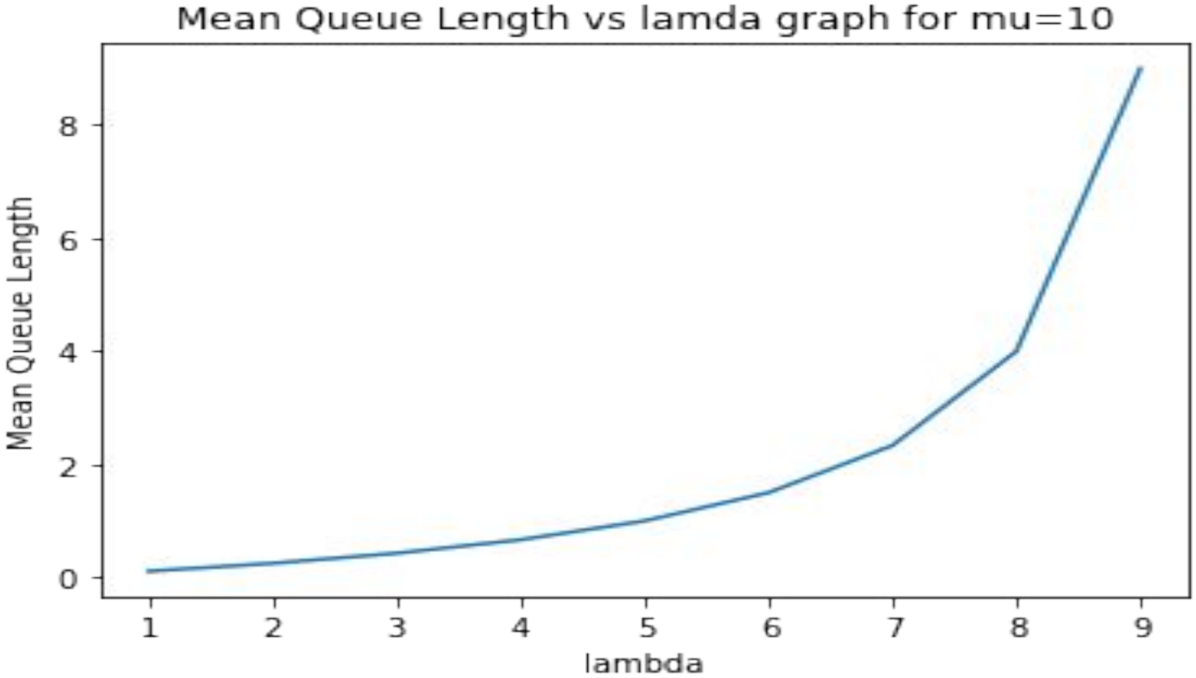
- NS3 is used to simulate modern network systems with real time scheduler many used for academic and research purposes.

Model Implementation

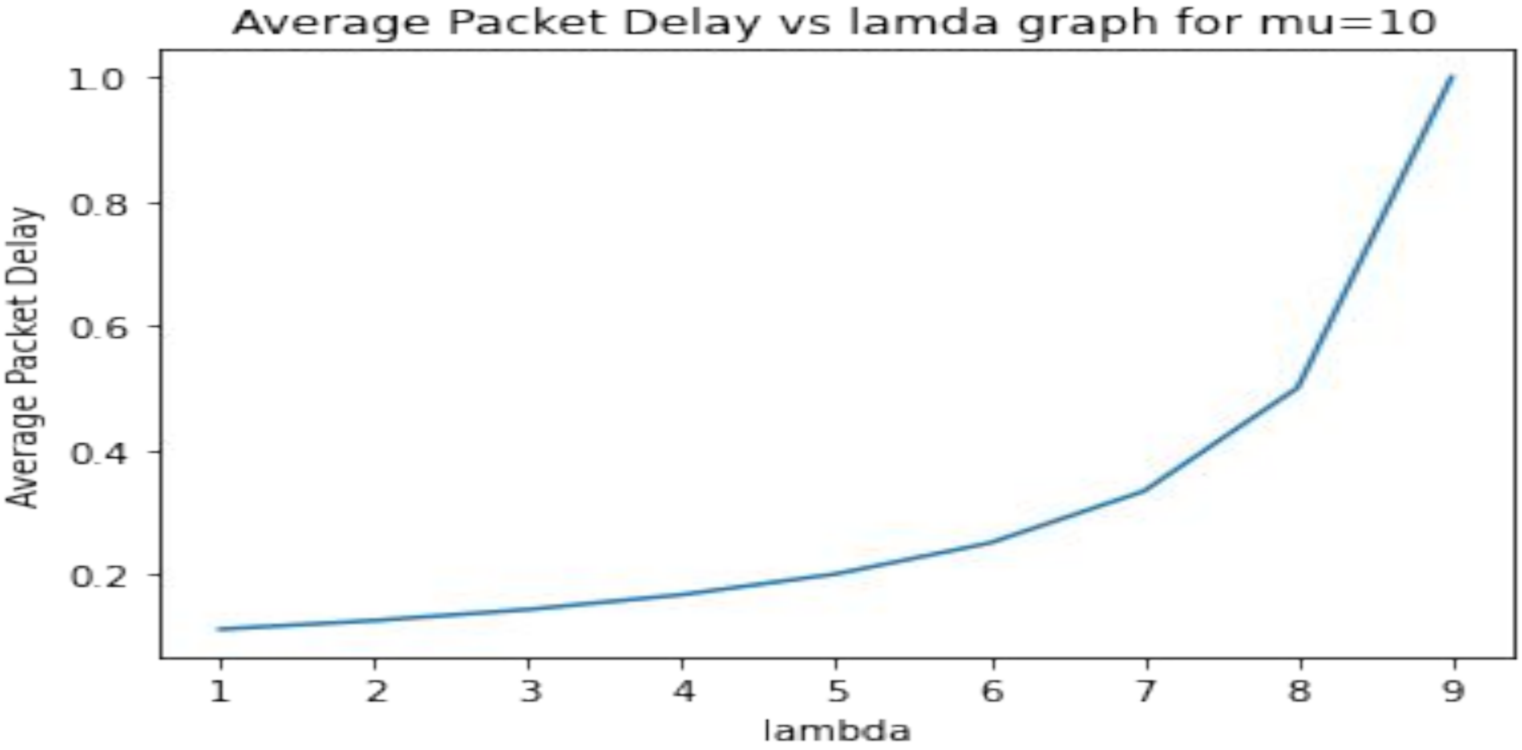
- It can take inputs like lambda referring to arrival rate, mu referring to service rate, initial packets referring to number of packets in queue initially, numpackets referring to number of packets to en-queue, queue limit referring to size of the queue.
- The packets are enqueued according to the arrival rate and dequeued according to the service rate.
- All activities are logged and traces are stored.



- Increasing lambda linearly decrease the idle proportion of the queue linearly following the formula P(idle) = 1 - ρ



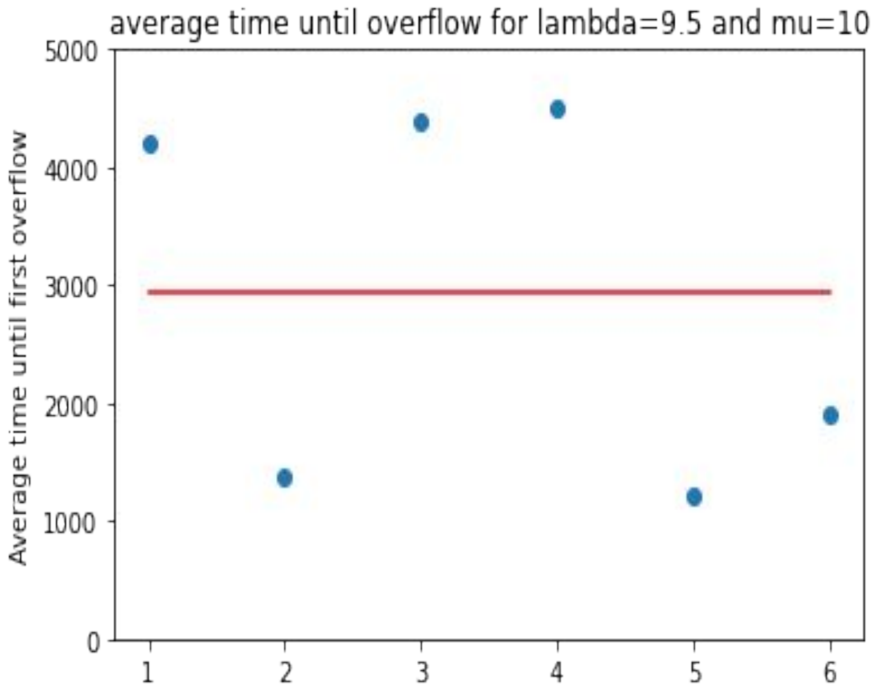
- Length of the system in steady state is consistent with the given by the formula: $L_s = \rho/(1-\rho)$



Mean Delay of packets is consistent with the formula $W_s = L_s/\lambda = 1/(\mu-\lambda)$

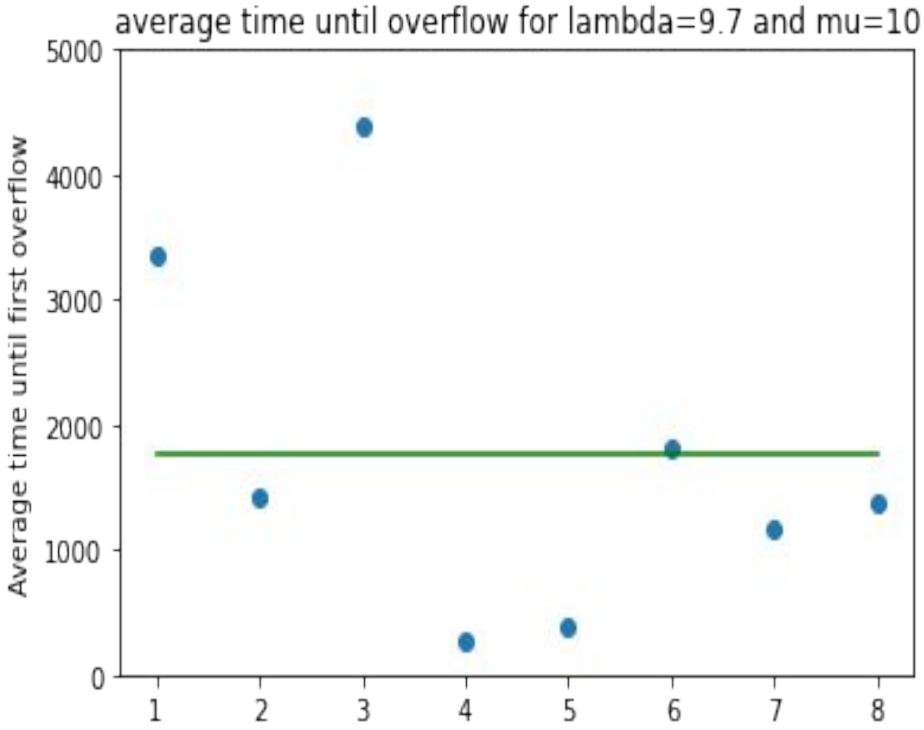
Buffer Overflow for λ = 9.5 and μ = 10

- The individual points represent the times at which the queue overflowed for the first time for different simulation.
- The horizontal line represents the average of the overflow times of these simulations.



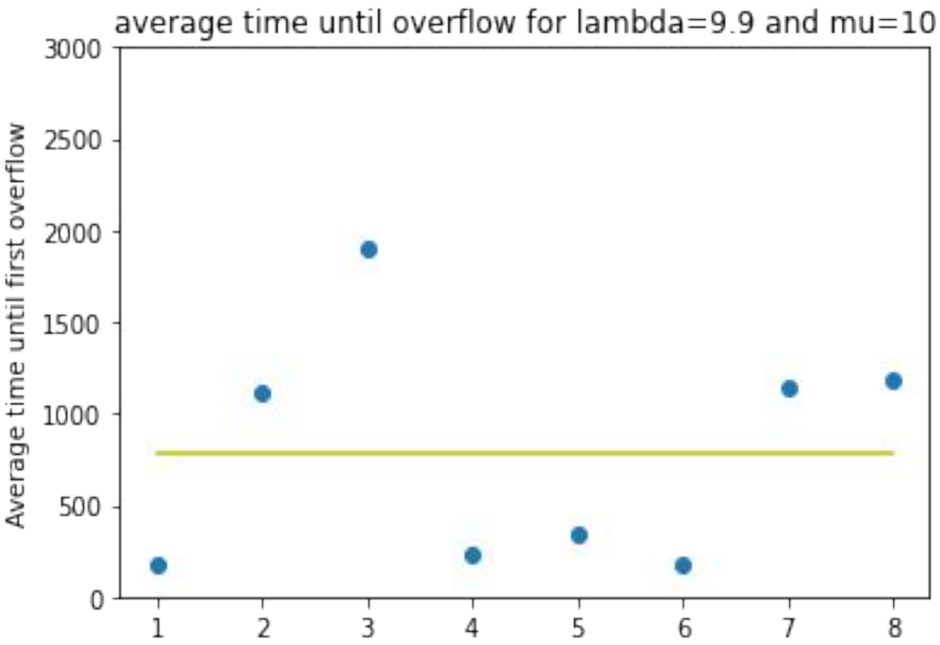
Buffer Overflow for λ = 9.7 and μ = 10

- The individual points represent the times at which the queue overflowed for the first time for different simulation.
- The horizontal line represents the average of the overflow times of these simulations.



Buffer Overflow for λ = 9.9 and μ = 10

- The individual points represent the times at which the queue overflowed for the first time for different simulation.
- The horizontal line represents the average of the overflow times of these simulations



Overflow times as ʎ approaches μ

- the average time until first overflow decreases with ʎ value coming close to μ.

- this behaviour is expected as the increased arrival rate congest the queue faster and probability of a packet dropped from the queue increase at lower times.

Summary and Conclusions

Summary

- We understood the basic performance metrics of queuing systems like mean length of queue/system, mean waiting time in a queue/system.
- We studied Nburst/M/1 model which takes into consideration of the burstiness of the packet arrivals in the network.
- We studied about self-similarity and see why an LTE network has self-similarity.
- We calculated Hurst parameter to determine the degree of self similarity

Summary

- We analyse a data set consisting of network trace at 4G cell stations which is collected over a certain time period and we use Variance-time plot to determine the Hurst parameter of the network traffic, this is done for several datasets.
- We implemented M/M/1 queue model in ns3 simulator and verified the previously studied results by plotting idle time proportion of the queue, mean length of the queue, mean delay of the packet for varying arrival rates.
- We also saw buffer overflow results for M/M/1 queue

Conclusions

- Since the calculated Hurst parameters for the analysed LTE dataset are close to 0.8 which is greater than 0.5, we proved that the LTE network traffic is self-similar ,and hence bursty in nature.

- So we can conclude that Nburst/M/1 model is a suitable model to analyse LTE networks.

Thank You