STUDYING PERFORMANCE OF LTE NETWORKS WITH QUEUING SYSTEMS

Prem Sujan(170123027)
&
Garvit Mehta(170123018)

Project Guide: Prof. N. Selvaraju

Objective

The main aim of the project is to study various Queuing Models in the context of LTE Networks and establish various performance metrics for LTE Networks. We will study NBurst/M/1 queuing model which has a close resemblance to practical LTE networks. First, we establish the results of M/M/1 model for the LTE network and further we simulate the actual network traffic using NBurst/M/1 model to confirm them.

Background of the problem

Long Term Evolution (LTE) started as a project in 2004 by Third Generation Partnership Project (3GPP) to provide technology focusing on packet-switched data.

This explosive growth of mobile data traffic and adoption of mobile connectivity by end users is increasing of global 4G LTE deployments and adoption.

Importance of the problem

While end users and application developers are expecting the networks to handle the high traffic rate, the heterogeneous nature of the traffic poses a challenge to understand the performance of the network.

So performance analysis of the network is crucial to improvements of the network.

Previous studies in the area

- Analysis of the real LTE network systems
- LTE network simulation using M/M/1 queuing model
- LTE network simulation using M/M/c queuing model

Queuing systems

Queuing systems deal with the mathematical study of congestion.

Generally, the study of servicing requests from customers/entities arriving in a queue fashion at a facility is called queuing theory.

Examples: Ticket booking, customer service centers.

Some mathematical terms

Markov Property

Poisson Process

Arrival rate

Service rate

Little's Law

Kendall notation

Traffic rate, L_s, L_q, W_s, W_q

Markov Property

A stochastic process has the Markov property if the conditional probability distribution of future states of the process (conditional on both past and present states) depends only upon the present state, not on the sequence of events that preceded it.

$$P(G > a + b | G > a) = P(G > b)$$

Exponential random variable follows markov property.

Also referred to as memoryless property.

Poisson process

The counting process $N(t)$, $t \in [0,\infty)$ is called a Poisson process with rate λ if it follows the condition

- $N(0)=0$
- $N(t)$ has independent increments
- The number of arrivals in any interval of length $\tau>0$ has Poisson($\lambda\tau$) distribution.

$$\mathbb{P}\{N(t)=n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Arrival rate

The rate at which customers/entities arrive at service stations with a request.

Units: $(T)^{-1}$

Service rate

The rate at which servers fulfill the requests produced by the customers.

Units: $(T)^{-1}$

Little’s Law

The fundamental relationship between the three parameters λ(arrival rate), L(length of the system), W(waiting time in the system)

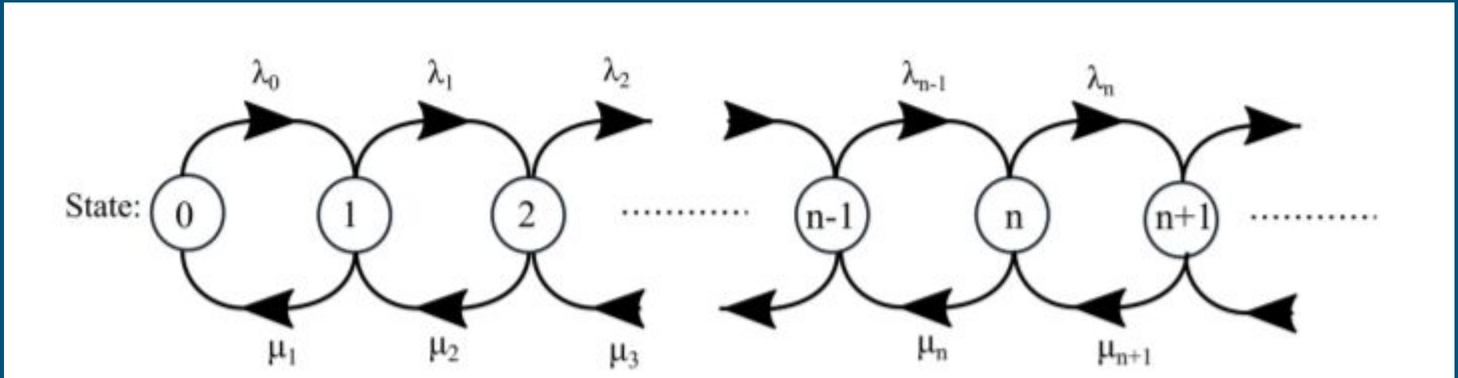
$$L = \lambda W$$

This follows from the Law of large numbers.

Birth death process

Birth death process is a variation of Markov process with discrete space where the system can transitions only to its neighbouring states as shown in the figure.

Transition to a lower state is referred to as death, and transition to a higher state is referred to as a birth.



Queuing models

Queuing models are mainly characterised by the process of arrival, the service mechanism, and queuing discipline.

Arrivals can be batch, continuous, poisson distributed, etc.

Queuing discipline can be FCFS, LCFS, Priority queuing, etc.

Different combination of these produce different analytical models.

We use a notation to describe queuing models

Kendall’s notation

Kendall notation is a standard system used in queuing systems to describe a queuing model with three factors A/S/c. Sometimes also referred to as A/S/c/K/N/D.

A: Arrival process.

S: Servicing process.

c: Number of service channels.

K: Length of the queue.

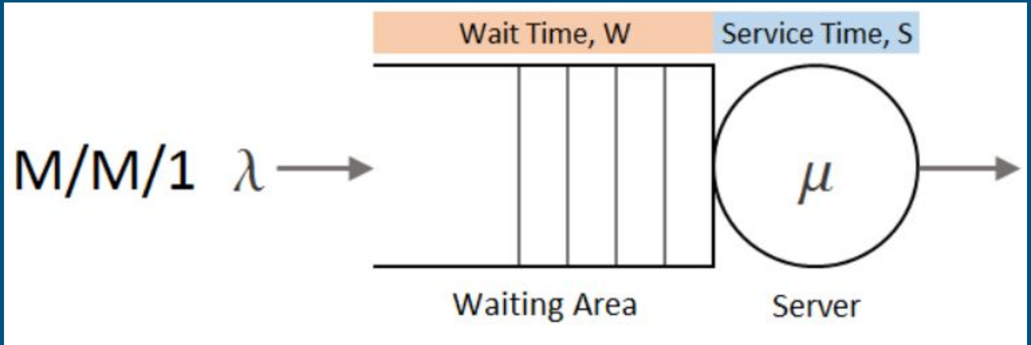
N: Customer population

D: Service discipline

M/M/1 queue

A simple single server queue system with interarrival time and service time following exponential distributions is called M/M/1 queue.

This is a special case of birth-death process with state independent transition rates.



Steady state probabilities

The probability of the system to stay in state n at time t is denoted by $P_n(t)$

From the birth-death process state diagram we can see that at time t+h (very small h), to stay in the same state n, there are 3 cases as follows

1. From state n-1 with one request and no service
2. From state n+1 with no request and one service
3. Remain in state n with no request and no service

   From poisson distribution we have $P(N(h)=1) = \lambda h$

The 3 cases when written mathematically,

$$P_n(t + h) = P_{n-1}(t)\lambda h(1 - \mu h) + P_{n+1}(t)(1 - \lambda h)\mu h + P_n(t)(1 - \lambda h)(1 - \mu h) \text{ for } n>0$$

$$(\lambda + \mu)P_n = P_{n+1}\mu + P_{n-1}\lambda \qquad (1)$$

$$P_0(t + h) = P_1(t)(1 - \lambda h)\mu h + P_0(t)(1 - \lambda h)$$

$$\lambda P_0 = \mu P_1 \qquad (2)$$

From (1) and (2) we have

$$P_n = (\lambda/\mu)^n P_0$$

Steady state probabilities

$$P_n = \rho^n P_0$$

An important parameter of the queueing system is traffic rate $\rho = \lambda/\mu$

Sum of steady state probabilities is

$$P_0 + P_1 + P_2 \ldots P_\infty = 1$$

$$P_0 = 1 - \rho$$

$$P_n = (1 - \rho)\rho^n$$

We can see that the $\lambda/\mu < 1$ for a positive steady probability for state n

Performance metrics(L_s,L_q,W_s,W_q)

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

Performance metrics (contd.)

$$L_s = \sum_{i=0}^{\infty} i P_i$$

L_s=(ρ/1-ρ)

$$L_q = \sum_{i=1}^{\infty} (i-1) P_i$$

L_q=(ρ^2/1-ρ)

Performance metrics (contd.)

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

$L_q = L_s\rho$ $\qquad$ $W_q = W_s\rho$

Why M/M/1 model fails for LTE systems

LTE system traffic originates from several terminal devices which request and upload data independently.

And the traffic generated by each device is further dependent on the application type used.

So we need a novel queue model to study LTE systems.

N-Burst Model for LTE Networks

Need For a New Model

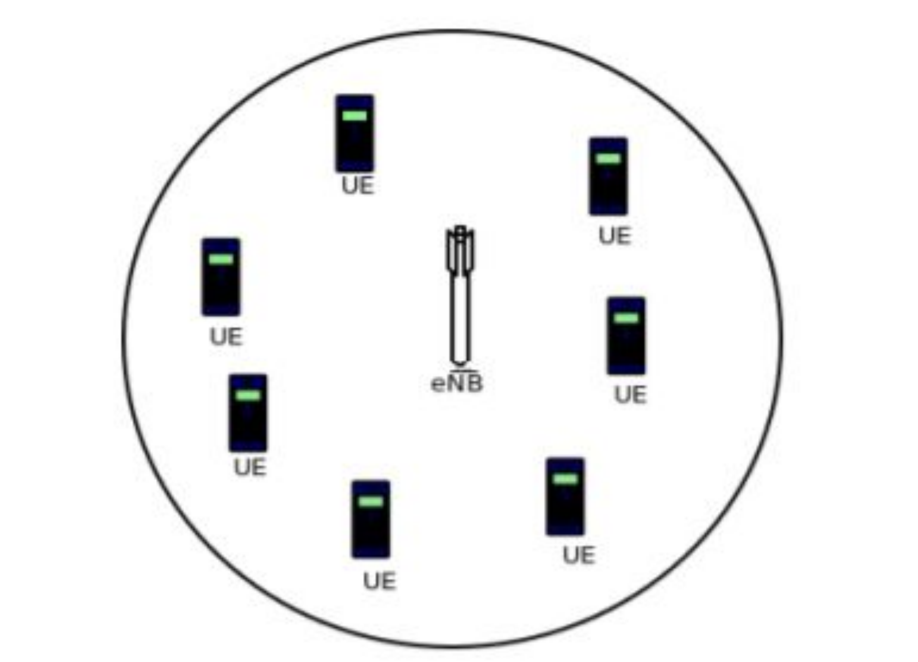
- There is an incredible growth in the number of wireless devices such as smart-phones, tablets and Internet of Thing (IoT) .
- In addition to the fast development of media streaming applications, IPTV, telemedicine and Internet gaming have led to a significant challenge to the design and deployment 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 come in bursts rather than continuously.
- 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.

LTE Network Topology

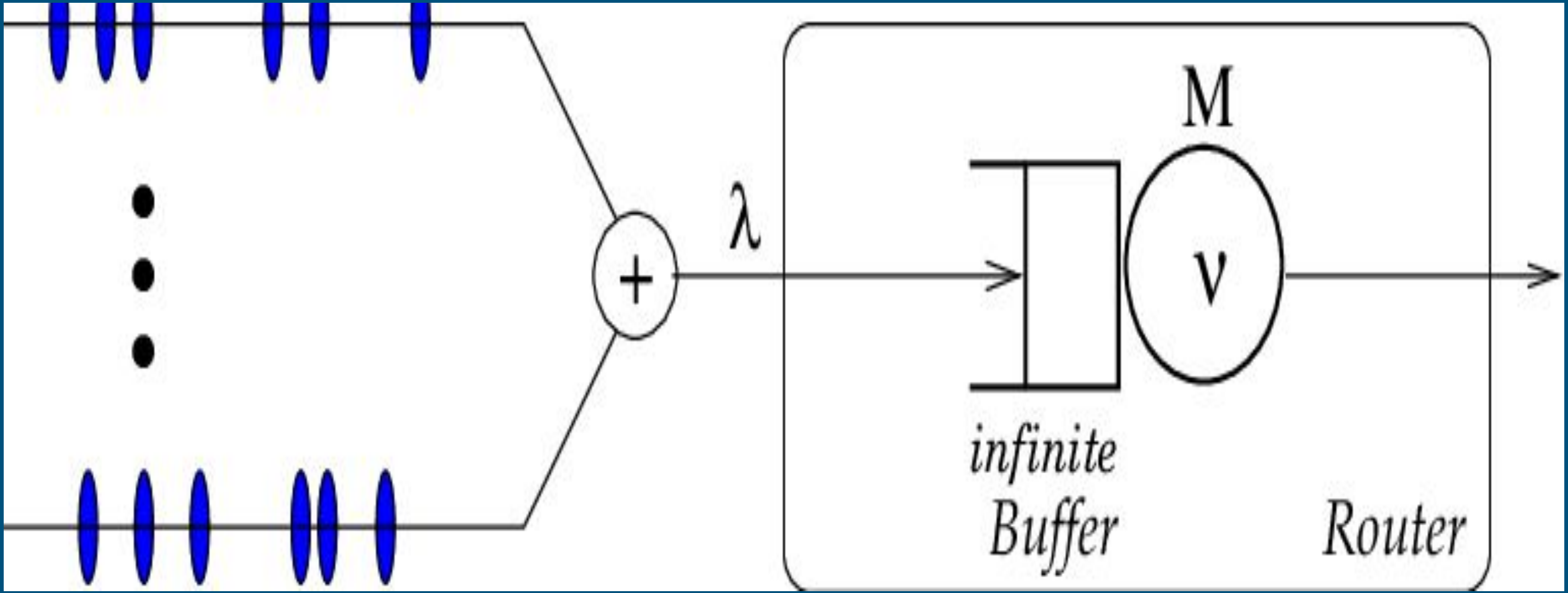
- A simple LTE network consists of a evolved nodeB, which provides service to a set of N wireless devices.
- These devices use the LTE technology for downloading data (down-link) and upload and request data (up-link).



Describing Network Queue

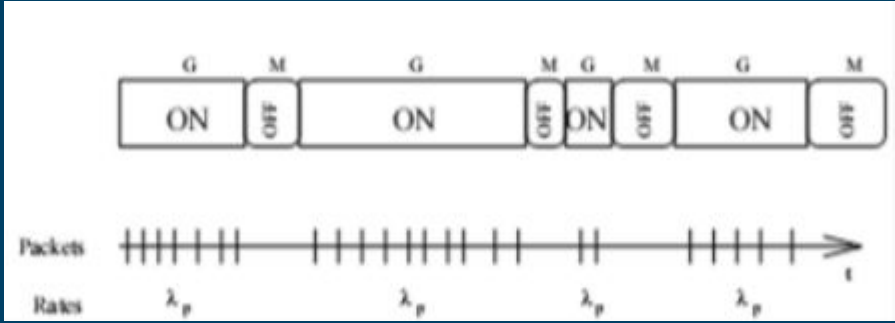
- The number of packets requested by a mobile device is considered as a random variable.
- The properties and distribution of this random variable is purely based on the mobile application used to request the data.
- This is the reason standard queuing models are not appropriate to understand telecommunication systems.

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



Description of Model

- 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.
- λ: The overall arrival rate that is produced by N sources i.e. $\lambda = \kappa N$
- λ_p: the peak transmission rate during a burst i.e. at ON time.
- λ_b: Mean burst arrival rate $=\lambda/N_b$.
- N_b: Average number of packets during a burst (at ON period)
- ON : the average time during which the node is active
- OFF : the average time during which the node transmission is OFF.

Essential Parameters of the Model (contd...)

- μ : the mean service rate of the router
- ρ = λ/μ : Load utilisation of the router
- The *Burstiness Parameter* b is defined below

$$b = \frac{\overline{OFF}}{\overline{ON} + \overline{OFF}} = 1 - \frac{\kappa}{\lambda_p} = \frac{\lambda}{N\lambda_p}$$

Since λ(the amount of data sent per unit time), and ρ can be held constant as b is varied over 0 to 1. So the parameter b can be considered as a shape parameter.

Two Simple Cases for b

Case 1 : b=0

- When b=0, it means that average time during which the node transmission is OFF is 0.
- In this case all the bursts come after one another and ON/OFF process reduces to a simple one.

Case 2 : b=1

- In this case, the average time during which burst comes is zero.
- It also means that all the packets in the burst arrive simultaneously i.e. They arrive as a bulk process.

Distribution of Sub-Processes

Considering the 1-Burst Model i.e. N=1. This model depends on four separate distributions , with random variables denoted by X_{ON}, X_{OFF} , X_{SV} and X_{INP} .

- **ON:** On time distribution have mean $\underline{ON}$. Therefore the mean number of packets during a burst is given by $N_b = \underline{ON}.\lambda_p$
- **OFF**: OFF time distribution have mean $\underline{OFF}$ (It will depend on how often the bursts are generated)
- **SV:** Service time distribution has mean 1/μ . (It will depend on distribution of packet sizes, service rate depends on packet size and router speed)
- **INP**: Inter packet time distribution is the distribution during a burst whose mean is $1/\lambda_p$

Limiting Case Analysis for b->0

- We can conclude that the the transmission of packets can occur at the slowest rate at ON times when OFF = 0 or b=0 ($\lambda_p = \kappa$).
- In this case the distribution of OFF time is irrelevant.
- ON distribution also does not have an impact on the system.
- Thus for b -> 0 the 1-Burst process reduces from $SM/G_{SV}/1$ queue to a $G_{INP}/G_{SV}/1$ queue. (X_{OFF} ,X_{INP} and X_{ON} are the distributions on which SM depends where SM denotes Semi Markov)

Limiting Case Analysis for b->1

- If we increase λ_p indefinitely we get b -> 1 i.e. $\underline{ON}$ -> 0
- It also means that all the packets in a given burst arrive simultaneously
- Also called *BULK ARRIVAL.*
- The bulk size is distributed according to ON-time distribution whose mean is N_b.
- The time in between two bursts is distributed according to X_{OFF}.
- 1-Burst process reduces to a $G_{(OFF)}^{[ON]}/G_{SV}/1$ queue where $G_{(OFF)}^{[ON]}$ represents the bulk arrival process whose Inter Arrival times are X_{OFF} and bulk sizes proportional to X_{ON}.

Concluding Remarks

Our base model should be simple, so we take X_{OFF} , X_{SV} and X_{INP} as exponential distributions. The ON distribution will be varied to obtain different results. Our model have parameters from each of these models.

$b \to 0$ the system becomes an $M_\lambda/M_\mu/1$ queue

$b \to 1$ the system becomes an $M_{\lambda_b}{}^{[(ON)]}/M_\mu/1$ queue

Delay Calculation For b=0

- Although, the analytic N-Burst model can be evaluated for all possible distributions,the calculations are not trivial.
- So we give mean packet delay for b=0

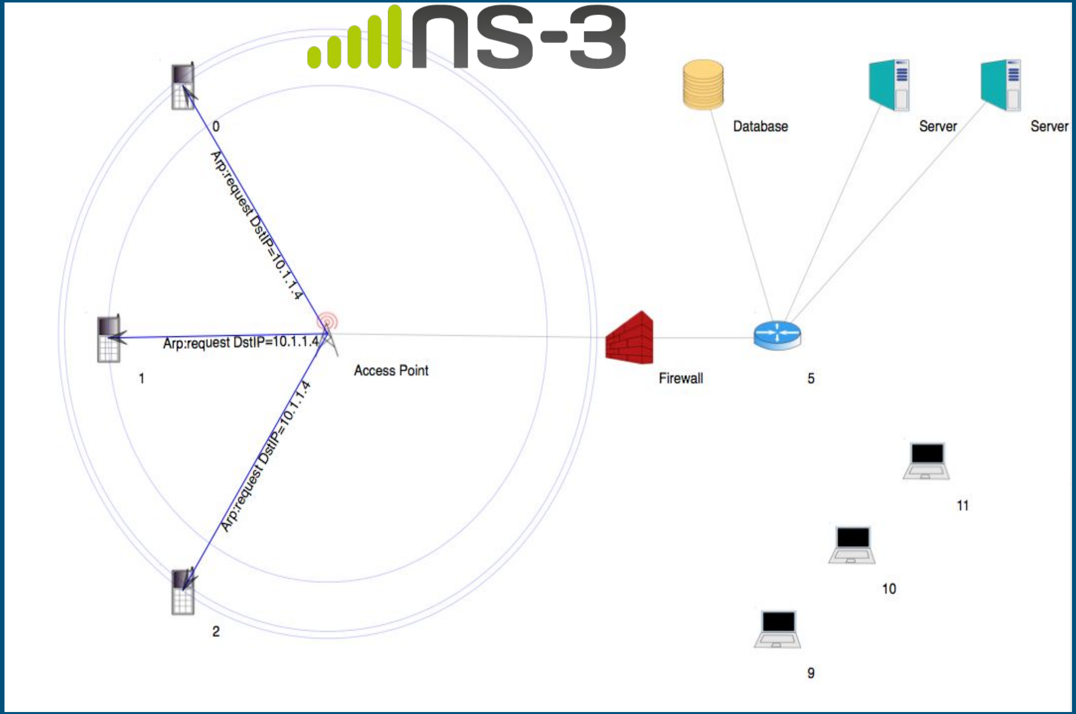
- *mdb0 represents mean packet delay for b=0.*

$$mdb0 = \frac{1/\mu}{1-\rho}$$

NS-3 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.



Future work

- Simulating M/M/1 model for LTE networks on NS3 with varying parameters to analyse packet loss, delay, throughput to confirm if they follow the model characteristics.
- Define LTE network topology and simulate a more appropriate model, NBurst/M/1 on NS3 with varying ON time distribution to analyse the performance of the network system.

Thank You