MM1 Queuing & VOIP

TOBENNA UZOMA, IHEGBOROW
A00267346

Aim: The aim of this experiment is to determine if the average time a packet waits in the queue exceeds the acceptable time limits of 20 seconds. Also, to observe the behaviour of the queue size, that is, if it increases monotonically or settles down. Finally, to generate a graphical representation of the waiting time on the link utilization on the network.

Objectives:

The objectives of this report will be broken down into three sections;

Part A

- Assemble and view the different types of statistics.
- Understand the process of applying filters to retrieved data during simulation.
- Mathematically scrutinize the statistical data from the simulation.

Part B:

- Creating a graphical representation of the Link Utilisation (Rho) in relation to the waiting time to indicate the existence of congestion within the network.

Part C:

- Answering the questions on VOIP traffic source.

Introduction/Backgrounds:

To begin this report, we need to discuss the Queuing Theory. Queuing Theory is the study of waiting lines or queues mathematically. It has been considered a branch of operations research because the outputs are used when making business decisions about resources needed to provide a service (Queueing Theory, n.d.).

The **M/M/1 queue** is a simple model where a single server serves jobs that arrive according to a Poisson process and have exponentially distributed service requirements (Queueing Theory, n.d.). In other places, it is known to be a stochastic process whereby the state spaces are the sets {0,1,2,3,4,5….} where the value equates to the amount of customers in the system, also adding any present in service (M/M/1 Queue, n.d.).

The **Link Utilization** is the amount in percentage of a networks bandwidth this is being used up by the network traffic (Hewlett-Packard Company, 1998). The application of queueing theory involves the network system capacity (C), which represents the optimum rate of work achievable. It is compared against 'R' which represent the average rate at which work is demanded of any system. If R > C, an indication that capacity was insufficient would be evident however if R < C then a basic fundamental rule would exist that the system would operate normally. Observation of link utilisation factor (Rho) in relation to Peak Packet Rate "R" and Service Rate "C" may be an indication that congestion exists within a network (Stewart, 2018).

Theoretical Link Utilisation (p) is expressed as per Eq (38):

$$Utilization\ (p) = \frac{\alpha . R}{C} = \frac{0.35\ x\ 167}{73} = 0.8 \quad (38)$$

Such that "α" is the activity factor between T(on) and T(off):

$$\alpha = \frac{T(on)}{T(on) + T(off)} \quad (39)$$

VOIP (Voice Over Internet Protocol) is also commonly known as Phone Service over the internet. If you have a reasonable quality Internet connection you can get phone service delivered through your Internet connection instead of from your local phone company. Some people use VOIP in addition to their traditional phone service, since VOIP service providers usually offer lower rates than traditional phone companies, but sometimes doesn't offer 911 service, phone directory listings, 411 service, or other common phone services. While many VoIP providers offer these services, consistent industry-wide means of offering these are still developing (What is VOIP, n.d.).

Finite State Machine (FSM), other times called a Finite State Automation, is a computational model that can be applied with hardware or software. It can also be applied to the stimulation of sequential logic and programs for computers. It creates regular languages. FSM can be used for the replication of problems in numerous fields – mathematics, artificial intelligence and linguistics. There are basically two types of FSMs - Deterministic and Non-deterministic finite state machines. There are slight differences between the way machines are displayed visually but the general idea behind them drives through an identical computational logic. Deterministic FSM identifies or allows regular languages. FSMs are often learned using languages comprising of binary strings that shadow a particular pattern. An example of a binary string language is: the language of all strings that have a 0 as the first character. In this language, 001, 010, 0, and 01111 are valid strings (along with many others), but strings like 111, 10000, 1, and 11001100 (along with many others) are not in this language (Moore, 2016).

Method:

In this experiment, I used the Node and Project editors in OPNET to build an M/M/1 queue system, retrieved data about the model, ran the simulation and studied the results.

- **Part A**
  - Opened a new file and created a Node Model.
  - Used the processor module to Define the Source model
  - Specified the processor's generation rate, distribution, average packet size, and packet size distribution
  - Created a queue model
  - Used the sink module to destroy packets when they are no longer needed
  - Connected the modules by packet streams
  - Saved the work

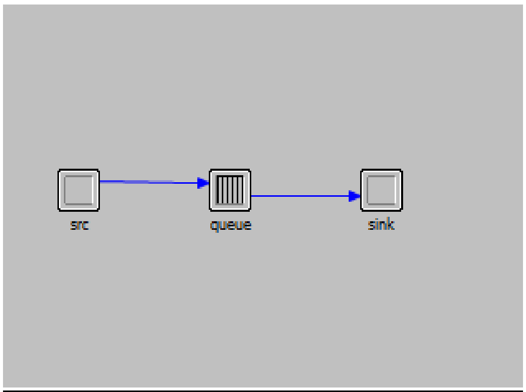


Figure 1: illustration of the node model

  - Set up a MM1 model by creating a project and scenario
  - Constructed a model list
  - Created the network model
  - Collected and compared the results

- **PART B**
  - Modified the service rate to have 5 variations in link utilization results between 0.4 and 0.9
  - Saved the changes
  - Generated a graph showing the 5 different simulation results of the waiting time against the corresponding link utilization value.

Theoretical Result

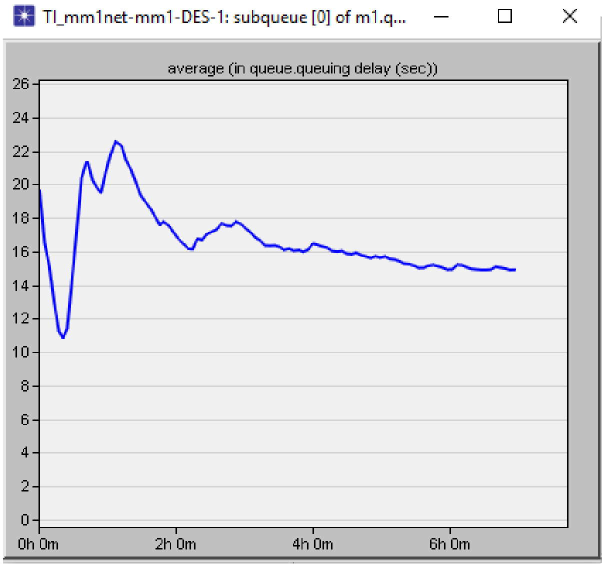


Figure 2: Queue size & delay

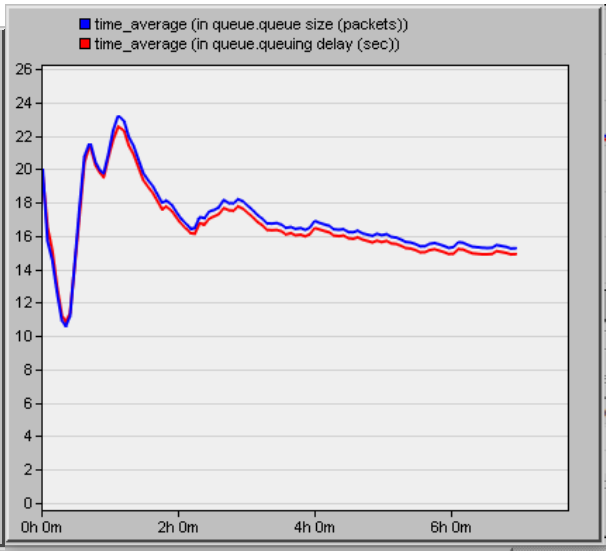


Figure 3: Average(queue-queue delay)

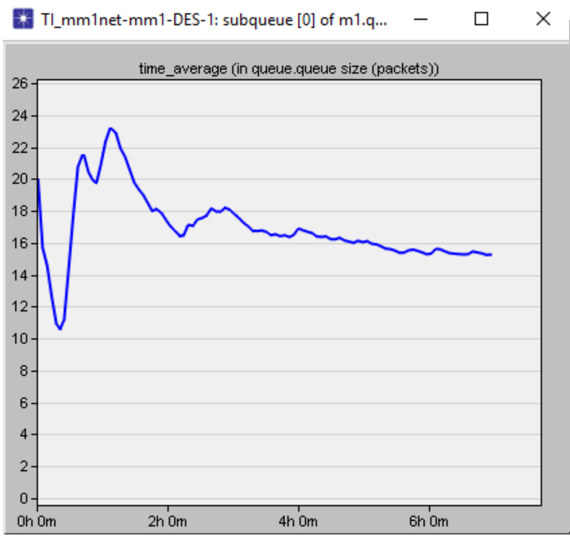


Figure 4: time average(queue-queue size)

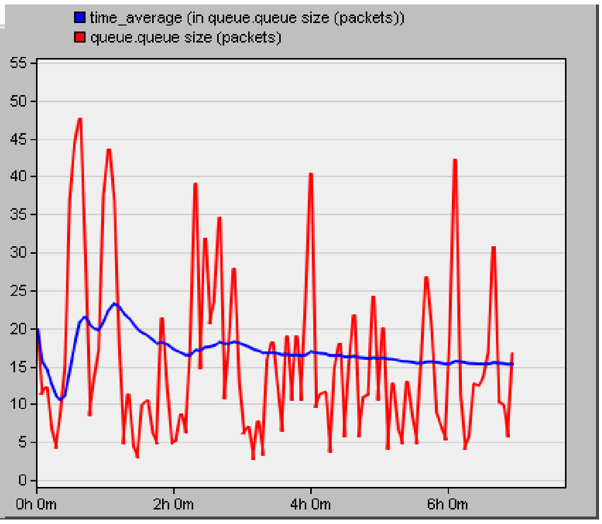


Figure 5: Queue size, Time-averaged and instantaneous

Queue Delay Calculations

Mean arrival rate:

$$I = \frac{1.0}{mean\ interarrival\ time} = \frac{1p}{1.0s} = 1(pEs)$$

Mean service requirement:

$$\frac{1}{m} = 9000(bEp)$$

Service capacity:

C = 9600(bEs)

Mean service rate:

$$mC = \approx \frac{1}{\ll 9000(b\ Ep)} \frac{\div}{y} \left(9600\ (b\ Es)\right) @1.067(p\ Es)$$

Mean Delay:

$$\overline{W} = \frac{1}{mC-1} = \frac{1p}{1.067(p\mathrm{E}s)-1(p\mathrm{E}s)} = \frac{1}{0.067(p\mathrm{E}s)} = 15s$$

Simulation Results

Calculate the Link Utilization Measurements range between 0 and 1

$Where\ \rho(\mathcal{R}ho) = \frac{\lambda}{\mu} \rightarrow \frac{Lambda}{Mu}$

$$\rho(\mathcal{R}ho) = \frac{9000}{9600} \quad = \quad 0.94$$

$$\rho(\mathcal{R}ho) = \frac{9000}{10000} \quad = \quad 0.90$$

$$\rho(\mathcal{R}ho) = \frac{9000}{11000} \quad = \quad 0.81$$

$$\rho(\mathcal{R}ho) = \frac{9000}{12000} \quad = \quad 0.75$$

$$\rho(\mathcal{R}ho) = \frac{9000}{13000} \quad = \quad 0.69$$

Table 1: Link Utilization and Waiting time

Link Utilization	Delay
0.94	15
0.9	9
0.81	4.5
0.75	3
0.69	2.5

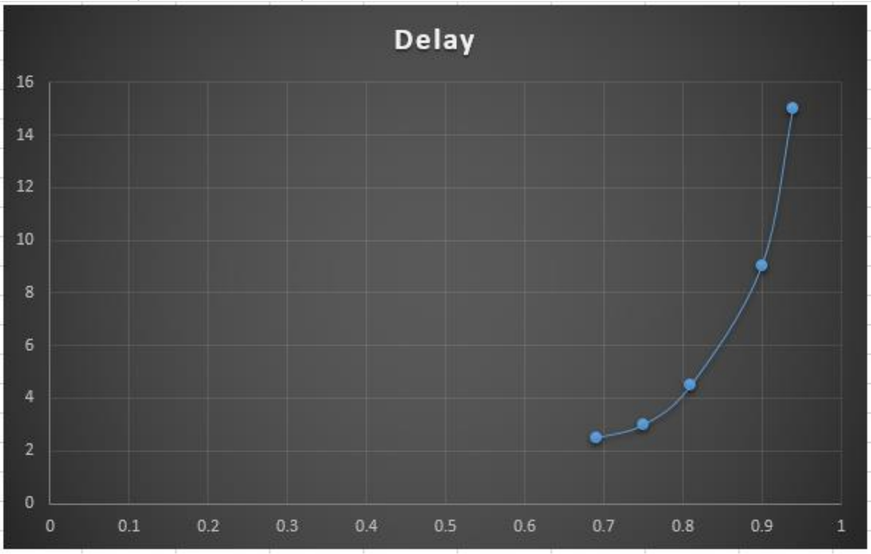


Figure 6: Waiting time vs Link Utilization

Discussion

- **Part A:**
  This part was basically concerned with building an MM1 queue system and comparing the average time a packet is delayed in the buffer (queue delay) with the average number of packets occupied in the queue (queue size). The average time a packet should be held in the queue must be at 20 seconds and the queue size has to increase monotonously before eventually levelling out showing that the server is not being overloaded. To verify this, I set the simulation runtime to 7 hours and set the seed to 195.

  Viewing the statistical result on the queue delay from *Figure 2* in the theoretical result section, we can observe that the network initially spiked above the average threshold limit which is 20 seconds and then reduced settling down at about 15 seconds which is an acceptable limit. The big differences previously in the simulation confirms the sensitivity of averages to the relatively little amount of samples gathered before eventually balancing out in the end.

  Again, looking at the statistical result on the time-averaged queue size from *Figure 4* in the theoretical result section, we can observe that the queue size increased monotonously initially before coming down to settle at about 5 hours. This clearly signifies that the server is performing at optimum capacity and is not being overloaded.

  The graph on *Figure 3* confirms the similarity of the average queue delay with the time-averaged queue size. This is as a result of the evenly spread queue injections and transfers that happened during the simulation of the model. The contrast between the end values of both observations are unimportant.

  The last graph in *Figure 5* shows views the instantaneous and the time-averaged queue size on the same graph. During the simulation, the queue size displays the instantaneous amount of packets in the queue while the time-averaged queue size displays the average amount of packets in the queue over a period of time. It can be concluded from here that the time-averaged queue size never exceeded the threshold limit of 20 seconds. Also the queue increases monotonously and then levels out at around 5 hours, indicating a very stable and functioning system.

- **Part B**
  This part was focused on showing the graphical representation of the waiting time on the link utilisation (Rho). In this experiment the mean size of each packet produced by the server is 9000 bits per second and the mean interval time between each packet is 1 second. The service rate setting is 9600 bits per second, hence the Link Utilization is 0.94.

  I then set up 4 different Link Utilization values where I changed the service rate values to 10000, 11000, 12000 and 13000 bits per second. The corresponding waiting times when observed in the global statistics were as follows – 9, 4.5, 3 and

2.5 seconds from *Table 1* in the simulation results above. The Link Utilization is below 1 because the network should never run at full capacity. If the network runs at the fullest capacity 1 and there is a burst of traffic, the servers will be overloaded and it will result in congestion and packet loss. As a result, for optimum network performance, the Link Utilization should be between 0.6 and 0.7 seconds and should never fall below 0.4 seconds. Looking at the graph in *Figure 6,* as the Link Utilization increases the delay on the network also increases.

Now, looking deeper into the code in detail, at the **Function Block** (FB) and the **Head-Of-Line Block** (HB), "*Op-pk-create-fmt(format-str)*" generates a new formatted packet with a structure already established in advance by the mentioned packet format. "format-str" is the format structure, which basically indicates the structure of the packet format that will be assigned to the packet. "*Op-pk-send(pkptr, ssc-strm-to-low)*" on the other hand, sends the mentioned packets through an output packet stream, plans the packets entry at the destination module for the current simulation time and frees ownership of the packet by supplicating process. "*pkpt*" indicates the packet of interest.

Conclusion

The node model was used to create an MM1 queue system where the average queue delay and the time-average queue size were compared. The time-average queue size did not pass the 20 seconds threshold, nor did the average queue delay grow monotonously without levelling out. It levelled out at 5 hours which indicated that the server was not being overloaded.

Also this report also proves that as the Link Utilization of the network increases, so also does the waiting time delay but all within a range of 0 and 1.

Part C

1. Is it possible to take traces of real traffic from a live network and model it in Opnet. How?
   Yes, this can be done on Opnet by a process known as Simulation Modelling. Simulation modelling is a fundamental stage in the evolution of any network deployment. The advantage of using simulation is the ability to analyse the network modifications pertaining to the real-world scenarios. Utilizing a model to determine network behaviour can be achieved using two main methods; Analytical & Simulation (Stewart, 2018).

2. How are simultaneous events handled in Opnet?
   Events that have been scheduled for the same simulation time are simulated sequentially, although they look like they occur synchronously. The simulation kernel engages one of two techniques to verify execution order;

   Natural Order technique: The first event to arrive in the list is handled first.
   Priority Factor: Modules and Events are given priority factors. In other words, events of higher priorities are handled before lower priority events.

3. Show and explain an appropriate “process model” for a VoIP traffic source. Using a suitable diagram, detail how such a model would operate.
   A VOIP source mimics behaviours of an ON/OFF source model and demonstrates the way traffic emanates from the end user in the form of packets generated at the lowest level. A multimedia or real-time application can generate “burst” traffic characteristics and dynamically change bandwidth requirements such as compressed video applications. Other real-time applications require constant bandwidth such as compressed videos (Stewart, 2018).

   Different Traffic types
   VBR (variable bit rate) traffic characteristics of “bursty” traffic flows are in contrast to *CBR (constant bit rates)* traffic streams generated by constant timed sensory real-time applications.

   Validation and Verification
   Validation of the source model quantifies the accuracy of the model to depict the given situation; however, verification determines if the simulation model has accurately been transposed to a computer program.

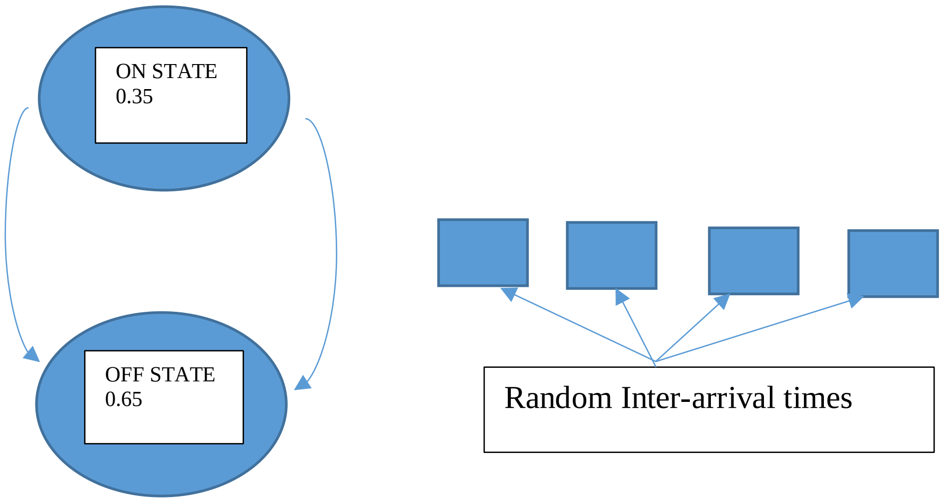


Figure 7: Illustration of Randomization of Packet Generation from source model

Application of the ON/OFF Source Model may be validated by the theoretical utilization of the public switched Telephone Network (PSTN) which produces traffic at a rate of 64kbps.

Based on the generation of a constant fixed medium size packet of 384bits at 64kbps, results in 167 packets generated per second.

<u>Utilization</u>

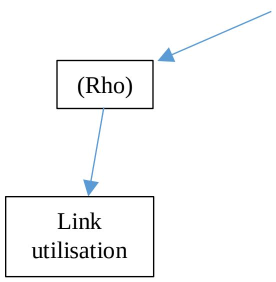
Average "ON" time of 0.35 for the active state
Average "OFF" time of 0.65 for the non-active state
Results in a Packet Rate (PCR) of 58 packets per second transmission.

<u>With a Utilization factor of 80%</u>

A single buffer is required to deal with 73 packets per second.

<u>Analytical Results</u>

The calculation applied to the simulation model with a utilization value of 0.8(80%) based on the ON/OFF source packet generation



$$P = \frac{\lambda}{u} = \frac{(384 \times 167) \times 0.35}{28{,}056 \text{ bits}} = 0.8$$

Packet size 384
Packet Cell Generation Rate = 167pks based on
Link Capacity = 64kbps ATM Cell link
Service Rate (capacity) = (22,444.8 ÷ 0.8)
= **28,056bit per sec**

Bibliography

Hewlett-Packard Company. (1998). *Utilization*. Retrieved from Hewlett-Packard Company: http://h17007.www1.hpe.com/device_help/HPJ3298A/utilization.htm

M/M/1 Queue. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/M/M/1_queue

Moore, K. (2016, April 1). *Finite State machines*. Retrieved from Brilliant: https://brilliant.org/wiki/finite-state-machines/

Queueing Theory. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Queueing_theory

Stewart, J. (2018). Assignment VOIP Source model [Recorded by J. Stewart]. Athlone, Westmeath, Ireland.

What is VOIP. (n.d.). Retrieved from Voip-info.org: https://www.voip-info.org/what-is-voip/