International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-3, Issue-12, December 2015

Simulation and Comparison of Markovian Exponential Queuing Models

L. N. Onyejegbu

***Abstract*— We all have experienced the discomfort of waiting in a queue. Unfortunately, this phenomenon is becoming increasingly common in urban societies with increasing population. One of the problems of Automated Teller Machine (ATM) machines is the long queue, which results in customer dissatisfaction. In this work, i determined the queuing model that will reduce customers' waiting time in an ATM facility. This was achieved, by simulating and comparing Markovian exponential queuing models: M/M/1 and M/M/S. In the notation, the M stands for Markovian; M/M/1 means that the system has a Poisson arrival process, an exponential service time distribution, and one server. While M/M/S is Markovian Poisson input, exponential service time model with s server. Queuing model M/M/S was recommended with two objectives of minimizing customer waiting time and percentage of idle time for the ATM. Simulating M/M/S using several sample sizes with our program (ATM Queue Simulator) produced the shortest waiting time and fastest service time when compared with simulation of M/M/1 queuing model. C++ programming language was used to implement this work.**

***Index Terms*— Queuing model, M/M/I, M/M/S**

I. INTRODUCTION

Queues (waiting lines) are a part of everyday life. We all wait in queues to buy a movie ticket, make a bank deposit, pay for groceries, mail a package, obtain food in a cafeteria, start a ride in an amusement park, etc. We have become accustomed to considerable amounts of waiting, but still get annoyed by unusually long waits.

The amount of time that a nation's populace wastes by waiting in queues is a major factor in both the quality of life there and the efficiency of the nation's economy. For example, before its dissolution, the U.S.S.R. was notorious for the tremendously long queues that its citizens frequently had to endure just to purchase basic necessities. Even in the United States, it has been estimated that Americans spend 37,000,000,000 hours per year waiting in queues. If this time could be spent productively instead, it would amount to nearly 20 million person- years of useful work each year [1].

Even this staggering figure does not tell the whole story of the impact of excessive waiting. Great inefficiencies also occur because of other kinds of waiting than people standing in line. For example, making machines wait to be repaired may result in loss of production. Vehicles (including ships and trucks) that need to wait to be unloaded may delay subsequent shipments. Airplanes waiting to take off or land may disrupt later travel schedules. Delays in telecommunication transmissions due to saturated lines may cause data glitches. Causing manufacturing jobs to wait to be performed may disrupt subsequent production. Delaying service jobs beyond their due dates may result in loss of future businesses.

Queuing theory is the study of waiting in all these various guises. It uses queuing models to represent the various types of queuing systems (systems that involve queues of some kind) that arise in practice. Formulas for each model indicate how the corresponding queuing system should perform, including the average amount of waiting that will occur, under a variety of circumstances.

Therefore, these queuing models are very helpful for determining how to operate a queuing system in the most effective way. Providing too much service capacity to operate the system involves excessive costs. But not providing enough service capacity results in excessive waiting and all its unfortunate consequences. The models enable finding an appropriate balance between the cost of service and the amount of waiting [2].

II. REVIEW OF RELATED LITERATURE

Queuing theory is a branch of mathematics that studies and models the act of waiting in lines. The theory of queues was initiated by the Danish mathematician A. K. Erlang, who in 1909 published "The theory of Probabilities and Telephone Conversation". He observed that a telephone system was generally characterized by either (1) Poisson input (the number of calls), exponential holding (service) time, and multiple channels (servers), or (2) Poisson input, constant holding time and a single channel. Erlang was also responsible in his later works for the notion of stationary equilibrium and for the first consideration of the optimization of a queuing system.

Applications of the theory to the telephony were soon appearing. In 1927, E. C. Molina published "Application of the Theory of Probability to Telephone Trunking Problems", and one year later Thornton Fry printed "Probability and its Engineering Uses" which expand much of Erlang's earlier work. In the early 1930's Felix Pollaczeck did some further pioneering work on Poisson input, arbitrary output, and single and multiple channel problems. Other names working in the same field during that period included Kolmogorov and Khintchine in Russia, Crommelin in France and Palm in Sweden. The work in queuing theory picked up momentum rather slowly in its early days, but in 1950 started to accelerate and there have been a great deal of work in the area since then. [3]. In this work we worked and simulated M/M/1 and M/M/s queuing models

L. N. Onyejegbu, **CPN, NCS, IEEE, TWOWS** Department of Computer Science, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria

III. QUEUING MODELS

Historically queuing theory began in 1913 with the work of A.K. Erlang on telephone traffic. In this work Erlang sought to answer such questions as how many telephone circuits and operators are required to satisfy a given demand. Applications to manufacturing, however, began largely with the later work of J.R. Jackson (1963) which outlines the now well-known Jackson queuing network. A solution for queue length probability distribution is provided for job shop-like queuing networks. External arrivals enter the first workstation according to the Poisson distribution and are subsequently routed either to the next process with probability *ij p* or out of the system with probability1− *pij* . Processing times are also Poisson, and there is infinite buffer capacity. Queue discipline must not rely on future routing or service time information, and thus is considered first come first serve (FCFS). Utilization of any station should not exceed 100%. Under these conditions, a product form solution exists stating that the probability that the network as a whole will be in a state, defined by the number of jobs waiting at each queue, is simply the product of the probabilities of each queue individually having said number of jobs waiting. However, the limitations of the necessary conditions led researchers to seek out adaptations of the method so as to reflect more realistic systems [4].

There are different types of Queuing models; Markovian single server queuing model M/M/1, Markovian Poisson input, exponential service time model with s servers M/M/s. Poisson input, general service time model with 1 server M/G/1. Poisson input, Erlang service time model with 1 server $M/E_k/1$. There is also M/M/C/K, queuing model where first M represents Markovian exponential distribution of inter-arrival times, second M represents Markovian exponential distribution of service times, C (a positive integer) represents the number of servers, and K is the specified number of customers in a queuing system. This general model contains only limited number of K customers in the system. However, if there are unlimited numbers of customers, it means K = Q, then our model will be labeled as M/M/C [5].

IV. M/M/1 QUEUING MODEL FORMULA

Here we will show how to model a single-queue single-server system with our ATM. In the notation, the M stands for Markovian; M/M/1 means that the system has a Poisson arrival process, an exponential service time distribution, and one server. Queuing theory provides exact theoretical results for some performance measures of an M/M/1 queuing system and this model makes it easy to compare empirical results with the corresponding theoretical results. [6].

$$L_s = \frac{\lambda}{\mu - \lambda} \qquad W_q = \frac{\lambda}{\mu(\mu - \lambda)} \qquad P_o = 1 - \frac{\lambda}{\mu}$$

$$P_{n>k} = \left(\frac{\lambda}{\mu}^{k+1}\right) \quad L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} \quad W_s = \frac{1}{\mu-\lambda} \quad \rho = \frac{\lambda}{\mu} \tag{1}$$

V. MODEL B (M/M/S): MULTIPLE-CHANNEL QUEUING MODEL

Multiple-channel queuing system is one in which two or more servers or channels are available to handle arriving customers. We still assume that customers awaiting service form one single line and then proceed to the first available server. Multichannel, single-phase waiting lines are found in many banks today: A common line is formed, and the customer at the head of the line proceeds to the first ATM.

The multiple-channel system presented assumes that arrivals follow a Poisson probability distribution and
that service times are exponentially distributed. Service is first-come, first-served, and all servers are assumed to perform at the same rate. Other assumptions listed earlier for the single-channel model also apply [7].

Table I. Parameters used in M/M/1 simulation

Where λ = mean number of arrivals per time period.
μ = mean number of people served per time period.
L_s = average number of units (customers) in the system (waiting and being served)
w_s = average time a unit spends in the system (waiting time plus service time)
L_q = average number of units waiting in the queue
W_q = average time a unit spends waiting in the queue
ρ = utilization factor for the system
P_o = probability of 0 units in the system (that is, the service unit is idle)
$P_{n>k}$ = probability of more than k units in the system, where n is the number of units in the system.

IV. M/M/S QUEUING MODEL FORMULA

The multiserver queue M/M/s is the model used most in analyzing service stations with more than one server such as banks, checkout counters in stores, checkin counters in airports and the like. The arrival of customers is assumed to follow a Poisson process, service times are assumed to have an exponential distribution and let the number of servers be *s* providing service independently of each other. We also assume that the arriving customers form a single queue and the one at the head of the waiting line gets into service as soon as a server is free. No server stays idle as long as there are customers to serve. Equation 2. gives M/M/S Formula.

$$L_s = L_q + \frac{\lambda}{\mu} \quad P_{n>k} = \left(\frac{\lambda}{\mu}\right)^{k+1} \quad L_q = L_s - \frac{\lambda}{\mu} \quad W_q = \frac{L_q}{\lambda} \quad W_s = \frac{L_s}{\lambda} \quad \rho = \frac{\lambda}{\mu} \quad P_o = \frac{1}{\sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!(1-\lambda/(s\mu))}}$$

(2)

Table II. Parameters used for M/M/S simulation

Where λ = mean number of arrivals per time period.
μ = mean number of people served per time period.
L_s = average number of units (customers) in the system (waiting and being served)
w_s = average time a unit spends in the system (waiting time plus service time)
L_q = average number of units waiting in the queue
W_q = average time a unit spends waiting in the queue
ρ = utilization factor for the system
n = the number of customers in queuing system
s = number of servers (channels)
P_o = probability of 0 units in the system (that is, the service unit is idle)
$P_{n>k}$ = probability of more than k units in the system, where n is the number of units in the system.

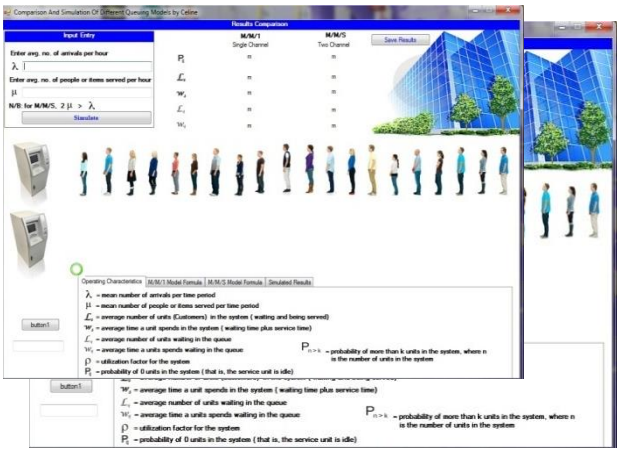

Fig 1. Behaviour of customers in the ATM
Figure 1 shows the behaviour of customers in the ATM

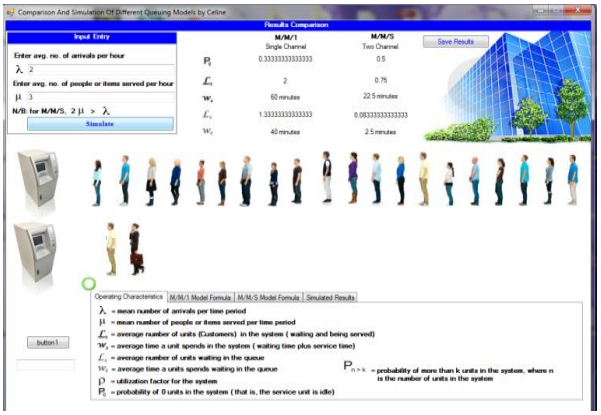

Fig 2: Simulation of M/M/1 and M/M/S

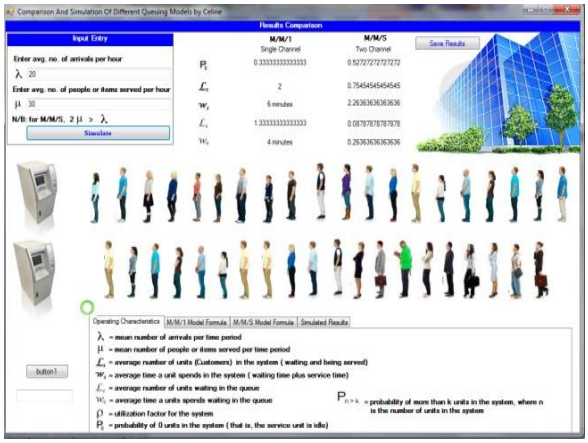

Fig 3: Simulation of M/M/1 and M/M/S with a large sample size

Fig 4.: Simulation of M/M/1 and M/M/S with minimal sample size

			M/M/1						M/M/S				
S/No	λ	μ	P_0	$L_{s(Hr)}$	$W_{s(Min)}$	L_q (Hr)	$W_{q(Min)}$		P_0	$L_{s(Hr)}$	$W_{s(Min)}$	$L_{q(Hr)}$	$W_{q(Min)}$
1	2	3	0.33	2	60	1.33	40		0.5	0.75	22.5	0.08	2.5
2	5	7	-0.3	2.5	30	1.79	21.43		0.5	0.82	9.89	0.11	1.32
3	10	12	-0.2	5	30	4.17	25		-0.45	1.03	6.15	0.19	1.15
4	20	25	-0.2	4	12	3.2	9.6		-0.47	0.97	2.9	0.17	0.5
5	18	20	-0.1	9	30	8.1	27		-0.43	1.16	3.86	0.26	0.86

Table III: Simulated results of M/M/1 and M/M/S queuing models
Table III is the simulated result from the model.

Fig 5: Graphical Representation of the time spent in M/M/1 and M/M/S Queuing Model

RESULTS DISCUSSION

I have been able to observe customers arrival time, waiting time in the queue, different behaviour of customers in the queue like balking, reneging, jockeying and service time with ATM machine. This Was observed for 2 months. Generally, arrivals do not occur at fixed regular intervals of times but tend to be clustered for a duration of one week. The Poisson distribution involves the probability of occurrence of an arrival at random and independent of all other operating conditions. The inter arrival rate (i.e., the number of arrivals per unit of time) λ is calculated by considering arrival time of the customers to that of the number of customers. Service time is the time required for completion of a service that is, it is the time interval between beginning of a service from ATM machine and its completion. I have calculated mean service time μ of customers by considering different service time for customers to that of the number of customers.

Based upon the tabulation and taking one day as a standard, I inferred that during weekday's prime hours there is heavy crowd in the bank ATMs. Which implies that the utilization factor is 1? It is vivid that the ATM is 100% utilized by the customers. In the non- busy hours, utilization factor is 50% for the bank. In weekend period the utilization factor is 62% for the Bank. The comparison between the waiting time in the queue and the system, by using simulation, shows more variation because the study was undergone with the observation of minimum number of customers with minimum duration. This study also reveals that the waiting time in queuing model (M/M/1) is more than that of Queuing Model (M/M/S).

REFRENCES

[1] M. A. Crane, and D. L. Iglehart, "Simulating stable stochastic systems, General Multiserver Queues," ACM 21, 2002, pp 103-105.

[2] S. S. Lavenberg, and D. R. Slutz, "Regenerative simulation of a queuing model of an automated tape library", 2010, pp 23-26.

[3] A. Dasgupta, and M. Ghosh, "Including Performance in a Queue via Prices: The Case of a Riverine Port." Journal of Management Science 46(11). 2000, pp 1466–1484.

[4] S. S. Lavenberg, "Efficient estimation via simulation of work rate in closed queueing networks", Proceedings in Computational Statistics, Physica Verlag, Vienna, Austria, 2003, pp 35-62.

[5] W.Whitt, "Predicting Queuing Delays:" journal of Management Science 45(6). 2002, pp 870–888.

[6] D. S. Hira, and P. K. Gupta, "Simulation and Queuing Theory," Operation Research, S.Chand and Company Ltd., 2004. New Delhi.

[7] Fishman, G. S. "Estimation in Multiserver Queuing Simulations," Oper. Res. 22, 2005, pp 72.