International Journal of Mathematics And its Applications

Volume 5, Issue 4–D (2017), 595–603.

ISSN: 2347-1557

Available Online: http://ijmaa.in/

International Journal of Mathematics And its Applications

Single and Multi Server Queuing Models: A Study

Research Article

Tariq Ahmad Koka[1]*, V.H.Badshah[1] and Riyaz Ahmad Shah[2]

1 School of Studies in Mathematics, Vikram University, Ujjain, Madhya Pradesh, India.

2 Department of Mathematics, Government Degree College, Kulgam, J & K, India.

Abstract: This paper describes queuing system and queuing networks which are successfully used for performance analysis of different systems such as computer, communications, transportation networks and manufacturing. Queues or waiting lines are a common phenomenon in everyday life. The essence of this phenomenon is the low efficiency of queuing system. In this paper we treat elementary queuing models. Attention is paid to methods for analysis of these models and also to application of queuing models. The queuing number, the service windows number, and the optimal service rate are investigated by means of queuing theory. The time of customer in a waiting line is reduced. The customer satisfaction is increased. By illustration, it is shown that the results are effective and practical.

Keywords: Queuing system, single server model, arrival rate, service rate, infinite and finite models.

1. Introduction

Most elementary queuing models assume that the inputs and outputs follow a birth and death process. Here the inputs mean arrivals and outputs mean departures. Any queuing model is characterized by situations where both arrivals and departures take place simultaneously. Understanding waiting lines or queues and learning how to manage them is one of the most important areas in operations management. In certain cases, a service system is unable to accommodate more than the required number of customers at a time. No further customers are allowed to enter until space becomes available to accommodate new customers. Such types of situations are referred to as finite (or limited) source queue. Examples of finite source queues are cinema halls, restaurants, etc. On the other hand, if a service system is able to accommodate any number of customers at a time, then it is referred to as infinite (or unlimited) queue. For example in a sales department here the customer orders are received; there is no restriction on the number of orders that can come in so that a queue of any size can form. Zhao [14] studied that, the China's commercial banks have done great efforts to increase the marketing, but most of them are facing a serious problem is customer queuing, which led to low service rate of the bank counters, poor business environment and number of high quality customers and potential customers are lost and so on. Yuejian [13] generalized that, the checkout stands were the service windows of supermarkets, which not only reflect supermarkets images but also associate with supermarkets service quality and business efficiency. Dharmawirya and Adi [3] gave a definition of queuing theory that, the queuing theory is the study of queues or waiting lines.

Elegalam [4] studied that the customers waiting for long time in the queue could become a cost to them. Caues and Cauas [6] were studied that, in general queues form when the demand for service exceeds its supply. Researchers, Brann and Kulick

[*] *E-mail: tariqkoka1920@gmail.com*

[1], Curin et. al. [2], Kharwat [7] have previously used queuing theory to the restaurant operation, to reduce cycle time in a busy fast food restaurant, as well as to increase throughput and efficiency. Vikas [12] mentioned two basic costs these are, cost associated with patients or customers having to wait for service (Waiting Cost) and service cost is the cost of providing service .These includes salaries paid to employees or servers while they wait for service from other servers. Xiao and Zhang [5] proved that a single line is better than more lines.

In the paper Vijay, Badshah and Koka [11] proved that, the single queue multi server model is better than multi queue multi server model and generalized the mathematical relations of the performance measures of both queuing models. The expected number for customers waiting in the queue (L_q) is less in the case of single queue as compare with case of s queues. The expected waiting time of the customer in the queue (W_q) is less in the case of single queue as compared to case of s queues. The expected waiting time of the customer in the system (W_s) is less in the case of single queue as compared to case of s queues. The expected number of customers waiting in the system (L_s) is greater in the case of single queue as compared to multiple queues. In our present paper we show that single queue multi is better than single queue single server.

Procedure for solution.

(1). List the alternative queuing system.

(2). Evaluate the system in terms of various times, length and costs.

(3). Select the best queuing system.

2. Single Server Queuing Models

2.1. (M/M/1(/FCFS) or M/M/1 -∞/∞ Model)

Infinite Queue Length Model (Exponential service-Unlimited Queue)

This model is based on certain assumptions about the queuing as:

(1). Arrivals are described by Poisson probability distribution and come from an infinite population.

(2). Single waiting line and each arrival waits to be served regardless of the length of the queue and no balking and reneging take place.

(3). Queue discipline is first come first serve

(4). Single server and service time follows exponential distribution.

(5). Average service rate is more than average arrival rate $\lambda < 1$

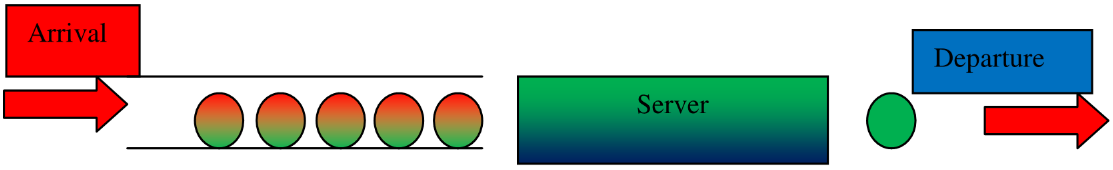

Figure 1. **Single Queue-Single server model**

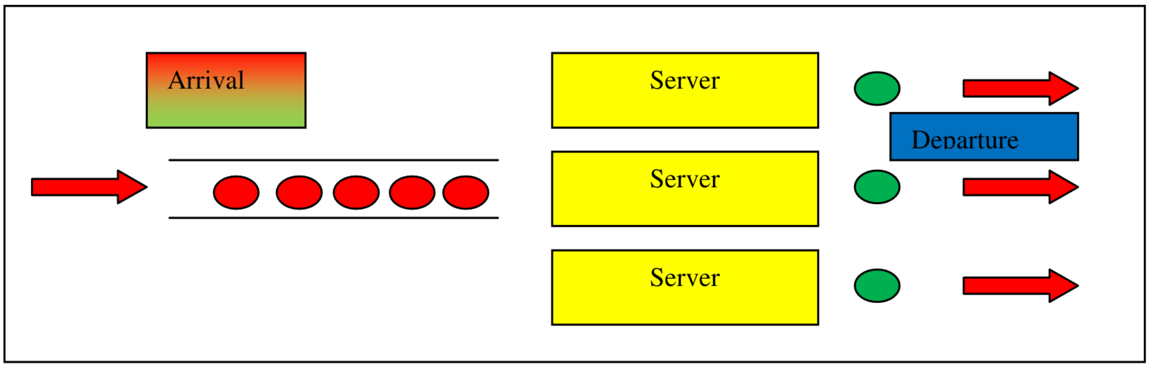

Figure 2. **Single queue Multi server model**

2.2. Case study

We know that $\frac{\lambda}{\mu} < 1$, particularly in infinite queue length models. Let $P_0, P_1, P_2 \ldots, P_\infty$ be the probabilities of zero person, one person, and two people and so on. From the customer point of view, the customer is ordinarily interested in four types of parameters.

(1). Length of system, L_s

(2). Length of queue, L_q

(3). Waiting time in the system, W_s

(4). Waiting time in the queue, W_q

Suppose h is the small interval of time, then we have $P_n(t+h) = P_{n-1}(t)$ (Probability of one arrival and no service) $(\lambda h) + P_{n+1}(t)$ (Probability of one service and no arrival) $(\mu h) + P_n(t)$ (Probability no arrival and no service) that is

$$P_n(t+h) = P_{n-1}(t)(\lambda h)(1-\mu h) + P_{n+1}(t)(\mu h)(1-\lambda h) + P_n(1-\lambda h)(1-\mu h) \tag{1}$$

Leaving the higher order terms, we get,

$$\begin{aligned} P_n(t+h) &= P_{n-1}(t)(\lambda h) + P_{n-1}(t)(\mu h) - P_n(t)(\lambda+\mu)h + P_n(t) \\ \Rightarrow \frac{P_n(t+h)-P_n(t)}{h} &= P_{n-1}(t)\lambda + P_{n+1}(t)\mu - P_n(t)(\lambda+\mu) \end{aligned} \tag{2}$$

Applying steady state condition, the rate of change of P_n with respect to interval h is zero. At steady state the probability is not going to be time independent, thus the L.H.S of equation (2) is zero.

$$\begin{aligned} 0 &= P_{n-1}(t)\lambda + P_{n+1}(t)\mu - P_n(t)(\lambda+\mu) \\ \Rightarrow \lambda P_{n-1} + \mu P_{n+1} &= P_n(\lambda+\mu) \end{aligned} \tag{3}$$

Again from (1). Put $n = 0$ we have

$$\begin{aligned} P_0(t+h) &= P_1(t) \;\; \text{one service and no arrival} + P_0(t) \quad \text{no arrival no service} \\ &= P_1(t)(1-\lambda h)(h\mu) + P_0(t)(1-\lambda h)(1) \end{aligned} \tag{4}$$

[Since P_0 means the system has already zero people and probability of no service is$1-0=1$] Leaving the higher order terms we have,

$$\frac{P_0(t+h)-P_0(t)}{h}=\mu P_1(t)-\lambda P_0(t)$$

Applying steady state condition we get,

$$\begin{aligned}\mu P_1-\lambda P_0&=0\\ \Rightarrow \mu P_1&=\lambda P_0 \qquad (5)\\ \Rightarrow P_1&=\left(\frac{\lambda}{\mu}\right)P_0\end{aligned}$$

Putting $n=1$ in (3) we get

$$\begin{aligned}\lambda P_0+\mu P_2&=P_1(\lambda+\mu)\\ \Rightarrow \lambda P_0+\mu P_2&=\lambda P_1+\mu P_1\\ \Rightarrow \lambda P_0+\mu P_2&=\lambda P_1+\mu\left(\frac{\lambda}{\mu}\right)P_0\\ \Rightarrow \lambda P_0+\mu P_2&=\lambda P_1+\lambda P_0\\ \Rightarrow \mu P_2&=\lambda P_1\\ \Rightarrow P_2&=\frac{\lambda}{\mu}P_1=\frac{\lambda}{\mu}\left(\frac{\lambda}{\mu}\right)P_0 \quad \text{(using equation (3))}\\ \Rightarrow P_2&=\left(\frac{\lambda}{\mu}\right)^2P_0\end{aligned}$$

Let $\frac{\lambda}{\mu}=\rho$. Then $P_1=\rho P_0$ and $P_2=\rho P_1=\rho^2P_0$, $P_n=\rho P_{n-1}=\rho^nP_0$. It is clear all the probabilities are dependent on P_0. Since we have

$$\begin{aligned}P_0+P_1+P_2+\cdots+p_\infty&=1\\ \Rightarrow P_0+\rho P_0+\rho^2P_0+\infty&=1\\ \Rightarrow P_0(1+\rho+\rho^2+\cdots+\infty)&=1\\ \Rightarrow P_0\left(\frac{1}{1-\rho}\right)&=1 \quad \text{(sum of a G.P)}\\ \Rightarrow P_0&=1-\rho \qquad (6)\end{aligned}$$

This is an important equation for $M/M/1-\infty/\infty$ Model. Therefore

$$P_j=\rho^jP_0=\rho^j(1-\rho);\quad j=0,1,2,\ldots$$

Calculations of L_s, L_q, W_s, W_q.

$$\begin{aligned}L_s&=\sum_{j=0}^{\infty}jP_j\\ &=\sum_{j=0}^{\infty}j\rho^jP_0\\ &=\rho P_0\sum_{j=0}^{\infty}j\rho^{j-1}\end{aligned}$$

$$
\begin{aligned}
&= \rho P_0 \sum_{j=0}^{\infty} \frac{d}{d\rho}(\rho^j)\\
&= \rho P_0 \frac{d}{d\rho} \sum_{j=0}^{\infty}(\rho^j)\\
&= \rho P_0 \frac{d}{d\rho}(1+\rho+\rho^2+\cdots+\infty)\\
&= \rho P_0 \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)\\
&= \rho P_0 \frac{1}{(1-\rho)^2}
\end{aligned}
$$

But $P_0 = 1-\rho$. Therefore,

$$L_s = \frac{\rho(1-\rho)}{(1-\rho)^2} = \frac{\rho}{(1-\rho)} \tag{7}$$

Now $L_s = L_q +$ expected number of people being served.

$$
\begin{aligned}
&\Rightarrow L_s = L_q + \frac{\lambda}{\mu}\\
&\Rightarrow L_q = L_s - \frac{\lambda}{\mu}
\end{aligned}
$$

Also we have, $L_s = \lambda W_s$ and $L_q = \lambda W_q$. These are called Little's equations.

2.3. Problem Section

Let $\lambda = 8/hr$ and $\mu = 9/hr$. Therefore

$$P_0 = 1-\rho = 1-\frac{8}{9} = 0.1111$$

It tells us that 11.11% of the times there is nobody in the system which implies that the server is free. So the probability that the server is not free is $1-P_0 = 1-0.1111 = 0.8889$. Probability that there is no queue.

$$
\begin{aligned}
&= P_0 + P_1\\
&= P_0 + \rho P_0\\
&= P_0(1+\rho)\\
&= 0.2098.
\end{aligned}
$$

This implies that about 21% of the times there is no queue. Probability that there are 10 people in the system is given by

$$
\begin{aligned}
P_{10} &= \rho^{10} P_0\\
&= (0.889)^{10}(0.111)\\
&= 0.0342
\end{aligned}
$$

That is 3.42% of the times there will be 10 people in the system when one is being served and remaining nine are waiting in the queue. Probability $(n \geq 2) = P_2 + P_3 + \cdots + P_\infty = 1-(P_0+P_1) = (1-21\%) = 79\%$. This implies that 79% of the times there will be two or more people in the system. Thus

$$L_S = \frac{\rho}{1-\rho} = \frac{8/9}{1-8/9} = 8$$

That is expected number of people in the system including the one which is being served is 8. Also

$$L_q = L_s - \frac{\lambda}{\mu} = 8 - \frac{8}{9} = 7.1111$$

and $W_s = \frac{L_s}{\lambda} = 1\ hour$; $W_q = \frac{L_q}{\lambda} = 7.1111/8 = 0.8889\ hour = 53\ min$.

2.4. M/M/1-N/∞ Model

In this model we have a restriction on the length of queue. Let $P_0, P_1, P_2 \ldots, P_N$ be the probabilities of zero person, one person, and two people and so on. Then

$$\begin{aligned}
P_0 + P_1 + P_2 + \cdots + P_N &= 1 \\
\Rightarrow P_0 + \rho P_0 + \rho^2 p_0 + \cdots \rho^N P_0 &= 1 \\
\Rightarrow P_0(1 + \rho + \rho^2 + \cdots \rho^N) &= 1 \\
\Rightarrow \frac{P_0 1(1-\rho^{N+1})}{1-\rho} &= 1 \quad \text{(Sum of a G.P. to N terms)} \\
\Rightarrow P_0 &= \frac{1-\rho}{1-\rho^{N+1}}
\end{aligned}$$

Since $P_n = \rho^n P_0$ $(n = 0, 1, 2, \ldots, N)$

$$\begin{aligned}
L_s &= \sum_{n=0}^{N} nP_n \\
&= \sum_{n=0}^{N} n\rho^n P_0 \\
&= \rho P_0 \sum_{n=0}^{N} n\rho^{n-1} \\
&= \rho P_0 \sum_{n=0}^{N} \frac{d}{d\rho}(P^n) \\
&= \rho P_0 \frac{d}{d\rho}(1 + \rho + \rho^2 + \cdots + \rho^N) \\
&= \rho P_0 \frac{d}{d\rho}\left(\frac{1-\rho^{N+1}}{1-\rho}\right) \\
&= \frac{\rho P_0}{1-\rho}\{-(1-\rho)(N+1)\rho^N + 1 - \rho^{N+1}\}
\end{aligned}$$

But,

$$\begin{aligned}
P_0 &= \frac{1-\rho}{1-\rho^{N+1}} \\
L_s &= \frac{\rho P_0}{1-\rho}\{1 + N\rho^{1+N} - (1+N)\rho^N\} \\
&= \frac{\rho(1-\rho)}{(1-\rho^{N+1})(1-\rho)^2}\{1 + N\rho^{1+N} - (1+N)\rho^N\} \\
&= \frac{\rho}{(1-\rho^{N+1})(1-\rho)}\{1 + N\rho^{1+N} - (1+N)\rho^N
\end{aligned}$$

The term $1 - \rho^{N+1}$ keeps coming because we have a restriction on N. Thus arrival rate of the customer is the effective arrival rate that is $\lambda_{eff} = \lambda(1 - \rho^N)$. Therefore, $L_s = \lambda_{eff} W_s$ and $L_q = \lambda_{eff} W_q$.

2.5. Illustration

Let $\lambda = 8/hr$, $\mu = 9/hr$ and $N = 10$

$$\begin{aligned}
P_0 &= \frac{1-\rho}{1-\rho^{N+1}} \\
&= \frac{1-8/9}{1-(8/9)^{11}} = 0.1530
\end{aligned}$$

So probability that there is no queue is given by

$$\begin{aligned} P_0 + P_1 &= P_0 + \rho P_0 \\ &= P_0(1+\rho) \\ &= 0.1530\left(1+\frac{8}{9}\right) = 0.2890 \end{aligned}$$

Probability that there are ten people in the system is

$$P_{10} = \rho P_0 = \left(\frac{8}{9}\right)^{10}(0.1529) = 0.0471$$

From this result we can say that the probability that someone does not join the queue is 0.0471 and probability that someone does join the queue is $10-(0.0471) = 0.9529$. So $\lambda_{eff} = \lambda(09529) = 7.6231 \leq 8$. Hence

$$L_s = \frac{\rho}{(1-\rho^{N+1})(1-\rho)}\{1+N\rho^{1+N}-(1+N)\rho^N\}$$

3. Model (Multi Server Queuing System)

The different authors like Taha [10], Gupta and Hira [8] and Sharma [9] generalized the M/M/S model. This model is based on certain assumptions about the queuing as:

(1). Arrivals are described by Poisson probability distribution and come from an infinite population.

(2). Multiple waiting line and each arrival waits to be served regardless of the length of the queue and no balking and reneging take place.

(3). Queue discipline is first come first serve.

(4). Service time follows exponential distribution.

(5). Average service rate is more than average arrival rate.

Here μ is the service rate of one server, but here we have c number of servers therefore $c\mu$ will be the service rate. If there are n customers in the queuing system at any point in time, then following two cases may arise:

(i). If $n < c$ (number of customers in the system is less than the number of servers), then there will be no queue. However, $(c-n)$ numbers of servers are not busy. The combined service rate will be: $\mu_n = n\mu$; $n < c$.

(ii). If $n = c$, (number of customers in the system is more than or equal to the number of servers) then all servers will be busy and the maximum number of customers in the queue will be $(n-c)$. The combined service rate will be: $\mu_n = c\mu$; $n = c$. Thus, we have $\lambda_n = \lambda$ for all $n = 0$

$$\mu_n = \begin{cases} n\mu, & n < c \\ \mu_n = c\mu, & n = c \end{cases}$$

The probability P_n of n customers in the queuing system is given by

$$P_n = \begin{cases} \frac{\rho^n}{n!}P_0, & n \leq c; \\ \frac{\rho^n}{c!c^{n-c}}P_0, & n > c. \end{cases}$$

$$P_0 = \frac{1}{\left[\sum\limits_{n=0}^{c-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{c!}\left(\frac{\lambda}{\mu}\right)^c \left(\frac{c\mu}{c\mu-\lambda}\right)\right]}$$

Expected number of customers waiting in the queue (i.e. queue length)

$$L_q = \sum_{n=c}^{8} (n-c)\, P_n \Rightarrow L_q = \left[\frac{1}{(s-1)!}\left(\frac{\lambda}{\mu}\right)^s \frac{\lambda\mu}{(s\mu-\lambda)^2}\right] P_0$$

Expected number of customers in the system

$$L_s = L_q + \frac{\lambda}{\mu}$$

Expected waiting time of a customer in the queue

$$W_q = \left[\frac{1}{(s-1)!}\left(\frac{\lambda}{\mu}\right)^s \frac{\lambda\mu}{(s\mu-\lambda)^2}\right] P_0 = \frac{L_q}{\lambda}$$

Expected waiting time that a customer spends in the system.

$$W_s = W_q + \frac{1}{\mu}$$

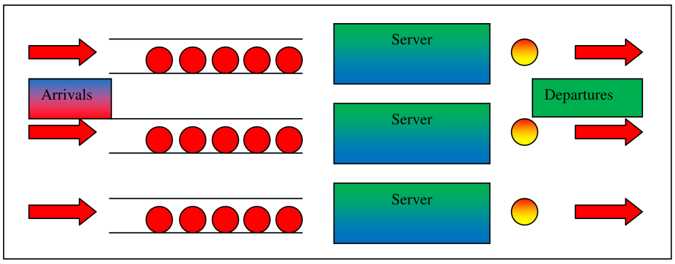

Figure 3. **Multi queue multi server model**

4. Conclusion

This paper gives a general look at the queuing theory. These models can be used to improve flow of customers, for evaluation of utilization throughput and response times. The most important information required to solve a waiting line problem is the nature of probability distribution of arrivals and service pattern. The important thing is that the ratio $\frac{\lambda}{\mu}$ must be less than one $\left(\frac{\lambda}{\mu} < 1\right)$ particularly in infinite queue length model but in finite queue length model we are not worried about the ratio $\frac{\lambda}{\mu}$ being greater than one. In general we find that Multi server queuing models is more beneficial than single server model though a bit costly.

References

[1] D.M.Brann and B.C.Kulick, *Simulation of Restaurant Operations using the Restaurant Modeling Studio*, Proceedings of the 2002 Winter Simulation Conference, IEEE Press, (2002), 1448- 1453.

[2] S.A.Curin., S.V.Jeremy, W.C.Erich and T.Omer, *Reducing Service Time at a Busy Fast Food Restaurant on Campus*, Proceedings of the 37th conference on Winter Simulation, IEEE Press, (2005), 2628-2635.

[3] M.Dharmawirya and E.Adi, *Case Study for Restaurant Queuing Model*, Conference on Management and Artificial Intelligence, Bali, Indonesia, 6(2011), 52-55.

[4] Elegalam, *Customer Retention versus Cost Reduction technique*, A Paper Presented at the Bankers Forum held at Lagos, (1978), 9-10.

[5] Huimin Xiao and Guozheng Zhang, *The queuing application in bank service optimization, Logistics Systems and Intelligent Management*, International Conference, IEEE press, 2(2010), 1097-1100.

[6] C.Kandemir Caues and L.Cauas, *An Application of Queuing Theory to the Relationship between Insulin Level and Number of Insulin Receptors*, Turkish Journal of Biochemistry, 32(1)(2007), 32-38.

[7] A.K.Kharwat, *Computer Simulation: An Important Tool in the Fast-Food Industry*, Proceedings of the 1991 Winter Simulation Conference, IEEE Press, (1991), 811-815.

[8] Prem Kumar Gupta and D.S.Hira, *Operations research*, Revised Edition, S. Chand, (2008), 903-910.

[9] J.K.Sharma, *Operations research theory and applications*, third edition, Macmillan India Ltd. New Delhi, (2007), 725-750.

[10] H.A.Taha, *Operations Research An Introduction*, 8th edition, McMillan Publishing Company, New York, (2007).

[11] S.Vijay Prasad, V.H.Badshah and A.K.Tariq, *Mathematical Analysis of Single Queue Multi Server and Multi Queue Multi Server Queuing Model: comparison study*, Global Journal of Mathematical Analysis, 3(3)(2015), 97-104.

[12] S.Vikas, *Use of Queuing Models in Healthcare: Department of Health Policy and Management*, University of Arkansas for Medical science at: http://works.bepress.com/cgi/viewcontent.cgi?article=1003&context=vikas_singh

[13] Yuejian Jie, *The optimal supermarket service*, International Journal of Business Management, 5(2)(2010), 128-129.

[14] X.X.Zaho, *Queuing theory with bank management innovation*, Modern Finance, 3(2007), 9-10.