Proceedings of IOE Graduate Conference, 2017
Volume: 5 ISSN: 2350-8914 (Online), 2350-8906 (Print)

Infinite Server Queueing Model with State Dependent Arrival and Service Rates

Sushil Ghimire [a], Gyan Bahadur Thapa [b], Ram Prasad Ghimire [c]

[a, b] *Department of Applied Sciences, Pulchowk Campus, IOE, TU, Nepal*
[c] *Department of Mathematical Sciences, School of Science, Kathmandu University, Nepal*
Corresponding Email: [a] sushil198@gmail.com, [b] thapagbt@ioe.edu.np, [c] rpghimire@ku.edu.np

Abstract

There are many situations in which people do not need to wait for the service. Servers are available more than the customers so that no one has to wait for service. Arrival follows a Poisson process with arrival rate λ and service time is exponentially distributed with rate μ. System capacity is only for the N number of customers. If there are more than N arrivals, they are rejected. We derive the formulas for the average number of customers in the system using recursive method to solve the mathematical model. Numerical results are presented to verify the results.

Keywords

Queueing, Server, Customer, Performance, Finite Capacity

1. Introduction

Most of us have experienced queue in our daily life in the bank, restaurant, supermarket, bus station and many other fields. All the arrivals wait for the service if the server/s are busy to serve the another customer. There are different types of Queueing models in which customers are served following different Queueing discipline. In some of the Queueing models, only one server is available whereas in some cases, number of servers are ready to serve the arrivals. It is quite natural that no one wants to spend time waiting in a queue though it is not always possible. But still, in Queueing theory, there is a discipline in which every arrival experiences immediate service. This is the system where arrivals are governed by a Poisson process and there are infinitely many servers, so jobs do not need to wait for a service. This is the special case of Queueing model where the number of servers which in general is denoted by c becomes very large with exponentially distributed service time. There are number of works performed by the different researchers in this field and some of them are as follows. Roijers et al. [1] analysed congestion periods of a $M/M/\infty$ queue presenting recursive relations through which all moments and covariance were obtained along with the formula for busy period. Kim and Lim [2] investigated customers' overtaking behaviour in a $M/M/c$ Queueing model applicable in flexible assembly system and telecommunication system for the study of explicit forms of probability distribution. Baek et al. [3] evaluated formulas for queue length, waiting time and probability of a busy period for a transient $M/D/1$ queue starting with a positive number of initial customers using the limiting property of an Erlang distribution. Ghimire et al. [4] derived and verified formulas for mean queue length and mean waiting time considering batch arrival of customers using generating function technique. Ammar [5] studied impatience customers and multiple vacations in a single server Queueing system in which impatience of a customer is because of the absentee of server upon arrival. Expressions for the time dependent probabilities, mean and variance of the system size some numerical illustrations were provided. Ramasamy et al. [6] contributed for violated First Come First Served $M/G/2$ Queueing system where both the servers are heterogeneous having mean service rates 1 for the first server and general service time distribution $B(t)$ for the second server to calculate the number of customers in the system and the actual waiting time by the customers in the system. Brandwajn and Begin [7] described reduced complexity in $M/Ph/c/N$ queues for which state of one server is represented explicitly, while the other servers were accounted through their rate of completions. Haviv and Oz [8] reviewed some existing observable Queueing mechanisms in which some were involved money transfers and suggested that the best ones are those for which customers have to make up their mind whether to join or not without inspecting the queue length. Barache et al. [9] dealt with the stationary characteristics of the $GI/M/\infty$ Queueing system by using an $M/M/\infty$ Queueing system to obtain the stability inequalities of the stationary distribution of the queue size to evaluate the performance of the proposed method by developing an algorithm which allowed to compute the various theoretical results. Sah and Ghimire [10] dealt with transient Erlangian Queueing system for the calculation of different performance measures and verified by using MATLAB simulation. Kim and Kim [11] proposed a model for multi-server batch arrival $M^X/M/c$ queue with impatient customers for the exact expression of loss probability in terms of the waiting time distribution with no impatience. Whitt [12] proposed a limit theorem supporting many busy servers where arrival occurs from a Poisson source to enhance the idea for heavy traffic limit and time varying arrival rate. Jiang et al. [13] dealt with a disaster $M/G/1$ queue in a multi-phase random environment in which the system stops working suddenly and resumes after exponential repair time. Along with some numerical results they calculated stationary queue at an arbitrary epoch, the sojourn time distribution and the length of the server's working time. Corral and Garcia [14] presented an $M/M/c$ retrial queue for the calculations of maximum queue length during a fixed time interval using splitting methods and eigenvalue, eigenvector technique. Whitt [15] examined steady state infinite server Queueing distribution where exponential service and sinusoidal arrival rate function is assumed. Schweer and Wichelhaus [16] studied non-parametric estimation of the service time distribution under partial information of discrete-time $GI/G/\infty$ queue using sequence of differences for the estimation of the resultant covariance kernel. Ghimire et al. presented [17] finite capacity time dependent multi-server Queueing model to calculate various performance measures verifying the results graphically using simulation. D'Auria [18] considered $M/G/\infty$ queue for the calculation of a stochastic decomposition formula for the number of customers in the system together with some examples in random environment. Gullu [19] analysed batch arrivals $M/G/\infty$ queue where whole arrival in a batch is served by the same server and the number of jobs in the system is characterized as a compound Poisson random variable.

The paper is classified in different Sections. Section 2 includes mathematical notations used in the model. Section 3 describes the derivation of mathematical model along with the formula for mean number of customers in the system. Numerical results are presented in Section 4 and finally Section 5 concludes the paper.

2. Mathematical Notations

The important part in the theory of queue is to present the mathematical model. Mostly, the same mathematical notations are used in the proposed Queueing model besides some other notations in some special models. In this paper, we use some standard mathematical notations which are common in Queueing theory and some of them are as follows:

(i) λ = Arrival rate

(ii) μ = Service rate

(iii) N = System Capacity

(iv) P_i = probability for state i

(v) L_s = Mean number of customers in the system

(vi) W_s = Mean waiting time in the system

Based on these notations, we propose the mathematical model in the next section. We use the recursive method for the derivation of the balanced equations established from the transient diagram for the different states.

3. Mathematical Model

For the notations described above, balanced equations are derived and solved by using recursive technique. Fig 1 is the transient diagram for our model and based on rate in is equal to rate out property, we have the following balanced equations for the different states:

for state zero $i.e.$ if there is no customers in the system

$$N\lambda P_0 = \mu P_0 \tag{1}$$

when there is one customer in the system

$$(N-1)\lambda P_1 + \mu P_1 = N\lambda P_0 + 2\mu P_2 \tag{2}$$

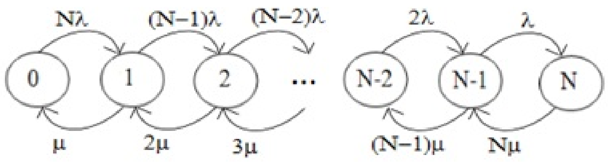

Figure 1: Transient diagram

Likewise, when there are two customers in the system

$$(N-2)\lambda P_2 + 2\mu P_2 = (N-1)\lambda P_1 + 2\mu P_3 \qquad (3)$$

Similar procedure is followed for all the states and when there are $N-1$ customers in the system, we have the following equation

$$\lambda P_{N-1} + (N-1)\mu P_{N-1} = 2\lambda P_{N-2} + N\mu P_N \qquad (4)$$

Finally, when there are N customers in the system

$$\lambda P_{N-1} = N\mu P_N \qquad (5)$$

All the above balanced equations can be separately represented in the form of P_0. From equation (1) we get,

$P_1 = \frac{N\lambda}{\mu} P_0$

Equation (2) can be written as

$P_2 = \frac{(N-1)\lambda}{2\mu} P_1 = \frac{N\lambda}{\mu} \frac{(N-1)\lambda}{2\mu} P_0$

Proceeding this way, equation (5) can be simplified as

$P_N = \frac{N\lambda}{\mu} \frac{(N-1)\lambda}{2\mu} \frac{(N-2)\lambda}{3\mu} \cdots \frac{\lambda}{N\mu} P_0$

The above equation for P_N can be written in the product forms as below:

$P_i = \prod_{k=0}^{i-1} \frac{\lambda(N-k)}{(k+1)\mu} P_0$

After suitable simplification, we get

$$P_i = P_0 \left(\frac{\lambda}{\mu}\right)^i \binom{N}{i} \qquad (6)$$

Now, the probability normalizing conditions is

$\sum_{i=0}^{N} P_i = 1$

Substituting the value of P_i from equation (6)

$\sum_{i=0}^{N} P_0 \left(\frac{\lambda}{\mu}\right)^i \binom{N}{i} = 1$

therefore, $P_0 = \left[\sum_{i=0}^{N} \left(\frac{\lambda}{\mu}\right)^i \binom{N}{i}\right]^{-1}$

The term inside the bracket is in the binomial form, hence

$p_0 = \frac{1}{\left(1+\frac{\lambda}{\mu}\right)^N}$

With this value of P_0 expression for P_i is

$$P_i = \begin{cases} \frac{\left(\frac{\lambda}{\mu}\right)^i \binom{N}{i}}{\left(1+\frac{\lambda}{\mu}\right)^N} & 0 \le i \le N \\ 0 & otherwise \end{cases}$$

Let L_s denote the average number of customers in the system, then it is calculated by using the formula for expectation of probability distribution, henceLet L_s denote the average number of customers in the system, then it is calculated by using the formula for expectation of probability distribution, hence

$L_s = \sum_{i=0}^{N} i.P_i = \frac{\left(\frac{\lambda}{\mu}\right)^i \binom{N}{i}}{\left(1+\frac{\lambda}{\mu}\right)^N}$

$= \frac{\frac{\lambda}{\mu}}{\left(1+\frac{\lambda}{\mu}\right)^N} \sum_{i=0}^{N} \frac{d}{d\left(\frac{\lambda}{\mu}\right)} \left(\frac{\lambda}{\mu}\right)^i \binom{N}{i}$

$= \frac{\frac{\lambda}{\mu}}{\left(1+\frac{\lambda}{\mu}\right)^N} \frac{d}{d\left(\frac{\lambda}{\mu}\right)} \left[\sum_{i=0}^{N} \left(\frac{\lambda}{\mu}\right)^i \binom{N}{i}\right]$

$= \frac{\frac{\lambda}{\mu}}{\left(1+\frac{\lambda}{\mu}\right)^N} \frac{d}{d\left(\frac{\lambda}{\mu}\right)} \left[\left(1+\frac{\lambda}{\mu}\right)^N\right]$

$= \frac{\frac{\lambda}{\mu}}{\left(1+\frac{\lambda}{\mu}\right)^N} \left[N\left(1+\frac{\lambda}{\mu}\right)^{N-1}\right]$

Since there are servers more than the customers, no one has to wait for the service providing that mean number of customers in the queue and mean waiting time in the queue are both zero. Mean waiting time in the system is merely the same as the average service time which is equal to $W_s = \frac{1}{\mu}$.

4. Numerical Results

To verify the model, we have used MATLAB simulation. We have plotted two graphs and for both of these graphs 20 is the system capacity (i. e. $N = 20$). Fig 2 is plotted for mean number of customers against arrival rate. It is the realistic phenomenon that whenever arrival increases number of customers in the system also increases. But after increasing the service rate, number of customers should decrease. This real nature is observed in Fig 2 in which for more arrival rate there are more customers and for more service rate there are less customers in the system. The top most graph is for the smallest service rate 12 and the graph close to the x-axis is for the highest service rate 14 concluding faster service decreases the number of requests in the system

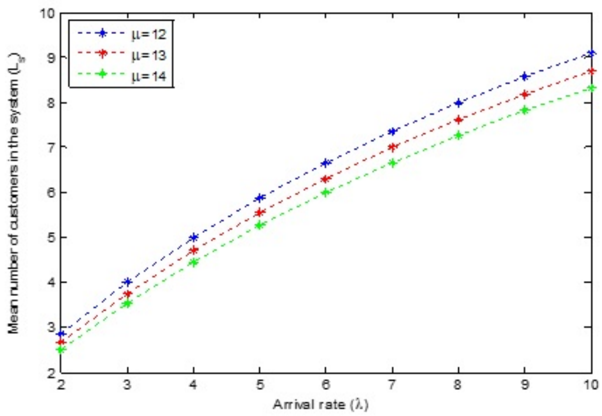

Fig 2: Mean number of customers in the system vs. Arrival rate

On the other hand, Fig 3 is the graph for mean number of customers vs. service rate. We can see that, for more service rate there are less number of customers in the system. For less arrival rate, less number of customers and for more arrival rate more number of customers in the system has been observed indicating that the model we established is appropriate. The graph at the bottom is for the least arrival rate 6 which indicates the less number of customers in the system whereas the number of customers are gradually increasing in the other two graphs for the arrival rates 7 and 8 respectively.

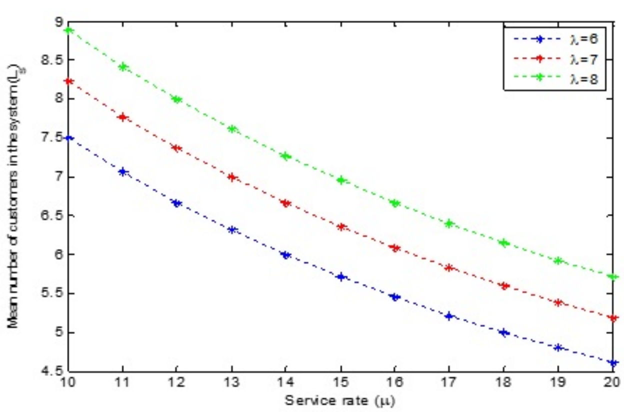

Fig 3: Mean number of customers in the system vs. Service rate

These two graphs obtained by using the MATLAB software are important to see the real applicability of the model in everyday life. Different values of arrival rate and service rate has been taken just to see the increased or decreased pattern of request in the system.

5. Conclusion

This is the special type of Queueing model which is rarely seen in practice but also it is an interesting part in the study of Queueing theory because it is the model for which no customers wait for the service. All the arriving customers are subject to receive the service without any waiting time. We have calculated the average number of customers in the system and plotted the graph for it against arrival rate and service rate. Since it is the finite capacity Queueing model the arriving customers exceeding N are not taken for the service. If we consider unlimited system capacity the study becomes more interesting and challenging. Moreover, including customers' behaviour like balking, reneging or jockeying makes the model more realistic.

We can see some of the applications of this type of model in the high class service station. Some of the telephone companies manage number of connecting towers so that no calls fail. In some rescue operations, one server is assigned to rescue one individual and in some occasions menu in the restaurant is ready before the customers arrive in the table. Likewise, some internet providers manage a very high bandwidth so that no users complain for the speed.

Acknowledgments

The first author is thankful to Erasmus Mundus SmartLink project for financial support as a PhD exchange student to carry out the work in Burgas Free University, Bulgaria from Sep 2016 – Sep 2017.

The second author is thankful to the Erasmus Mundus LEADERS Project for funding him as a Post Doc Research Fellow to carry out the work in Department of Mathematics, University of Evora, Portugal from Nov 2016 – Aug 2017.

References

[1] Frank Roijers, Michel Mandjes, and Hans Van Den Berg. Analysis of congestion periods of an $M/M/\infty$ queue. *Performance Evaluation*, 64:737–754, 2007.

[2] Woo-sung Kim and Dae-eun Lim. Computers & Industrial Engineering Analysis of overtaking in $M/M/c$ queues. *Computers & Industrial Engineering*, 101:177–183, 2016.

[3] Jung Woo, Ho Woo, Soohan Ahn, and Yun Han. Exact time-dependent solutions for the $M/D/1$ queue. *Operations Research Letters*, 44(5):692–695, 2016.

[4] Sushil Ghimire, R P Ghimire, and Gyan Bahadur Thapa. Mathematical Models of $M^b/M/1$ Bulk Arrival Queueing System. *Journal of the Institute of Engineering*, 10(1):184–191, 2014.

[5] Sherif I Ammar. Transient analysis of an $M/M/1$ queue with impatient behavior and multiple vacations. *Applied Mathematics and Computation*, 260:97–105, 2015.

[6] Sivasamy Ramasamy, Onkabetse A Daman, and Sulaiman Sani. An $M/G/2$ queue where customers are served subject to a minimum violation of FCFS queue discipline. *European Journal of Operational Research*, 240(1):140–146, 2015.

[7] Alexandre Brandwajn and Thomas Begin. Reduced complexity in $M/Ph/c/N$ queues. *Performance Evaluation*, 78:42–54, 2014.

[8] Moshe Haviv and Binyamin Oz. Regulating an observable $M/M/1$ queue. *Operations Research Letters*, 44(2):196–198, 2016.

[9] M Gi, M Gi, and Aicha Bareche. Approximate analysis of an $GI/M/\infty$ queue using the strong stability method. *IFAC-PapersOnLine*, 49(12):863–868, 2016.

[10] Shyam Sundar Sah and Ram Prasad Ghimire. Transient Analysis of Queueing Model. *Journal of the Institute of Engineering*, 11(1):165–171, 2015.

[11] Bara Kim and Jeongsim Kim. A batch arrival $M^X/M/c$ queue with impatient customers. *Operations Research Letters*, 42(2):180–185, 2014.

[12] Ward Whitt. A Poisson limit for the departure process from a queue with many busy servers. *Operations Research Letters*, 44(5):602–606, 2016.

[13] Tao Jiang, Liwei Liu, and Jianjun Li. Analysis of the $M/G/1$ queue in multi-phase random environment with disasters. *Journal of Mathematical Analysis and Applications*, 430(2):857–873, 2015.

[14] A Gómez-corral and M López García. Maximum queue lengths during a fixed time interval in the $M/M/c$ retrial queue. *Applied Mathematics and Computation*, 235:124–136, 2014.

[15] Ward Whitt. The steady-state distribution of the $M(t)/M/\infty$ queue with a sinusoidal arrival rate function. *Operations Research Letters*, 42(5):311–318, 2014.

[16] Sebastian Schweer and Cornelia Wichelhaus. Nonparametric estimation of the service time distribution in the discrete-time $GI/G/\infty$ queue with partial information. *Stochastic Processes and their Applications*, 125(1):233–253, 2015.

[17] Sushil Ghimire, Ram Prasad Ghimire, and Gyan Bahadur Thapa. Performance Evaluation of Unreliable $M(t)/M(t)/n/n$ Queueing System. *British Journal of Applied Science & Technology*, 7(4):412–422, 2015.

[18] Bernardo D Auria. Stochastic decomposition of the $M/G/\infty$ queue in a random environment. *Operations Research Letters*, 35:805–812, 2007.

[19] Reÿk Gullu. Analysis of an $M/G/\infty$ queue with batch arrivals and batch-dedicated servers. *Operations Research Letters*, 32:431–438, 2004.