

## **Waiting Time Reduction in a Local Health Care Centre Using Queueing Theory**

Mala<sup>1</sup>, S.P. Varma<sup>2</sup>

<sup>1</sup> (CSIR-NET), University Dept. of Mathematics, BRA Bihar University, Muzaffarpur-1, Bihar, India.

<sup>2</sup> Ex-Associate Prof., University Dept. of Mathematics, BRA Bihar University, Muzaffarpur-1, Bihar, India.

---

**Abstract:** *Queueing Theory is a mathematical approach to the study of waiting lines. Long waiting time in a health care system indicates the lack in management of the system. As a result of this, a health care centre lost the patients satisfaction and also it decreases the enhancement of the system. This paper is an attempt to analyse the use of queueing theory in a local health care clinic and the calculations performed in this paper is based upon the actual observed data collected from a local health care clinic located in Muzaffarpur city of Bihar, India. The paper summarizes a range of queueing theory results in the following areas. Traffic intensity, Average waiting time in queue, Average of time spent in system, Average queue length, Average number of individuals in the system. The central objective of this paper is to reduce the waiting time of patients and also to increase the efficiency of the clinic. The paper also considers the clinic as a single server queueing system following Poisson arrival based upon the discipline first come first serve and exponential service rate.*

**Keywords:** *exponential service, Poisson arrival, queueing theory, waiting line, waiting time.*

---

### **I. Introduction**

Queueing Theory has its origin in research by a Danish telephone engineer Agner Krarup Erlang when he generated models to describe the Telephone Exchange of Copenhagen. "The Theory of Probability and Telephone Conversations" is the first paper on queueing theory published in 1909 by Erlang [1]. He pondered the problems of determining the number of telephone circuits necessary to provide phone services that would prevent customers from waiting too long for an available circuit. In developing a solution to this problem he began to realize that the problem of minimizing waiting time was applicable to many fields and began developing the theory further. Therefore, A.K. Erlang is considered the father of queueing theory.

The queueing theory is one of the most celebrated problems of Operations Research which has attracted the attention of researchers, scientists, mathematicians and social scientists. A lot of research work has been dedicated to the application of this theory in health care systems, construction industries, human resource management, transportation, traffic and many other such systems.

A lot of contribution and application of queueing theory in the field of health care are found in the literature. In an era of health care reform, queueing theory is applied in improving quality, safety and decreasing health care cost. Whenever it is used appropriately, the results are often remarkable in saving time, increasing revenue and increasing staff and patient satisfaction. Therefore, applying queueing theory to health care sector is a necessary step towards improving quality of care and enhancement of the systems. Queueing theory has been studied in health care settings since 1952 [2]. A considerable body of research has shown the use of queueing theory in real world health care situation. McClain [3] reviews research on models for evaluating the impact of bed assignments policies on utilization, waiting time and the probability of turning away patients. Nosek and Wilson [4] review the use of queueing theory in pharmacy application with particular attention to improving customer's satisfaction. Customer satisfaction is improved by predicting and reducing waiting times and adjusting staffing. Preater [5] presents a brief history of the use of queueing theory in health care. Green [6] applied the queueing theory in health care. She discusses the relationship amongst delays, utilization and the number of servers; the basic M/M/S model, its assumptions and extensions; and the application of the theory to determine the required number of servers. Agnihotri, S.R and Taylor, P.F. [7], Khan, M.R. and Callahan, B.B. [8] and many others have exploited the queueing theory in the different areas of health care viz. waiting time and utilization analysis, system design, appointment systems, out-patient appointment systems, the emergency cardiac in- patient flow and others. They successfully established the applicability of queueing theory in the field of health care.

Thus, queueing theory is a mathematical approach to the study of waiting lines. Long waiting times in any health care centre affect the improvement of the centre as well as the nation's economy. Therefore, to reduce the waiting times of arriving patients is a major challenge for services not only in India but all over the world especially in developing countries. While considering improvement in services, centre must measures the

cost of providing a given level of service against the potential costs from having patients wait. Queueing theory has increasingly become a universal tool of management for decision making in a local health care clinic.

This chapter is focused on a case study of a local health care clinic “Venkatesh Nursing Home” located at Maripur of Muzaffarpur city of Bihar, India where Dr. Awadhesh Kumar (M.S.) is providing treatment of surgery and consultancy as a single server. This chapter analyzes the available resources of the clinic as well as it focuses on the utilization of queueing models for the enhancement of the clinic. The case study is on the basis of actual observed data of 22 days, collected from the clinic and the performance parameters have been used for calculations.

## II. Description of the Models

### Single Channel Queueing System

Consider a single server queueing system, (M/M/1) in which arriving customers is following Poisson’s process with arrival rate  $\lambda$  and the service process is following exponential distribution with service rate  $\mu$ . Here customers are identified as arriving patients. We assume that the services in all phases are independent and identical and only one patient at a time is in the service mechanism.

When a patient enters the system and at a time if the system is free, his/her service time starts at once and when the system is not free, the patient joins the queue and wait for their turn/number for service. After completion of services, the patient is free from queue if there is not any further extended service facility. If the server is busy then the arriving patients goes to orbit and becomes a source of repeated calls. This pool of source of repeated calls may be viewed as a sort of queue. The time it takes to serve every patient is an exponential random variable with parameter  $\mu$ .

A pictorial representation of a single server queueing system, (m/m/1) is given below in which patients are standing in queue, waiting for the server to be free for providing service. In the situation of congestion of patients in a health care system, there is a very less possibility for the patients arriving in end to get treatment as there is a single server rendering services. In worst situation, patients may leave the system without being served.

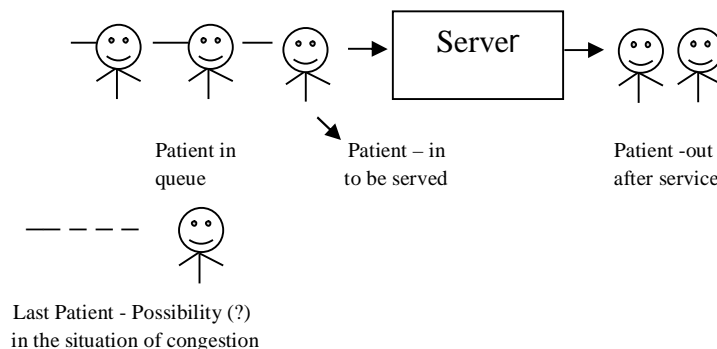


Fig1: single channel queueing system

### Multichannel Queueing System

The multichannel queueing model is known in the Kendall’s notation as the M/M/m model, where M signifies a Poisson distribution and m is number of parallel service channels in the system. This is commonly used to analyze the queueing problem. This model computes the average wait times and queue lengths, given arrival rate, number of servers and service rates. This particular model applies, in which there are multiple channels served by a single queue as at a bank teller or many airline tickets counters. The outputs of the model are as follows:

1. Expected waiting time per patient in the system (health centre).
2. Expected waiting time of patients in the queue.
3. Expected number of patients in the system (health centre).
4. Expected number of patients in the queue.

The exact calculation of these measures requires knowledge of the probability distribution of the arrival rate and service times. Moreover, successive inter- arrival times and services times are assumed to be statistically independent of each other. In this system, there are multiple servers with all sharing a common waiting line. A waiting line is created when all the servers are busy in rendering service. As soon as one server

becomes free, a customer (patient) is dispatched from the waiting line using the dispatching discipline in force for being served. These are obvious from the pictorial representation of the multichannel queuing system which is given below.

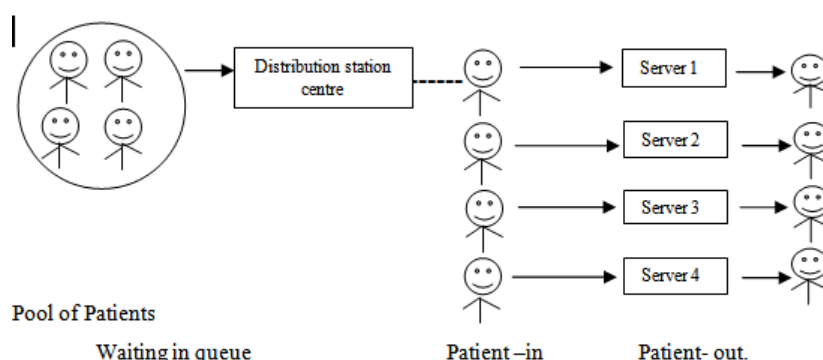


Fig. 2: multi-channel queuing system

### III. Model Assumptions

This research is based on the following assumptions.

- a) The finding obtained after investigation from one unit of the medical health centre should be valid in the other units.
- b) The patients are almost well familiar with the organization system of the medical centre.
- c) The arrival rate of the patients to queue and service rate are compatible to Poisson distribution or in other words the time interval between two consecutive arrivals and time service both follow exponential distribution.
- d) The queueing discipline is such that the first patient goes to the server which is ready for the service.
- e) In case of multichannel queueing system, it is assumed that none of the servers are unattended.

### IV. Details of the Case Study of Local Health Care Clinic

This chapter is a case study of Dr. Awadhesh Kumar’s clinic which is a local health care centre “Venkatesh Nursing Home” Ram Raji Road, Maripur located in Muzaffarpur city of Bihar (India). The study adopts a descriptive, actual observational case study approach. The goal of the study is to reduce the waiting time of patients in the clinic and also to optimize its existing capacity.

At the registration counter three persons, named Md. Imam, Mr. Jai Prakash and Mr. Suman are employed for providing services and they perform works according to their schedule in a week. In this local health care clinic, working doctor Dr. Awadhesh Kumar (M.S) is treated as a single server. He performs general surgery, laparoscopic operations and consultancy. He attends the patient mostly in morning and afternoon session, timing 8:00 to 9:00 AM and 1:30 to 3:00 PM respectively. Congestion of patients is low on Saturday and Sunday in comparison to other days of a week. An average of 8 to 10 minutes is taken for a single patient in the first check up and an average of 30 to 40 minutes for a surgery. After examining, some of the patients leave the centre with the advice to follow the prescribed medicine and some of them were told to be present at a specified time on a particular day for different types of the surgery.

The case study is based upon the actual observed data collected in 22 days. The arrival time as well as the service time began and ended for around 125 patients in the local health care clinic. The waiting time is obtained by subtracting the arrival time from the time of service began for each of the day.

### V. Model Parameters

#### Traffic Intensity

It is obtained from dividing the average arrival rate  $\lambda$  (in time) to the average service rate  $\mu$ .

$$\text{i.e. } \rho = \lambda/\mu \quad \text{----- (i)}$$

Whenever  $\lambda$  is larger, the arrival of patients will increase and the system will work harder and queue will be longer. On the contrary, whenever the  $\lambda$  is smaller, the queue will be shorter but in this case, the use of the system will be low. If the arrival rate of patients in the system were more than service rate i.e.  $\lambda > \mu$  then  $\rho > 1$ , which means the system capacity is less than the arriving patients, therefore the queue length is increased. In this queueing system the average arrival rate is less than the average service rate i.e.  $\lambda < \mu$ .

**Average Waiting Time in Queue**

The average waiting time in queue (before services is rendered) is equal to the average time which a patient waits in the queue for getting service. Its formula is

$$\frac{\rho}{\mu(1-\rho)} = \frac{\lambda}{\mu(\mu-\lambda)} \quad \text{----- (ii)}$$

**Average Time Spent in the System**

The average time spent in a system (on queue and receiving service) is equal to the total time that a patient spends in a system which includes the waiting time and service time. Its formula is

$$\frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda} \quad \text{----- (iii)}$$

**Average Number of Patients in the System**

The average number of patients in the system is equal to the average number of patients who are in the line or server. It is defined as

$$\frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda} \quad \text{----- (iv)}$$

**Average Queue Length**

The average queue length is composed of the average number of patients who are waiting in the queue. It is defined as

$$\frac{\lambda^2}{\mu(\mu-\lambda)} \quad \text{----- (v)}$$

**The Probability of not Queuing on the Arrival**

$$= 1 - \rho \quad \text{----- (vi)}$$

**VI. Calculations Performed for the Case Study**

The arrival time as well as the time service began and ended for 125 patients in the local health care clinic “Venkatesh Nursing Home”. There are two types of services; consultancy and surgery. A total of 22 days were used for the data collection. On the basis of actual observed collected data, we find that

Total waiting time of 125 patients for 22 days = 1584 minutes

Total service time of 125 patients for 22 days = 1320 minutes

Using the model parameters for the single channel queueing model, we arrive at the following results:

The arrival rate,  $\lambda = \frac{\text{Total number of patients}}{\text{Total waiting time}}$   
 $= \frac{125}{1584} = 0.0789 \quad \text{----- (i)}$

The service rate,  $\mu = \frac{\text{Total number of patients}}{\text{Total service time}}$   
 $= \frac{125}{1320} = 0.09469 \quad \text{----- (ii)}$

The average time spent in queue (before service is rendered)  
 $= \frac{\lambda}{\mu(\mu-\lambda)} = \frac{0.0789}{0.09469(0.09469-0.0789)}$   
 $= 52.9530 \equiv 52 \text{ minutes} \quad \text{----- (iii)}$

The average time spent in the system  
 $= \frac{1}{\mu-\lambda} = \frac{1}{0.09469 - 0.0789}$   
 $= 63.3312 \equiv 63 \text{ minutes} \quad \text{----- (iv)}$

Average number of patients in the system

$$= \frac{\lambda}{\mu - \lambda} = \frac{0.0789}{0.09469 - 0.0789}$$

$$= 4.9968 \equiv 5 \quad \text{----- (v)}$$

Average queue length (i.e. Average number of patients in queue)

$$= \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho^2}{(1 - \rho)}$$

$$= \frac{0.6942}{0.167} = 4.1570 \equiv 5 \text{----- (vi)}$$

The probability of queueing on arrival i.e. Traffic

$$= \rho = \frac{\lambda}{\mu}$$

$$= \frac{0.0789}{0.09469} = 0.8332 \text{----- (vii)}$$

The probability of not queueing on the arrival

$$= 1 - \rho$$

$$= 1 - 0.8332 = 0.167 \text{----- (viii)}$$

### VII. Results And Discussion

A single channel queueing system is used to represent the local health care clinic where doctor is treated as a single server and the model parameters are applied for calculations. The case study is on the basis of the actual observed data collected in 22 days of service for 125 patients.

The traffic intensity,  $\rho = \frac{\lambda}{\mu} = 0.8332$  obtained in (vii) shows the probability of patients queueing on arrival. This reveals the congestion of patients waiting for treatment as doctor is engaged in rendering service to the patients that has earlier been given appointment either for surgery or consultancy. This represents the inadequate service system of the clinic.

Also from the results (iii) and (iv) it is obvious that the average time spent in the clinic (in queue and in receiving treatment) is greater than the average time spent in the queue before providing treatment. Thus, there will always be a queue of patients in the clinic which is also very clear from the results (v) and (vi). The result (viii) shows that there is a very less possibility of providing services to new arriving patients.

### VIII. Conclusion

In course of the case study it was observed that in the clinic, patients have to spare tremendous waiting time for service, and a result of which emergency patients may suffer.

It is observed that the congestion occurs due to patients waiting for the doctor who has been engaged in surgical operation of the patient earlier given appointment. But in case of some emergency, the earlier given appointments are postponed to another schedule and this event extend the waiting times of other patients including both those waiting for consultation and surgery. This congestion may be reduced either if one more server that is a doctor be appointed for first checkup of the patient or the single server that is the doctor may be able to fix the number of patients for a day. The second case will not be beneficial for any health care centre improvement, it may be termed as denial of providing service for an ailing individual.

### References

- [1] A.K. Erlang, The Theory of Probabilities and Telephone Conversations, *NytTidsskrift for Matematik*, 20(B), 1909, 33-39.
- [2] N.T.J. Bailey, A study of queues and appointment systems in hospital out-patient departments with special reference to waiting times, *Journal of Royal Statistical Society, Netherlands*, 14, 1952, 185-199.
- [3] J.O. McClain, Bed planning using queueing theory models of hospital occupancy: a sensitive analysis, *Inquiry*, 13, 1976, 167-176.
- [4] R.A. Jr. Nosek and J.P. Wilson, Queueing theory and customer satisfaction: a review of terminology, trends and applications to a pharmacy practice, *Hospital Pharmacy*, 3, 2001, 275-279.
- [5] J. Preater, Queue in health, *Health Care Management Science*, 5, 2002, 283.
- [6] L.V. Green, Queueing analysis in healthcare, in *patient flow: Reducing delay in healthcare delivery.*( Hall, R.W., Springer, New York, 2006a), 281-308.
- [7] S.R. Agnihotri and P.F. Taylor, Staffing a centralized appointment scheduling department in Lourdes Hospital, *Interfaces-Journal Online*, 21, 1991, 1-11.

- [8] M.R. Khan, and B.B. Callahan, Planning laboratory staffing with a queueing model, *European Journal of Operational Research*, 67(3), 1993, 321-331.
- [9] R.A. Adeleke, O.D. Ogunwale, O.Y. Halid, Application of Queueing Theory to Waiting Time of Out-Patients in Hospitals, *Pacific Journal of Science and Technology*, 10(2), 2009, 270-274.
- [10] S.Fomundam, J.W. Herrmann, A Survey of Queueing Theory Applications in Healthcare, *ISR Technical Report* 2007, 24.
- [11] A. Irfan, A. Bashir and A. Shakeel, Performance analysis of networks of queues under active queue management scheme, 15, 2007, 416-425.
- [12] Mala and S.P. Varma, Application of Queueing Theory in Health Services, *IJAIR*, Vol.3, Issue 2, 2014, 121-124.
- [13] Mehandiratta, R., Application of Queueing Theory In Health Care, *International Journal of Computing and Business Research* 2 (2), 2011, 2229-6166.
- [14] M. Tabari, M.Seifi-Divkolaii and Y. Gholipour-Kanani, Human Resource Management with using of Analyses Queue Approach, *Indian Journal of Education and Information Management*, 1, 2012, 79-87.
- [15] M. Tabari, Y. Gholipour and M. SeifiDivkolaii, Application of the Queueing theory to Human Resource Management, 2012, 1211-1218.