Review Paper ISSN NO. 2456-3129

International Journal of Engineering, Pure and Applied Sciences,
Vol. 3, No. 2, June-2018

Study about Queuing Theory and its Applications

Amita Devi
Department of Mathematics, IEC University, Baddi, H.P, INDIA
Email: amitaranote08@gmail.com

Abstract- In Queuing theory we analyze the basic features of queuing theory and its applications. There are many situations in daily life when a queue is formed. Queuing theory is the mathematical study of waiting lines and it is very useful for analyzing the procedure of queuing of daily life of human being. Queuing theory applies not only in day to day life but also in sequence of computer programming, networks, medical field, banking sectors etc.
Queuing models can be used to model characteristics of basic call centres, i.e. multi-server systems with time dependent arrival rates, general service time distribution and state-dependent on arrival(balking).The discrete time modeling approach which has previously been used for modeling the time-dependent behaviour of multi-server queues is extended to incorporate state-dependent balking. A simple formula to estimate the mean number in the system is derived for busy systems with balking.

Keywords- Queuing theory, Queuing system, Queuing Networks.

1. INTRODUCTION

Queuing theory is a branch of mathematics that studies and models the act of waiting in lines. This paper will take a brief look into the formulation of queuing theory along with examples of the models and applications of their use. The goal of the paper is to provide the reader with enough background in order to properly model a basic queuing system into one of the categories we will look at, when possible. Also, the reader should begin to understand the basic ideas of how to determine useful information such as average waiting times from a particular queuing system.
The first paper on queuing theory, “The Theory of Probabilities and Telephone Conversations” was published in 1909 by A.K. Erlang, now considered the father of the field. His work with the Copenhagen Telephone Company is what prompted his initial foray into the field. He pondered the problem of determining how many telephone circuits were necessary to provide phone service 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.
Erlang’s switchboard problem laid the path for modern queuing theory. The chapters on queuing theory and its applications in the book “Operations Research: Applications and Algorithms” by Wayne L. Winston illustrates many expansions of queuing theory and is the book from which the majority of the research of this paper has been done.
In the second section of this paper, we will begin defining the basic queuing switchboard problem laid the path for modern queuing. We will begin by reviewing the necessary probabilistic background needed to understand the theory. Thus we will move on to discussing notation, queuing disciplines, birth-death processes, steady-state probabilities, and Little’s queuing formula.
A queuing system which consists of the customers and the servers. Waiting line or queues are in the schools, hospitals, bookstores, libraries, banks, post office, petrol pumps, theatres etc., all have Queuing problems. Queues are very familiar in our daily life. Queuing theory is a branch of operations research because the results are used for making decisions about the resources needed to provide service. Many valuable applications of the queuing theory are traffic flow (vehicles, aircraft, people, communications), scheduling (patients in hospitals, jobs on machines, programs on computer), and facility design (banks, post offices, supermarkets). A.K.Erlang (1878-1929).Danish Engineer who is called the father of Queuing theory. He published his articles relating to the study of congestion in telephone traffic. A queuing theory is the Mathematics of waiting lines. A queuing system can be described by the flow of units for service, forming or joining the queue, if service is not available soon, and leaving the system after being served. Waiting in lines or queues seems to be a general phenomenon in our day to day life. Think about the many times you had to wait in line in the last month or year and the time and frustration that was associated with those waits. Whether we are in line at the grocery store checkout, the barbershop, the stoplight, bank waiting our turn is part of our everyday life. Queuing theory is the formal study of waiting in line and is an entire discipline within the field of operations management. The purpose of this article is to give the reader a general background into queuing theory and queuing systems, its associated terminology, and how queuing theory relates to customer or customer satisfaction. Also, past and

IJEPAS

International Journal of Engineering, Pure and Applied Sciences, Vol. 3, No. 2, June-2018

present applications of queuing technology and what staffs can do to manage customer or customer queues more effectively will be discussed. Finally, automated queuing technology will be described. Queuing theory utilizes mathematical models and performance measures to assess and hopefully improve the flow of customers through a queuing system. Queuing theory has many applications and has been used extensively by the service industries. Queuing theory has been used in the past to assess such things as staff schedules, working environment, productivity, performance, customers waiting time, and customers waiting environment. In bank, queuing theory can be applied to assess a multitude of factors such as registration fill-time, customer waiting time, customer counselling time, and receptionists and technician staffing levels. The application of queuing theory may be of particular benefit in receptionists with high-volume out customer workloads and/or those that provide multiple points of service, such as those in the Department of Veterans Affairs (VA), Department of Defence (DoD), university health systems, and managed care organizations. Problematic queuing systems (i.e, long lines) can lead to the customer's perceptions of excessive, unfair, or unexplained waiting time—resulting in significant detrimental effects on the customer's overall satisfaction with the service transaction. Waiting in lines seems to be part of our everyday life. At the bank, filling station, bus stop, or in the canteen, "waiting our turn". Queues form when the demand for a service exceeds its supply (Kandemir-Cavas and Cavas, 2007). In banks, customers can wait minutes or hours before been attended to. For many customer or customers, waiting in lines or queuing is annoying (Obamiro, 2003) or negative experience (Scotland, 1991). The unpleasant experience of waiting in line can often have a negative effect on the rest of a customer's experience with a particular bank. The way in which managers address the waiting line issue is critical to the long term success of their firms (Davis et al, 2003) Queuing has become a symbol of inefficiency of publicly funded bank in the world and Nigeria is not an exception. Managing the length of the line is one of the challenges facing most banks. A few of the factors that are responsible for long waiting lines or delays in providing service are: lack of passion and commitment to work on the part of the bank staff (Belson1988) overloading of available staff, bank officials attending to customers in more than one section etc. These put bank managers under stress and tension, hence tends to European Journal of Business and Management www.iiste.org ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online) Vol.7, No.35, 2015 91 dispose off a customer without attending to their needs, which often leads to customer dissatisfaction (Babes and Sarma, 1991). This paper is based on the understanding that most of these difficulties can be managed by using queuing model to determine the waiting line performance such as: average arrival rate of expectant, average service rate expectant, system utilization factor, cost of service and the probability of a specific number of customers in the system. The purpose of this study is to provide insight into the general background of queuing theory and its associated organization performance, and how queuing theory can be used to model good service delivery and organization performance of a First Bank Plc Gwagwalada Abuja, Nigeria. The resultant performance variables can be used by the policy makers to increase efficiency, improve the quality of service and performance, as well as decrease cost in bank organizations and services.

.

1.3 Applications

- **Library management**

A library is an organized collection of books, and some special materials like audio or visual materials, CDs, cassettes, video tape, DVDs, e-books, audio books and many other types of electronic resources. Scope of Queuing application in Libraries is circulation of books counters service and allied services like reprography. The basic tasks in library are stacks maintenance, membership management, selection of library materials, and planning the acquisition of materials

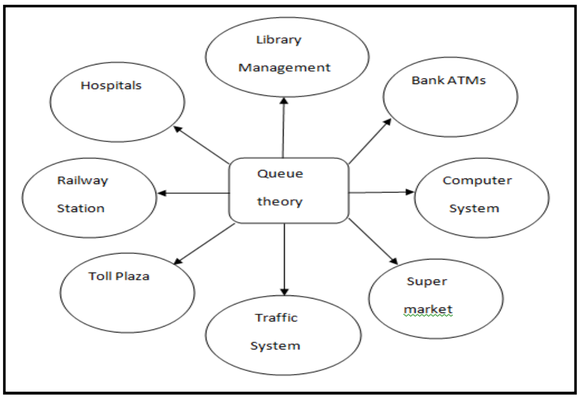

Fig. 1 structure for applications of queuing theory

- **Bank ATMs**

The Automated Teller Machine (ATM) is one of the several electronic banking channels used in the banking industry of Ghana. According to Aldajani and Alfares (2009), automated teller machines are among the most important service facilities in the banking industry. In Ghana, ATM as a banking instrument has enjoyed widespread acceptance and usage. In a survey conducted by Abor (2005), more than half of respondents revealed their preference for ATM as a conduit to conducting transactions. ATMs particularly when installed off-site serve to keep customers away from bank halls. Unfortunately most ATMs in Ghana are on-site operating very much as another department

IJEPAS
International Journal of Engineering, Pure and Applied Sciences,
Vol. 3, No. 2, June-2018

of the bank. The issue of queue control in bank halls via ATM is essentially defeated because ultimately access to these facilities is limited to customers going to the bank. ATMs themselves have as a result become subjects of large service demands which directly translate to queues for services when these demands cannot be quickly satisfied. This situation becomes compounded and more evident during festive periods and month endings, around which time demand for cash is high.

In ATM, bank customers arrive randomly and the service time i.e. the time customer takes to do transaction in ATM, is also random. We use queuing model to derive the arrival rate, service rate, utilization rate, waiting time in the queue and the average number of customers in the queue. Queuing can help bank ATM to increase its quality of service, by anticipating, if there are many customers in the queue.

- **Hospitals**

Queuing models using for estimating waiting time of a patient, utilization of service, models system design, and models for evaluating appointment systems . A queuing system helps minimizing the waiting time of patients and maximizing the utilization of the servers i.e. Doctors, nurses, hospital beds etc. Queuing is not new but recently hospitals has begun to use it effectively

- **Traffic system**

The vehicular traffic flow and explore could be minimized using queuing theory in order to reduce the delay on the roads. The role of transportation in human life cannot be overemphasized. A basic model of vehicular traffic based on queuing theory. It will determine the best times of the red, amber, and green lights to be either on or off in order to reduce traffic.

- **Supermarket**

Supermarket is a important application of queuing theory. Two important results need to be known from the data collected in the supermarket by the mathematical model: one is the ‘service rate’ provided to the customers during the checking out process, and the other is the gaps between the arrival times (inter arrival time) of each customer per hour. In order to get an overall perspective of the customer’s quality of service, the questionnaires which indicate the result in percentages, are also used to get the evaluation from the customers directly.

There are five counters in ICA Borlänge at one place, which means consisting of five servers with five queues in terms of Queuing Theory. A queue forms whenever current demand exceeds the existing capacity to serve when each counter is so busy that arriving customers cannot receive immediate service facility. So each server process is done as a queuing model in this situation.

- **Toll plaza**

Computer simulation is one of the popular approaches to the design of toll plazas. Toll plaza configurations such as toll collection methods, number of toll booths, and types of vehicles have been used here. Toll plaza performance measures such as average queue length, average waiting time, maximum queue length, and maximum waiting time at the tolls were compared between two different types of representations of projected traffic volumes. Toll plaza designing factors such as lane selection options, electronic toll collection(ETC)rates, and number of manual tolls were combined with traffic flow measure the specified toll performances .Finding appropriate values of input parameters for a traffic simulation model is always a challenge to simulation model builders as well as to traffic engineers. For generating traffic flow in a simulation model, deterministic traffic counts for a time period can be used as an input parameter into the model rather than considering a probabilistic distribution.

- **Railway station**

In the country like India where Railway is one of the most popular and cheapest means of transportation, it is always difficult to book confirmed tickets for the journey. The population that the country has it doesn’t match up with number of trains running various routes especially those connecting the metro cities. Indian Railway is trying to meet the ever increasing demand of over billion people. The queuing system is used to avoid the inconvenience of passengers and it is feasible and the results are effective and practical.

- **Computer system**

Many jobs arrive sequentially at a computer system in accordance with a Poisson arrival process, and the execution time of a job is a random variable. Jobs are executed in the order of arrival, if the computer is busy when a job arrives; the job is placed in a queue. In the terminology of queuing theory the computer is the “server” and the jobs are “customers”. The logical structure of the single server queuing model can be restored with a simple device. Let the server represent the combined resources of computer and operator.

2. QUEUING ANALYSIS

One of the major issues in the analysis of any traffic system is the analysis of delay. Delay is a more subtle concept. It may be defined as the difference between the actual travel time on a given segment and some ideal travel time of that segment. This raises the question as to what is the ideal travel time. In practice, the ideal travel time chosen will depend on the situation; in general, however, there are two particular travel times that seem best suited as benchmarks for comparison with the actual performance of the system. These are the travel time under free flow conditions and travel time at capacity.

Most recent research has found that for highway systems, there is comparatively little difference

IJEPAS
International Journal of Engineering, Pure and Applied Sciences,
Vol. 3, No. 2, June-2018

between these two speeds. That being the case, the analysis of delay normally focuses on delay that results when demand exceeds its capacity; such delay is known as queuing delay, and may be studied by means of queuing theory. This theory involves the analysis of what is known as a queuing system, which is composed of a server; a stream of customers, who demand service; and a queue, or line of customers waiting to be served.

2.1 Queuing System

Figure 1 shows a schematic diagram illustrating the concept of a queuing system. Various components are discussed below.

❖ **Input parameters**

- Mean arrival rate
- Mean service rate
- The number of serve
- Queue discipline

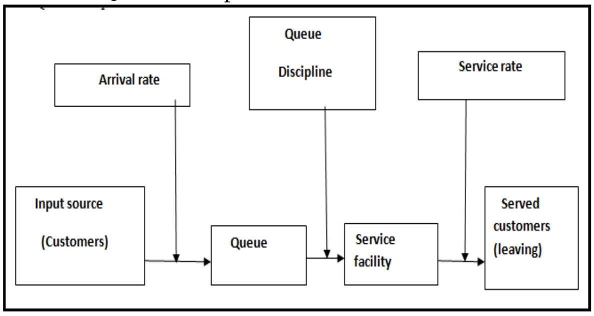

Fig.2 Components of a basic queuing system

These are explained in the following sections.

- **Mean Arrival rate (λ)**

It is rate at which customers arrive at a service facility. It is expressed in flow (customers/hr or vehicles/hour in transportation scenario) or time headway (seconds/customer or seconds/vehicle in transportation scenario). If inter arrival time that is time headway (h) is known, the arrival rate can be found out from the equation:

$$\lambda = 3600 / h$$

Mean arrival rate can be specified as a deterministic distribution or probabilistic distribution and sometimes demand or input are substituted for arrival.

- **Mean arrival rate (μ)**

It is the rate at which customers (vehicles in transportation scenario) depart from a transportation facility. It is expressed in flow (customers/hr or vehicles/hour in transportation scenario) or time headway (seconds/customer or seconds/vehicle in transportation scenario). If inter service time that is time headway (h) is known, the service rate can be found out from the equation:

$$\mu = 3600 / h$$

- **Number of servers**

The number of servers that are being utilized should be specified and in the manner they work that is they work as parallel servers or series servers has to be specified.

- **Queue discipline**

Queuing discipline is a parameter that explains how the customers arrive at a service facility. The various types of queue disciplines are

(1) First in first out (FIFO)
(2) First in last out (FILO)
(3) Served in random order (SIRO)
(4) Priority scheduling
(5) Processor (or Time) Sharing

(1) **First in first out (FIFO)**: If the customers are served in the order of their arrival, then this is known as the first-come, first-served (FCFS) service discipline. Prepaid taxi queue at airports where a taxi is engaged on a first-come, first-served basis is an example of this discipline.

(2) **First in last out (FILO)**: Sometimes, the customers are serviced in the reverse order of their entry so that the ones who join the last are served first. For example, assume that letters to be typed, or order forms to be processed accumulate in a pile, each new addition being put on the top of them. The typist or the clerk might process these letters or orders by taking each new task from the top of the pile. Thus, a just arriving task would be the next to be serviced provided that no fresh task arrives before it is picked up. Similarly, the people who join an elevator first are the last ones to leave it.

(3) **Served in random order (SIRO)**: Under this rule customers are selected for service at random, irrespective of their arrivals in the service system. In this every customer in the queue is equally likely to be selected. The time of arrival of the customers is, therefore, of no relevance in such a case.

(4) **Priority Service**: Under this rule customers are grouped in priority classes on the basis of some attributes such as service time or urgency or according to some identifiable characteristic, and FIFO rule is used within each class to provide service. Treatment of VIPs in preference to other patients in a hospital is an example of priority service.

International Journal of Engineering, Pure and Applied Sciences,
Vol. 3, No. 2, June-2018

(5) **Processor (or Time) Sharing**: The server is switched between all the queues for a predefined slice of time (quantum time) in a round-robin manner. Each queue head is served for that specific time. It doesn't matter if the service is complete for a customer or not. If not then it'll be served in its next turn. This is used to avoid the server time killed by customer for the external activities (e.g. preparing for payment or filling half-filled form).

2.2 Basic elements of queue

The basic elements of a queuing system can be classified by:

1) The Arrival process
2) Queue
3) The service mechanism
4) The Number of Customers allowed in the system
5) The Number of service channels

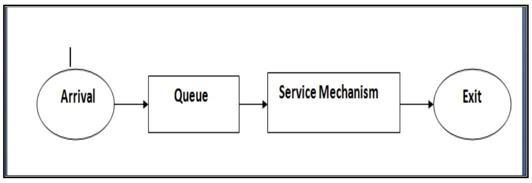

Fig. 3 Basic elements of queue

- **The Arrival process**

The simplest arrival process is one where we have completely regular arrivals (i.e. the same constant time interval between successive arrivals). A Poisson stream of arrivals corresponds to arrivals at random. In a Poisson stream successive customers arrive after intervals which independently are exponentially distributed.

The Poisson stream is important as it is a convenient mathematical model of many real life queuing systems and is described by a single parameter - the average arrival rate. Other important arrival processes are scheduled arrivals; batch arrivals; and time dependent arrival rates (i.e. the arrival rate varies according to the time of day).

- **Queue**

There can be limitations with respect to the number of customers in the system. For example, in a data communication network, only finitely many cells can be buffered in a switch. The determination of good buffer sizes is an important issue in the design of the networks.

- **The Service Mechanism**

A description of the resources needed for service

to begin how long the service will take (the **service time distribution**) The number of servers available whether the servers are in series (each server has a separate queue) or in parallel (one queue for all servers) Whether pre-emption is allowed (a server can stop processing a customer to deal with another "emergency" customer).

Assuming that the service times for customers are independent and do not depend upon the arrival process is common. Another common assumption about service times is that they are exponentially distributed.

- **The Number of Customers allowed in the System**

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 type of situations is 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) source 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.

The Number of Service Channels The more the number of service channels in the service facility, the greater the overall service rate of the facility. The combination of arrival rate and service rate is critical for determining the number of service channels. When there are a number of service channels available for service, then the arrangement of service depends upon the design of the system's service mechanism.

Parallel channels means, a number of channels providing identical service facilities so that several customers may be served simultaneously. Series channel means a customer go through successive ordered channels before service is completed. A queuing system is called a **one-server model,** i.e., when the system has only one server, and a **multi-server model** i.e., when the system has a number of parallel channels, each with one server.

2.3 Customers behaviour

The customers generally behave in five ways:

(1) **Patient Customer:** Customer arrives at the service system, stays in the queue until served, no matter how much he has to wait for service.

(2) **Customer:** Customer arrives at the service system, waits for a certain time in the queue and leaves the system without getting service due to some reasons like long queue before him.

IJEPAS

International Journal of Engineering, Pure and Applied Sciences,
Vol. 3, No. 2, June-2018

(3) **Balking:** Customer decides not to join the queue by seeing the number of customers already in service system
(4) **Reneging:** Customer after joining the queue waits for some time and leaves the service system due to delay in service.
(5) **Jockeying:** Customer moves from one queue to another thinking that he will get served faster by doing so.

2.4 State of System

The basic concepts in the analysis of queuing theory are that of a state of the system. It involves the study of a system's behaviour over time .The state of system may be classified as follows:

- **Transient state:**

A system is said to be in transient state when its operating characteristics are dependent on time. Thus a queuing system is in transient state when the probability distributions of arrivals waiting time and servicing time of customers are dependent on time. This state occurs at the beginning of the operation of the system.

- **Steady state:**

A system is said to be in steady state when its operating characteristics become independent of time. Thus a queuing system is in a steady state when the probability distributions of arrival, waiting time and servicing time of the customers are independent of the time. This state occurs in the long run of the system. Let P_n (t) denote the probability that there are n units in the system at time t, then the system acquires steady state as t→ ∞ if $\lim_{t\to\infty} P_n(t) = P_n$ (independent of time t).

- **Explosive state:**

If the arrival rate of the system is more than its servicing rate, the length of the queue will go on increasing with time and will tend to infinity as t→ ∞ .

3. Queuing Networks

- **Open Network system**

In an open system, customers enter and are being serviced at various stations and leave the system. Reimaan, M.I. [1984] applied open queue network in heavy traffic. Research on using queuing networks in modeling a number of complex systems such as manufacturing, computer and communication networks and the evaluation of performance measures of networks and the design of such systems has grown very rapidly in the last few years. Lee and Zipkin [1992] considered a tandem multistage production process where demand is a Poisson process and the unit production times are exponentially distributed. Markovian queuing networks have been studied by many authors. Brandon & Yechiali [1991] studied a Markovian tandem queuing network with feedback (with Poisson arrivals, Exponential service times, and Bernoulli routing). Bambos & Wasserman [1994] studied a stationery tandem queuing networks with job feedback. Frenk et al. [1994] presented two algorithms for server allocation problems in manufacturing queuing network.

For restricted networks, blocking may occur due to restrictions on each queue length, resulting in a decrease in maximum utilization. Queuing networks with blocking have proved useful in modeling stochastic systems such as computer systems, manufacturing systems, telecommunication system, and production lines, and quality control systems with a sequence of inspection stations.

During the last three decades researcher concentrate on the approximation techniques for the solutions of queuing problems and in particular to queuing networks. Brandwajn and Jow [1988] proposed an approximation method for the solution of tandem queues with blocking. In his paper, Altiok [1989] studied a system of tandem queues with phase-type service times and finite buffers. He considered three cases: (1) the first queue has an infinite buffer; (ii) the first queue has finite buffer; (iii) the first server is always busy. He used an approximation method to compute the steady-state probability distributions of the number of customers in each queue. This method decomposes the system into individual queues with revised arrival & service rates and with queue capacities. Then each queue is treated separately with an iterative scheme which relates these queues to each other. Lee & Pollock [1989] proposed an approximation method for analyzing the general configuration of a queuing network with blocking. In their paper, Song & Takahs [1991] studied problems arising in applications of the cross aggregation method to tandem queuing systems with production blocking.

- **Closed network system**

In a closed network a fixed and finite number of customers circulate through the network and no arrivals or departures are allowed.

A survey of queuing networks theory literature reveals that some consideration has been devoted to the study of closed systems. Most of the published paper concentrated on the study of the two-stage model, where a customer upon completion of his service in one of the two stations can enter service immediately on the other, or he has to wait if there is a queue in front of him.

Kumar Shanti & Yao [1989] studied the equilibrium behavior of the queue lengths in a product form closed queuing network when the service rates at a subset of stations are increased. Signman [1989] discussed the stability of closed queue network. Multiclass closed queuing networks was studied by

IJEPAS

International Journal of Engineering, Pure and Applied Sciences,
Vol. 3, No. 2, June-2018

Chevalier and Wein [1993] who considered the problem of finding an optimal dynamic sequencing polity to maximize the mean throughout rate. They generalized the results of Horrison and Wein [1990] from the system containing two-station network to a system containing a finite number of stations. Closed queuing networks with blocking have been studied by many authors. Suri & Diehl [1986] studied a single class closed queuing networks with exponential service times and blocking. They used an approximation technique based partly on analysis and partly on heuristic arguments. In the paper, Van Dijik and Lamond [1988] investigated a new methodology to obtain simple and computationally attractive bounds for non-product form queuing systems to be applied to finite single-server exponential tandem queues. Their methodology is based on modifying the original system into product form systems that provide bounds for some performance measures. They studied a two stage tandem queue with blocking.
Akyildiz [1988] presented an algorithm for throughput analysis of closed restricted systems. He states that if two queuing networks, one with blocking and the other without blocking, have an almost equal number of states, then their throughputs are almost equal. For an excellent review paper for closed systems with blocking, see also Onvural [1990].

4. QUEUING MODELS

Queuing theory uses models to represent the various types of queuing systems. There are various kinds of queuing models. These queuing models have a set of defined characteristics like some arrival and service distribution, queue discipline, etc. Formula for each model indicates how the related queuing system should perform, under a variety of conditions. The queuing model are very powerful tool for determining that how to manage a queuing system in the most effective manner. The queuing theory is also known as the random system theory, which studies the content of: the behaviour problems, the optimization problem and the statistical inference of queuing system.
The queuing models are represented by using a notation which is discussed in the following section of queue notation.

4.1 M/M/1 model

In this model the arrival times and service rates follow Markovian distribution or exponential distribution which are probabilistic distributions, so this is an example of stochastic process. In this model there is only one server. The important results of this model are:

1. Average number of customers in the system = L = $\frac{\rho}{1-\rho}$
2. Average number of customers in the system = Lq = $\frac{\rho^2}{1-\rho}$
3. Expected waiting time in the system W = = L/ λ = (1/λ) λ/ μ−λ = 1/ μ−λ
4. Expected waiting time in the queue Wq = Lq/ λ = 1/ λ × λ^2 / μ (μ−λ) = λ/ μ(μ−λ)

4.2 M/M/N model

The difference between the earlier model and this model is the number of servers. This is a multi -server model with N number of servers whereas the earlier one was single server model. The assumptions stated in M/M/1 model are also assumed here. Here μ is the average service rate for N identical service counters in parallel. For x=0

$$P(0)=[\sum_{x=0}^{N-1}(\frac{\rho^x}{x!}+\frac{\rho^N}{(N-1)!(N-\rho)})]^{-1} \quad \text{.......... (1)}$$

The probability of x number of customers in the system is given by P(x). For 1 ≤ x ≤ N

$$P(x)=\frac{\rho^x}{x!}*P(0) \quad \text{........... (2)}$$

For x > N

$$P(x)=\frac{\rho^x}{N!N^{x-N}}*P(0) \quad \text{............ (3)}$$

The average number of customers in the system is

$$E[X]=\rho+[\frac{\rho^{N+1}}{(N-1)!(N-\rho)^2}]P(0) \quad \text{........... (4)}$$

The average queue length

$$E[Lq]=[\frac{\rho^{N-1}}{(N-1)!\,(N-\rho)^2}]P(0) \quad \text{.......... (5)}$$

The expected time in the system

$$E[T]=E[X]/\lambda \quad \text{........ (6)}$$

The expected time in the queue

$$E[Tq]=E[Lq]/\lambda \quad \text{............... (7)}$$

4.3 Examples

For M/M/1 model

Example . In Gupta health clinic Nalagarh, the average rate of arrival of patients is 12 patients per hour. On an average, a doctor can serve patients at the rate of one patient every four minutes.

a) Assume, the arrival of patients follows a Poisson distribution and service to patients follows an exponential distribution.
b) Find the average number of patients in the waiting line and in the clinic.
c) Find the average waiting time in the waiting line or in the queue and also the average waiting time in the clinic.

International Journal of Engineering, Pure and Applied Sciences,
Vol. 3, No. 2, June-2018

Solution: Poisson arrivals, Exponential service and a single doctor on service, follows a M|M|1 Queuing model.

The average rate of patient arrival, λ = 12 patients per hour

The average rate of serving a patient, μ = 1 in 4 minutes = 15 patients per hour

The utilization factor ρ = (12/15) = 0.8

Average number of patients in the clinic

$$L_s = \frac{\lambda}{\mu-\lambda} = 4 \text{ patients}$$

Average number of patients in the waiting line

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = 3.2 \text{ patients}$$

Average waiting time in the clinic,

$$W_s = \frac{1}{\mu-\lambda} = 0.267 \text{ hours}$$

Average waiting time in the queue,

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)} = 0.333 \text{ hours.}$$

For M/M/2 model

Example - A small internet café has two computer terminals. The arrival rate of internet users in the café is 10 users per hour. Each user spends 10 minutes on the computer. The arrival and service process follow exponential distribution

- Find the following:

a) What is the probability that both computers are free?

b) What is the probability that a customer can use the computer after arriving?

c) What is the probability that a customer will find no queue on arrival?

d) Find the average number of customers in the system?

Solution The internet café has two computers, means c=2

The arrival and the service time follow exponential distribution, hence it is a M|M|2 queuing model

(a) Probability that both the computers will be free: We are interested to find the probability of no customers in the café i.e.

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!}\left(\frac{1}{1-\frac{\rho}{c}}\right)\right]^{-1}$$

$$= [1 + \rho + \frac{\rho^2}{2\left(1-\frac{\rho}{c}\right)}]^{-1}$$

c=2, λ=10 users per hour, μ=6 users per hour

ρ = 10/6 = 1.666

P_0 = 0.0909 or 9.1%

(b) P (a customer can use a computer after arriving)

= P (no customer in the café) +P (one customer in the café)

= P_0 + ρP_1 = 0.2424 or 24.2%

(c) P (No queue on arrival)

= $P_0 + P_1 + P_2$

= $P_0 + \rho P_0 + \left(\frac{\rho^2}{2}\right) P_0$

= 0.3687 or 36.9%

(d) Average number of customers in the system

$$L_s = L_q + \rho$$

$$= \frac{\rho^{c+1}}{(c-1)\times(c-\rho)^2} \times P_0$$

= 3.787 users

5. BENEFITS AND LIMITATIONS OF QUEUE THEORY

- **Benefits of queue theory**

The techniques of queuing theory is applied for the solutions of a large problems :

1) It is used in the problem of scheduling of work and jobs in production control.
2) It is used for scheduling of mechanical transport fleets.
3) It is used for minimization of congestion due to traffic delay at tool booths.
4) It is used for routing and scheduling of salesman and sales efforts.
5) It is used for scheduling of aircrafts at landing and take-off from busy airports.
6) It is used for scheduling of issue and return of tools by workman from tool cribs in factories.
7) Queuing theory provides models that are capable of influencing arrival pattern of customers or determines the most appropriate amount of service or number of service stations.
8) Queuing models essentially relates to study of behaviours of waiting lines via mathematical techniques utilizing concept of stochastic process.
9) Queuing theory attempts to formulate, interpret and predict for purpose of better understanding the queues and for the scope to introduce remedies such as adequate service with tolerate waiting.

- **Limitations of queue theory**

Following are the limitations of queuing theory:-

1) Queuing problems are generally complex in nature and cannot be easily solved. The element of uncertainty is there in almost all queuing situations.
2) On multi – channel queuing situations, the analysis become more difficult when the departure from one queue becomes the arrival of the another queue.
3) In most of the cases, the observed distributions of service times and time between arrivals cannot be fitted by

International Journal of Engineering, Pure and Applied Sciences, Vol. 3, No. 2, June-2018

mathematical distributions as we generally assume while solving queuing problems

4) Queuing models with variable arrivals and / or service times are so complicated in nature that they cannot be handled by formulae and mathematical methods. In such situations Monte – Carlo method of simulation is a better answer

.

6. CONCLUSION

Queuing theory can be applicable in many real world situations. For example, understanding how to model a multiple-server queue could make it possible to determine how many servers are actually needed and at what wage in order to maximize financial efficiency. Or perhaps a queuing model could be used to study the lifespan of the bulbs in street lamps in order to better understand how frequently they need to be replaced.

REFERENCES

[1]. **Y. Baba, on the M[X]/G/1** Queue with Vacation Time, *Operations Research Letters*, 5 (1986): 97 - 98.

[2]. **N.T.J. Bailey,** On Queuing Process with Bulk Service, *Journal of the Royal Statistical Society*, 16 (1954): 80 – 87.

[3]. **U.N. Bhat Embedded Markov Chain** Analysis of Bulk Queues, *Journal of the Australian Mathematical Society*, 4 (1963): 244 – 263.

[4]. **P. Borthakur and J. Medhi, Poisson** Input Queuing System with Startup Time and Under Control-Operations Policy. *Computers and Operations Research*, 14(1987): 33-40.

[5]. **K.M. Bretthaner, M.J. Cote**, Nonlinear Programming for Multiperiod Capacity Planning in a Manufacturing System, *European Journal of Operational Research*, 96(1) (1997): 167-179.

[6]. **J.J. Buckley**, Elementary Queuing Theory Based on Possibility Theory, *Fuzzy Sets and Systems*, 37 (1990): 43 – 52.

[7]. **J.J. Buckley, T. Feuring, Y. Hayashi**, Fuzzy Queuing Theory Revisited, *Int. J. Uncertain, Fuzzy* 9 (2001): 527 – 537.

[8]. **J. Buzacott and J. Shanthi Kumar**, *Stochastic Models of Manufacturing Systems*, Englewood Cliffs, NJ(1993) : Prentice Hall.

[9]. **S. Chanas, M. Nawakowshi**, Single Value Simulation of Fuzzy Queues, *Fuzzy Sets and Systems*, 46 (1992): 321-330.

[10]. **X. Chao and Y. Zhao**, Analysis of Multiple-Server Queue with Station and Server Vacation, *European Journal of Operational Research*, 110(1998): 392-406.