DETERMINANTS OF STRUCTURE OF SERVICE SYSTEM THROUGH A GENERALISED QUEUE SYSTEM APPROACH

BY

MUSTAPHA, Adeniyi Mudashiru (Ph.D): Department of Business Administration, Al-Hikmah University Ilorin, Kwara State; E-mail: mustymud@yahoo.com

ISHAQ, Adam Ajimoti (Ph.D): Department of Physical Sciences, Al-Hikmah University Ilorin, Kwara State; E-mail: kadam4best@gmail.com
&

LAWAL, Azeez Tunbosun (Ph.D): Department of Business Administration, Al-Hikmah University Ilorin, Kwara State; E-mail: atlawal@alhikmah.edu.ng

Abstract
The issue of appropriate structure of service system to adopt is becoming a serious challenge to the designer of a system. A good system design has the tendency of ensuring high productivity and efficiency. This paper examines the determinants of structure of service system through queue system approach. This study is one of the several works that recognises that the cost element must be taken into consideration whenever an organisation attempts to design a structure for the service system. To design a system that is more effective, the paper concludes that the format depends on the volume of customers served, randomness of arriving population and partly on the restrictions imposed by sequential requirements as to the order in which it must be performed. The paper recommends that to design a system that is more effective in terms of being able to reduce the waiting time in the queue and ensure a quick service delivery an appropriate system structure must be put in place based on size, volume of calling population and cost.
Keywords: Customers, Service System, Structure of Service and Queuing Model.

Introduction

Queues or waiting lines are observed in everyday life in order to bring orderliness into the society in which we live. Speaking in a similar vein, Grupta and Hira (2016) posited that waiting lines or queues are omnipresent. They observed further that business of all times, industries, schools, hospitals, cafeterias, bookstores, libraries, banks, post offices, petrol pumps, theatres, all have queuing problems.

Davis, Mckeown and Rakes (1986) reported that to complete the operating characteristics, the user must specify certain parameters of the queuing system, such as how the units arrive to be served and how the actual service is handled. Description rather than optimization is the objective of queuing models, and any optimization that takes place must be done by the user varying the system parameters to obtain different sets of operating characteristics.

Lawrence and Pasternack (2002) in their introductory note to queuing models posited that the objective of a queuing analysis is to design system that enables an organization to perform optimally according to some criterion. One possible criterion is to maximize profit. It is not possible to maximize profit through queuing model. One cannot expect the criteria to be the same for a consumer product industry and defence industry or ordinance factory, an outpatient check up ward and a casualty section in a hospital (Chary, 2007).

The main goal of queuing management is to maximize the level of satisfaction with the service provided. Therefore, the primary issue in queuing management and customer satisfaction is not the actual amount of time a customer waits for service but the customer's perception about that wait and the associated level of dissatisfaction. A highly satisfied customer will very likely provide repeat business and spread the positive experience by word of mouth (advertising), resulting in increased revenues and profitability (Dabholkan, Thorpe, & Rentz, 1996). Conversely, a dis-satisfied customer will most likely not provide repeat business and will be more than willing to share his or her bad experience with whoever will listen. This will have an obvious negative impact on revenues and profits (Davis & Heinete, 1994; Hannagam, 1995; Prabhu, 1997)

Speaking in the same vein, Fournier and Mick (1999) expressed that customer satisfaction is an important, theoretical as well as practical issue for most marketers and consumer researchers. Taylor and Baker (1994) posited that customer satisfaction is widely recognized as the key influence in the formation of customer future purchase intention. The above views have been reinforced by File and Prince (1992) when they submitted that satisfied customers are likely to tell others about their favourable experience and thus engage in positive word of mouth advertising. Dissatisfied customers on the other hand, are likely to switch brand and engage in negative word of mouth advertising. In addition, behaviour such as repeat purchase and word of mouth directly affect the viability and profitability of a firm (Dabholkar, Thorpe, & Rentz, 1996).

Giffin (1978) reported that the objective of queue modelling analysis is to minimize the ultimate total cost of time lost by the customers and the units waiting for service; and time lost and its opportunity cost in providing the service to arriving units. It is apparent that with increases in the number of servers or service facilities, the average time spent by each unit or customer in a service system can be reduced. Thus, the goal of queuing modelling according to Vohra (2007) is the achievement of an economic balance between the cost of providing service and the cost associated with the required time for that service.

The specific objective of this study is to examine the determinants of structure of service system, with a view to provide alternative options to management while designing a structure. According to Daellenbach and George (1978) the physical structure of waiting lines consists of three components, which are one or several sources of arrivals, queues and a service facility consisting of one or several parts.

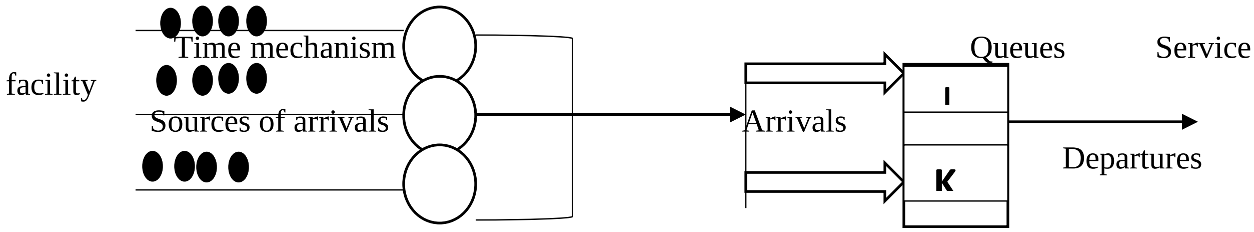


Figure 1: Diagram showing the physical structure of waiting lines.
Source: Adopted from Daellenbach and George, 1978.

The above diagram indicates possible structures that an organisation could adopt based on the source of arriving population and volume. Where the source of calling population is one and the size of the organisation is small, single channel single phase structure will be appropriate. However, where the size of the organisation is large, with high traffic of calling population, multiple channels and multiple phase service structure will be more appropriate. Arrival may originate from one or several sources or pools of potential customers, referred to as the calling population. Also assume that each source has a well-defined arrival pattern over time. Arrival maybe uniformly spaced overtime, that is, the inter-arrival times are constant, or randomly spaced over time with a known inter-arrival time probability distribution. A calling population may be inexhaustible in the sense that the number of potential customers in the sources is assumed to be always very much larger than the number of units in the system, or a calling population may be sufficiently limited in size that the arrival pattern varies as a function of that size.

Arrival process according to Vohra (2007) may be classified according to source which can be infinite or finite. It could also be according to numbers in which the customer may arrive for service individually or in groups. Also according to time where the customers may arrive is known (regular or otherwise) times, or they might arrive in a random way. Arrival might follow any pattern, the frequently employed assumption, which adequately support real world situations, is that the arrival are Poisson distributed. Lawrence and Pasternack (2002) gave the following condition for a Poisson arrival process which is orderliness, stationarity and independence.

There may be nine, one, or several queues. The manner in which units take from the queue is called queue discipline. It may be on a first-come-first-serve basis, random, or subject to service priorities, as for instance in an emergency clinic. It may also be possible to switch queues, and the choice of which queue to join may be open to arrival. There may also be bribing or cheating by some customers for queuing position (Vohra, 2007). Furthermost, the maximum queue length may be unlimited or finite. Potential arrivals may balk if the queue length become excessive and decide not to join, or may join the served. In either case, they are lost to the system. If no units are waiting most of the time, then the services facility will tend to be idles for a large portion of time. If there are costs associated with the idle service channels, then this is undesirable. On the other hand, if the service facility is busy and queue exist most of the time, arrivals will frequently have to wait prior to service (Vohra, 2007). If the waiting times are long, this may again result in tangible or intangible costs, such as lost in production time, deterioration of certain attributes of arrivals or loss of good will.

The service facility may consist of one or several situations or channels. They may operate either in parallel, in which case an arrival has to go through one channel only before being discharged from the system, or they may operate in series, in which case an arrival has to go through several channels in sequence before being discharged. The service time at each channel may be constant or random with a known service time distribution. There are two aspects of a service system. They are structure of the system service and the speed of service (Vohra, 2007). Structure of the service system means how the service facility exists. There may be a single service facility, a multiple, parallel facility with single queue, multiple parallel facility with multiple queues and services facilities in a series. In a queuing system, the speed with which service is provided can be expressed in either of two ways; as service rate and as service time. The service rate describes the number of customers serviced during a particular time, while the service time indicates the amount of time needed to service customers. Service rates and times are reciprocals of each other and either of them is sufficient to indicate the capacity of the facility (Bunday, 1976).

Stafford (1979) while describing the elements of the queuing situation revealed that individuals arrive at the end of the queue, wait in the queue, receive the service they have been waiting for and they leave the system. Schematically, the situation is given thus:

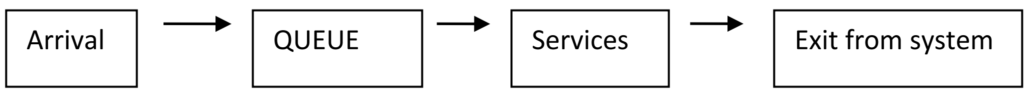


Figure 2: Diagram showing Sequence of getting service
Source: Adopted from Stafford, 1979.

While contributing to the characteristics of queuing model, Chase and Aquilano (1991) identified six major components of queuing theory: the source population, the way customers arrive at the service facility, the physical line itself, the way in which customers are selected from the line, the characteristics of the service facility itself (such as how the customers flow through the system and how much time it takes to serve these customers) and the condition of the customers' existing in the system. An analysis of queuing is done considering calling population, actual waiting lines and service facility. Salami (1989) reveals that the number of potential customers, which could be finite or infinite, is referred to as "calling population". The analysis of queuing being considered is that of infinite calling population. Usually, an infinite calling population can be assumed if the rate of arrival of future customers is not affected by the number of customers already on the queue.

Levin, Rubbin and Stingson (1986) asserted that the arrival pattern could be either random or in an organized pattern. When arrival are random, it is required that the probability distribution describing arrivals, specifically inter-arrival times must be known. Arrivals following random patterns are usually described by the statistical poison distribution (which may not be the case always). The statistical Poisson distribution (T) being referred to, represents the total time period under consideration.

It approximates a normal distribution, and is skewed to one side. Trueman (1977) expressed that the population source in terms of its size and behaviour shapes the demand placed on a service system. He indicated that when the number of potential arrivals is dependent on the number of units currently in the system, (those being served plus those in queue), the calling population is limited, or finite. The calling population is considered to be unlimited (no inherent limitation) or infinite, when the number of arrival are independent of the number of units currently in the system.

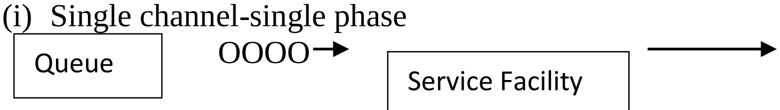
The behaviour of the calling population can be described by the arrival pattern and by how units act before and after they join a waiting line. An arrival pattern refers to the size of arrivals between successive arrivals. Units may arrive in a facility one at a time. The time between successive arrivals maybe constant, or variable, but usually the time between arrivals varies at random according to some probability distribution. For many service facility we can assume that units arrive for service according to Poisson distribution with a means rate equal to л. The Poisson distribution is suitable for low probability events that have repeated opportunities to occur. It is very useful in many decision areas in operations management such as inventory control, maintenance, e.t.c.

The Poisson assumption is appropriate whenever the probability of an arrival at any instant is very small but there are many opportunities for it to happen. Units arrive at random independently of each other, with the same average rate л. The probability of two or more arrivals in a very short time, say one second is almost zero. Whenever units arrive at a service facility according to a Poisson distribution with a mean arrival rate equal to λ, the times between arrivals follow a negative experimental distribution with a mean equal $1/\lambda$, and vice-versa. The units exhibit different attitude when getting into a queue. The varying attitude has been identified as reneging, balking or jockeying. Arriving customers are considered to be patient when they will wait for the service regardless of the number in queue. A customer, dissatisfied by the delay in service, who leaves the queue (reneges), is termed impatient. Customers who move from one queue to another hoping to receive service more quickly are said to be jockeying. A customer who balks will not wait for service if the queue is too long. He may or may not return later. When arrivals collaborate to reduce overall waiting time, their behaviour is termed collaborate.

Queue are analyzed in terms of the maximum length it is capable of having. This length can either be limited or otherwise. When queue are limited, it is a result of lack of space or attitude of the units in calling population. However, in more realistic queuing analysis, queue lengths are assumed to grow at an infinite rate. Cohen (1985) indicated that the characteristics of all queuing system in which customers require just one service are the arrival pattern, service pattern, queue discipline or service discipline and maximum number of customers allowed in the system.

According to Cohen (1985), the three characteristics outlined above, arrival pattern, service pattern and queue discipline, must be identified before a given system can be analyzed. The arrival pattern describes how the queue grows, the services pattern determines how fast it diminishes, while the queue discipline specified how the items or people taken from the queue are being processed and attended to. In his contribution to the structure of queues, Buffa (1972) highlighted four basic structure of waiting line situation that described the general conditions of all servicing facility. The simplest situation is the arriving from a single line to be serviced by a single processing facility, for instance, one man barbershop. This is called the single channel or single phase case.

(i) Single channel-single phase



(ii) Multiple channel-single phase

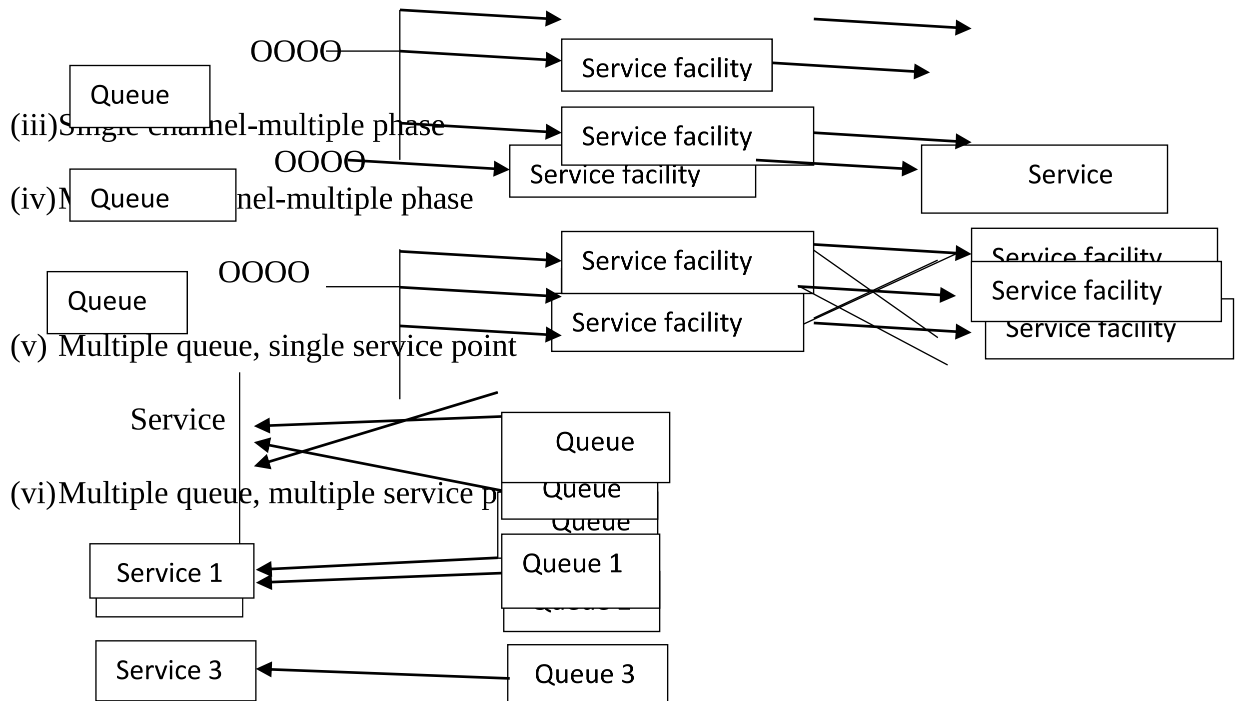


Figure 3: Diagrams (i-vi) highlight the general conditions of all servicing facilities available for an organisation
Source: Adapted from Buffa, 1972.

Determinants of the Structure of Service System
The following features are considered when examining servicing facilities, the physical layout of the system, queue discipline and the appropriate probability distribution describing service times. The physical flows of items to be served may go through a single line, multiple lines, or some mixture of the two. The choice of format depends partly on the volume of customers served and partly on the restrictions imposed by sequential requirement as to the order to which it must be performed.

The queue disciplines in a service facility are important to the analyst of queues. When the first customer in line is the next customer served, we have a first-come-first-served (FIFO) queue discipline (Gross et al, 2013; Fakokunde, Adeniyi & Aremu, 2017). When the last customer in line is the next customer served, we have a last-come-first-served (LIFO) queue discipline. When the units requiring service receive partial service in sequential order, such as programs being run on time shared terminals, we have a round-robin queue discipline. The queue discipline refers to which, how and what units in the calling population receives what service how and when. They observed that queue discipline can be categorized into two; priority and First come, first served.

The first-come-first-served queue discipline is very much common in most contemporary service systems, for example banks, supermarkets and hospitals. The priority discipline can be regrouped into two (pre-emptive and non-pre-emptive). Pre-emptive discipline permits and allows units to interrupt/break the continuity of units already receiving service; on the contrary, non-pre-emptive queue discipline arranges the queue so that the unit with the greatest priority rating gets served before others (Fakokunde et al., 2017).

Decision on the structure and the nature of queues would be incomplete and remain an arm chair academic exercise without considering the associated probability distribution describing the service times. Service time may be constant or random in nature. Again, where service time are random by distribution, they are better described by the experimental probability distribution (the actual distribution of some services pattern may however be different).

From the structure discussed above, the cost elements were not taking into consideration. Scholars are now beginning to look at the implication of cost on the structure of queues. The queuing system also (known as processing system) can be characterized by four main elements: the arrival, the queue discipline, the service mechanism, and the cost structure (Fakokunde, et al., 2017). The arrival is the way by which customer arrives and enter the system for service (Cooper, 1972). The queue discipline is the rule for determining the formation of the line or queue and the order in which jobs are processed. The service mechanism describes how the customers are served. The cost structure specified the payment made by the customer and the various operating costs of the system. Other elements that impact the queue structure and performance include the number of service stations and the number and speed of servers.

Methodology

The model for this study is queuing model. The model is considered appropriate because in general, components of queuing theory consist of calling population, actual waiting lines and service facility. Supporting this position, Cohen (1985) opined that before a given system can be analyzed, the three characteristics, arrival pattern, service pattern and queue discipline must be identified.

Conclusion

To design a system that is more effective, the paper concludes that the format depends on the volume of customers served, randomness of arriving population and partly on the restrictions imposed by sequential requirements as to the order in which it must be performed. Organisations, whether profit making or not- for-profit aspires to achieve efficiency and effectiveness in terms of its ability to meet customers' demand. Poor system design has led to frustration, irritation, health hazard and general disenchantment of customers as opined by Gormey, (1981). To achieve the objective of a queuing analysis, management must therefore designs a system that enables an organization to perform optimally according to some criterion.

Recommendations

1. The paper recommends that to design a system that is more effective in terms of being able to reduce the waiting time in the queue and ensure a quick service delivery
2. An appropriate system structure must be put in place based on size, volume of calling population and cost. Management should therefore be guided by the physical queue discipline and the appropriate probability describing service times.
3. The physical flows of items to be served may go through a single line, multiple lines, or some mixture of the two.

References

Aquilano, N., & Chase, R. (1991). *Fundamentals of Operations Management*. Homewood, IL: Irwin.

Buffa, E.A. (1994). *Meeting the Competitive Challenges-Manufacturing Strategy for US Companies*. Hannewood III: Dowjones Irwin.

Bunday, J.P. (1976). *An Introduction to Queuing Theory*. New York; Housted Press.

Chary, S.N. (2007). *Production and Operations Management.* 3rd Edition. New Delhi : Tata McGraw Hill Publishing Company Ltd.

Cohen, S.S. (1985). *Operations Research*. London: Edward Arnold.

Cooper, R.B. (1972). *Introduction to Queuing Theory*.(2nd Ed.), New York: Macmillian

Dabholkan, P.A., Thorpe, D.I. & Rentz, J.O. (1996). A measure of service quality for retail stores: Scale development and validation. *Journal of the Academy of marketing Science*. *24*(1), 2-16.

Daellenbach, R. T & George, J.A (1978). *Introduction to Operations Research Techniques. Boston:* Allyn and Bacon Inc.

Davis, K.R. & McKeown, P.G. & Rakes, T.R. (1986). Management Science: An Introduction. Boston, Mass: Kent Pubublishing Company

Davis, M.W. & Heinete, J. (1994). Understanding the role of customer and the operation for better queue management. *International Journal of Operations and Production Management in Hospital Pharmacy. 36*(3), 21-34

Fakokunde, T.O., Adeniyi, A.M. & Aremu, M.A. (2017). Understanding the queuing theory for improved service delivery : An entrepreneurial-mindset approach . *EJBE, 6*(1), 33-47.

File, K.M., & Prince, R.A. (1992). Positive word of mouth; Customer satisfaction and buyer behaviour. *Internal Journal of Bank Marketing ,10*(1), 25-29.

Fournier, S., & Mick, D.G. (1999). Recovering satisfaction. *Journal of Marketing*. 63(4), 5-23.

Giffin, W. (1978). *Queuing: Basic theory and Application*. Ohio: Grid.

Gormey, L. (1981). *Queuing Theory: A problem Solving Approach*. New York: Petrocoili Books Inc.

Gross, D., Shortie, J.F, Thompson, J.M. & Harris, C.M. (2013).*Fundamentals of Queuing Theory*. New York: John Wiley and Sons

Gupta, P.K., & Hira, D.S. (2016). *Operations Research.* New Delhi, India: S Chand and Company Limited

Hannagam, T. (1995). *Management Concepts and Practices*. London: Pitman Publishing.

Lawrence, J.A. Jr. & Pasternack, B.A (2002). *Applied Management Science Modelling Spread Sheet Analysis and Communication for Decision Making*. John Wiley & Sons, Inc.

Levin, R.I, Rubbin, O.S. & Stingson, J.P. (1986). *Quantitative Approach to Management*. International Edition. Singapore: McGraw Hill

Prabhu, N.U. (1997). *Foundations of Queuing Theory*. N.Y.: Huwer Academic Publishers.

Salami, O.S. (1989). *Application of Queuing Theory to Service System*. A case study to the outpatient department, Ilorin: An unpublished Research Project report Submitted to the Department of Business Administration, University of Ilorin.

Stafford, L.W.T. (1979). *Business Mathematics (2nd Ed*). Pittman publication Ltd.

Taylor, S.A. & Baker, T.L. (1994). An Assessment of the relationship between service quality and customer satisfaction in the formation of customers purchase intentions. *Journal of Relating*. 70(2),163-178

Trueman, R. E. (1977). *An Introduction to Quantitative Methods for Decision Making (2nd Ed).* New York: Holt, Rodnehart and Winston.

Vohra, N.D. (2007). *Quantitative Techniques in Management*. (3rd Ed), New Delhi, India : Tata McGrow-Hill Publishing Company Ltd.