Queuing Model as a Technique of Queue Solution in Nigeria Banking Industry

Anichebe Nnaemeka Augustine[1]
08033536590: Emekanichebe @Yahoo.Co
Management, University Of Nigeria, Enugu Campus

Abstract

The study of "Queuing model as a Technique of Queue solution in Nigeria Banking Industry" was carried out. The obvious cost implications of customers waiting range from idle time spent when queue builds up, which results in man-hour loss, to loss of goodwill, which may occur when customers are dissatisfied with a system. However, in a bid to increase service rate, extra hands are required which implies cost to management. The onus is then on the management to strike a balance between the provision of a satisfactory and reasonable quick service and minimizing the cost of such service. Thus, the management should evaluate performance of different queuing structures and strike a middle ground between cost on one hand and benefits of improved service on the other hand, which is the main thrust of this study. Therefore, this study attempts to look at the problem of long queues in blanks, why bank managers find it difficult to eliminate queues and the effect of queuing model as a technique of queue solution in Nigerian Banking Industry. Descriptive research method was employed in carrying out the study at United Bank for African Pie., Gariki Branch, Enugu, through observation, interview and questionnaire administration. The variables measured include arrival rate (λ) and service rate (μ). They were analyzed for simultaneous efficiency in customer satisfaction and cost minimization through the use of a multi-channel queuing model, which were compared for a number of queue performances such as; the average number of customers in the queue and in the system, average time each customer spends in the queue and in the system and the probability of the system being idle. It was discovered that, using a three-server system was better than a 2-server or 4-server systems in terms of the performance criteria used and the study inter-alia recommended that, the management should adopt a three-server model to reduce total expected costs and increase customer satisfaction.

Keywords: Queuing Model, Queue Solution, Banking Industry

Introduction

Every relationship is a game and banker-customer relationship is not an exception. As the world turns to a global village characterized by intense and ever increasing competition, bank operation managers continue to experience wrenching challenges, which they must keep up with for survival. Banks being major component of the financial system, intermediating between the surplus and deficit sectors of the economy are always center of attraction to many customers that want to carry out one transaction or the other through the services provided by these banks. However, the major problem is the inability of the banks to match their service facilities to the needs of customers without much delay. The common experience in Nigeria is that most banks do not have the facilities and capacities to service the number of customers without much delay on the part of the customers. The problem in this regard had been that though bank customers for instance, have always been desirous of spending the least possible time in banking transactions, this age-long desire is yet to be met by the banks. Banks on the other hand, want to attract, retain customers and at the same time optimize profit. Profit making in banks is a function of management ability to provide efficient services to customers at little or no time wastage (Agbadudu, 1995).

One major recurring problem in Nigerian Banks is the overcrowding banking halls, this had led to the movement of customers from one bank to the other, where they can obtain banking services without much delay. The modem day computerization, (such as Online Real Time, Automated Teller Machine (ATM) etc) by banks in attempt to minimize waiting line problem have not yielded the much desired result due to frequent breakdown of such computerization and networking arrangements. Hence, long queue persisted in all Nigerian banks. More time and money for one thing, means less time and money for other things. The time available to any economic agent has alternative uses, for leisure or for work (Ashley, 2006).

The number of hours devoted to work determines the individual's wage. Apart from situation of work or leisure, economic agents sometime commit considerable amounts of their time when they come into service stations for service. Typical of Nigeria, customers wait for hours to get service without the service providers feeling that there is an opportunity cost for the wasted time. Maybe, the lack of economic growth in the country is traceable to this economic waste and the economy will begin to grow if this waste can be minimized.

Statement of the Problem

Waiting for services is part of our daily life. We wait to eat at restaurants, we "queue up" at the check-out counters in grocery stores and we "line up" for service in post offices, banks and petrol filling stations. The waiting phenomenon is not an experience limited to human beings only; jobs wait to be processed on a machine and cars stop at traffic lights. Unfortunately, we cannot eliminate waiting without incurring inordinate expenses. In fact, all we can hope to achieve is to reduce the adverse impact of waiting to acceptable levels. In a traditional non-queuing environment, customers can be left confused as to what line to stand in, what counter to go to when called and distracted by noisy crowded environment.

In situations where facilities are limited and cannot satisfy the demand made upon them, bottlenecks occur which manifest as queue but customers arc not interested in waiting in queues. When customers wait in queue, there is the danger that waiting time will become excessive leading to the loss of some customers to competitors (Kotler, 1999). But allowing them to serve themselves so easily is a key factor in both keeping and attracting customers (Michael, 2001).

The days of a customer adopting one product or company for life are long gone. With easy access and global competitiveness, customers are often swayed by advertising and a chance at a better deal. Quality levels and features between competing brands and organizations are often comparable. The thing that separates competitors is their level of service. It is not unusual for customers to switch back and forth between products or organizations simply because of pricing, a bad first impression or lack of quality service.

The obvious cost implications of customers waiting range from idle time spent when queue builds up, which results in man-hour loss, to loss of goodwill, which may occur when customers are dissatisfied with a system. However, in a bid to increase service rate, extra hands are required which implies cost to management. The onus is then on the management to strike a balance between the provision of a satisfactory and reasonable quick service and minimizing the cost of such service. Thus, the management should evaluate performance of different queuing structures and strike a middle ground between cost on one hand and benefits of improved service on the other hand, which is the main thrust of this study. Therefore, this study attempts to look at the problem of long queues in banks, why bank managers find it difficult to eliminate queues and the effect of queuing model as a technique of queue solution in Nigerian Banking Industry.

OBJECTIVES OF THE STUDY

The primary objective of this study in line with the identified problems is to determine whether the present capacity level in the banking industry, using United Bank for Africa Pie. as a case study, strikes a balance between cost of waiting and cost of providing service. This was carried out by measuring:

(a) The number of customers waiting for service.
(b) The processing time and
(c) The probability that the facility will be idle.

The study specifically aims to determine:

(i) The amount of waiting time a customer is likely to experience in a system;
(ii) How the waiting time will be affected, if there are alterations in the facilities; and
(iii) Make policy recommendations based on the findings from the study.

Review of Related Literature

Evolution of Queuing Theory

Queuing theory was developed to provide mathematical models to predict behaviour of systems that attempt to provide service for randomly arising demands can trace its origin back to a pioneer investigator, A.K. Eriang a Danish Mathematician, who in 1909, published "'I he 'theory of Probabilities and Telephone Conversations", based on his work for the Danish Telephone Company in Copenhagen, Denmark (Gross and Harris, 1998). Operations Researchers took over the vanguard of advancing the analytic development of queuing theory. The terminology of queuing theory reflects the origins of the field (Prieger, 2001). Although the early work in queuing theory picked up momentum rather slowly, the trend began to change in the 1950s, when the pace quickened and the application areas broadened well beyond telephone systems. There are many valuable applications of the theory, including traffic flow (vehicles, aircraft, communication), scheduling (Patients in hospitals, jobs on machines, programs on computer), facility design (banks, post offices, fast-food restaurants). Today, we encounter a myriad of queues everyday of our lives, and queuing theory when it can, helps us to navigate around these (Gross and Harris, 1998).

Conceptual Farmework

A queuing system can simply be described as customers arriving for service, waiting for service if it is not immediate and if having waited for service, leaving the system after being served. The term "Customers" is used in a general sense and docs not imply necessarily a human customer. For example, a customer could be a ball bearing waiting to be polished, an airplane waiting in line to take off, a computer program waiting to be run, or a telephone call waiting to be answered (Taha, 2001). Queuing theory is the formal study of waiting line and is an entire discipline in operation management. Queuing theory utilizes mathematical models and performance measures to asses and hopefully improve the flow of customers through the waiting line (Bunday, 1996; Prabhu, 1997 and Gomey, 1981).

Queuing theory is also a set of tools and techniques for analyzing such problems, concerned with providing service to customers so as to have balance of the cost of waiting and cost of servicing customers in a line (Agbadudu, 1996).

Generally, a queuing or waiting-line problem arises whenever the demand for customer service cannot perfectly be matched by a set of well-defined service facilities that is, there is more demand for service than there is facility available for service. There may be many reasons such as shortage of available services, economically infeasible for a business to provide the level of service necessary to prevent waiting or limitation to the amount of service that can be provided.

Generally, these limitations can be removed with the expenditure of capital. To know how service should be made available, one need to know answers to such questions as "How long will a customer wait?" and "How many people will form the line?" Queuing theory attempts to answer these questions through detailed mathematical analysis and in many cases it succeeds (Gross and Harris, 1998).

In Nigeria, a study conducted by Oladapo (1988), revealed a positive correlation between arrival rates of customers and bank's service rates. He concluded that the potential utilization of the banks service facility was 3.18% efficient and idle 68.2% of the time. However, Ashley (2006), asserted that even if service system can provide service at a faster rate than customers arrival rate, waiting lines can still form if the arrival and service processes are random.

One week survey conducted by Elegalam (1978), revealed that 59.2% of the 390 persons making withdrawals from their accounts spent between 30 to 60 minutes, while 7% spent between 90 and 120 minutes. Baale (1996), while paraphrasing Alamatu and Ariyo (1983), observed that the mean time spent was 53 minutes, but customers prefer to spend a maximum of 20 minutes. Their study revealed worse service delays in urban centres, averaging 64.32 minutes, compared to 22.2 minutes average in rural areas. To buttress these observations, Jawah (1986), found out that customers spend between 55.27 to 64.56 minutes making withdrawals from their accounts. Efforts in this study are directed towards application of queuing models in capacity planning to reduce customer waiting time and total operating costs.

Classification of Queuing System

A queue for the purpose of this study is the aggregation of customers waiting a service function. It is an everyday occurrences and results when the number of calling units exceed the number of available service centres. Queues are integral parts of any service system, which refers to the whole situation from arrival of inputs to their departure.

The variants of queuing models that can be applied to waiting problems include; simple system, multi-channel system, constant service and limited population models (Ashley, 2000). A simple system is a single line and single server system, which consists of items forming a single queue, which is served by a single facility. A Multi-channel system is a system where two or more servers or channels are available to handle arriving customers needs. Here, a common line is formed and the customer at the head of the line proceeds to the first free server. There is also a Constant Service Time Model. This is where the customers service time is constant instead of exponential distributed times and lastly, the Limited Population Model. The Limited Population Model is a dependent relationship between the length of the queue and the arrival rate. Among all these models, those that can be applied to solving customer waiting problems in banks are the Simple and Multi-channel Models.

Characteristics of a Waiting Line System

The three main characteristics that determine the appropriateness of a queuing model are; the arrival to the system, queue discipline and the service facility (Safe Associate, 2002). Arrival characteristics of a waiting line is related to size, pattern of arrival and service time distribution. The size of the source population may be unlimited (infinite), when the number of arrivals on hand at any given moment is just a small portion of the potential arrival or limited when the population size is relatively known.

The pattern of arrival at the system is considered random, when arrivals are independent of one another and their occurrence cannot be predicted exactly. Also, the number of arrivals per unit of time in queuing problems can be estimated by a Poisson probability distribution. The service time distribution may be constant when it takes the same amount of time to take care of each customer or random when the reverse holds. If arrival rate is Poisson distributed, a negative exponential probability distribution is assumed for random service times.

Queue Discipline

This refers to the priority rule by which customers are served, that is, the order in which items received service. According to Jay and Barry (1993), there are two main categories, that include:

(1) Preemptive Priority**:** This is common in emergency situations, which allows customers that arrive at any time to replace customers that are being served, example in patient treatment in hospitals.

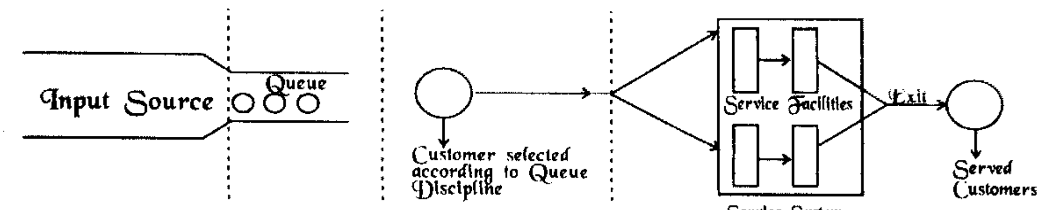
(2) Non-Preemptive Priority**:** Here, items in the queue are arranged so that the item with the highest priority in the system is served first and there is no displacement of items in service. The methods here include:

- (i) FIFO (First In-First Out): Allows the first item to enter the system (items at the head of the queue), to be served first. It is the most frequently applied discipline because it is believed to be fairer than the other types of rules.
- (ii) LIFO (Last In-First Out): Here, the last item of the queue or that which enters the system last is served first.

Queue Structure

This describes how the customers are served. It is characterized by the configuration of the service facility and the service distribution. The configuration may be designed in terms of channels and phases; The channels (number of entry point/s to the service system), may be single or multiple. The phases (the number of service stations through which the customer must pass before the service is considered complete), may also be single or multiple phases depending on the nature of service.

Hypothetical Structure of a Queue



Source: Smith (1999).

Safe (2002), limited the characteristics of a single server to include:

(a) A single service channel and variables arrival following a Poisson distribution While Service Time Is Exponentially Distributed.

(B) Fifo Queue Discipline Is Adopted Because Of Infinite Population Of Potential Customers That Has No Simultaneous Arrival And There Is No Balking Or Reneging.

Characteristics of Multi-Server System

According To Safe (2002), A Multi-Server System Has All The Features Of Simple Queue And In Addition, Assumes No Limit To The Permissible Length Of The Queue. Also, All Servers Are Assumed To Perform At The Same Rate. This Is A Model In Which Two Or More Servers Or Channels Are Available To Handle Arriving Customers. Customers Awaiting Service Form One Single Line And The Customer At The Head Of The Line Proceeds To The First Available Free Server. The Equations Are Obviously More Complex, But They Are Used In Exactly The Same Manner And Provide The Same Type Of Information As The Simple Model. Computer Software Packages Are Most Useful In Solving Multiple Channel As Well As Other Queuing Problems.

Method and Materials

Research Method

The Nigerian Banking Industry Most Bedeviled With Customer Waiting Problems Is Studied Here In United Bank For Africa Plc., Through Observation, Interview And Questionnaire Administration. The Variables Measured Include Arrival Rate (X,) And Service Rate (A). They Were Analysed For Simultaneous Efficiency In Customer Satisfaction And Cost Minimization Through The Use Of Multi-Channel Queuing Models, Which Were Compared For A Number Of Queue Performances Such As; The Average Time Each Customer Spends In The Queue And In The System, Average Number Of Customers In The Queue And In The System And The Probability Of The System Being Idle.

In The Realization Of These Objectives, Primary Data In Respect Of Customer Arrival Rate And Cashier/Tellers' Service Rate Were Used And Were Obtained Through Observations While Customer Attitude Survey Was Carried Out Through A Total Of Two Hundred And Forty Six (246) Copies Of Questionnaire Administered On Two Hundred And Forty Six (246) Stochastically Selected Customers For That Purpose. The Population Of This Study Is Infinite And Arrived Rate Was Random And Exponentially Distributed. The Sample Size Of 246 Was Derived Using The Freund And Williams Formular For An Infinite Population And It States Thus:

$$N = \frac{Z^2 pq}{E^2}$$

Where N = Sample Size

Z = Standard Error Of The Mean (1.96)

P = Probability Of Success (0.8)

Q = Probability Of Failure (0.2)

E = Level Of Significance (0.05)

The Variables Were Analyzed Using The Queuing Models Presented In The Table Below.

Note That λ = Arrival Rate And μ , = Service Rate.

Since There Were (5 X 52 Weeks) I.E. 260 Working Days In A Year, With The Exclusion Of Public Holidays, A 251 Working Days Is Used In This Study, While The Working Hours Per Day Is Eight (8).

Formulae for Computing Performance Measures of Queuing System

Multi Channel System Formulae:

Multiple service points with single queue.

1. Traffic intensity = p = λ /Cμ; where C = Number of service points

ISSN 2224-607X (Paper) ISSN 2225-0565 (Online)
Vol.3, No.8, 2013

2. $$P_o = \frac{C!(1-p)}{(pC)^c + C!(1-p)\left\{\sum_{n=0}^{C=1} \frac{1}{n!}(pC)^n\right\}}$$

Where n = integers from zero to one less than the number of channels

3. $$W_s = \frac{(pC)}{C!(1-p)^2\ C\mu} P_o + \frac{1}{\mu}$$

4. $$W_q = \frac{(pC)^c}{C!(1-p)^2\ C\mu} P_o$$

5. $$L_s = \frac{p(pC)^c}{C!(1-p)^2} P_o + pC$$

6. $$L_q = \frac{p(pC)^c}{C!(1-p)}$$

Source: Okeke, 2002.

Data Analyses and Discussion of Results

Analyses of Data

Table 4.1 Presentation Of Results

	Performances Measures	**Multi-Server (2 Server)**	**Multi-Server (3 Server)**	**Multi-Server (4 Server)**
1	Probability Of The System Being Idle (P_o)	3.3%	5.72%	3.96%
2	Average Number Of Customers In The System (L_s)	15.95	3.3	3.52
3	Average Time A Customer Spends In The System (W_s)	58.3	12.1	12.92
4	Average Number Of Customers In The Queue (L_q)	13.86	0.54	-0.006
5	Average Time A Customer Spends In The Queue (W_q)	50.6	1.97	-
6	Total Economic Cost	N18,414.00	N2,069.50	N2,166.91

Source: SPSS, 2010.

Table 4.1 Above Shows That A 3-Server System Is Better Than A 2-Server Or 4-Server Systems In Terms Of The Performance Criteria Used.

The Analyses Of This Study Is Based On Eighteen Thousand Naira (N18,000.00) Minimum Wage. With An Eight (8) Hour Working Day, We Have A Total Of One Hundred And Sixty (160) Hours A Month At One Hundred And Twelve Naira Fifty Kobo (N112.50), Per Hour. For Instance, In Terms Of Cost Considerations, A 3-Server System Recorded Least Cost Of Two Thousand And Sixty Nine Naira Fifty Kobo (N2,069.50), Compared To A 2-Server And 4-Server Systems That Recorded Eighteen Thousand Four Hundred And Fourteen

Naira (N18,414.00) And Two Thousand One Hundred And Sixty Six Naira Ninety One Kobo (N2,166.91) Respectively. The Average Time A Customer Spends In The System And In The Queue Were 12.1 Minutes And 1.97 Minutes Respectively For A 3-Server System. These Were Lower Than That Of A 2-Server System With 58.3 And 50.6 Minutes In The System And In The Queue Respectively. This Was Further Better Than That Of A 4-Server System That Recorded 12.92 Minutes And Negligible Time Spent In The System And In The Queue. In Terms Of Average Number Of Customers In The System, A 2-Server System Had 15.95, While A 3-Server And A 4-Server Systems Had 3.3 And 3.52 Customers Respectively. The Probabilities Of Idleness Were 3.3%, 5.72% And 3.96% Respectively For A 2, 3 And 4 Server Systems Respectively.

Discussion of Results

Multi-Server Models Were Compared And It Was Observed That;

(A) Using A Three-Server System Was Better Than A Two-Server System In All Ramifications. For Instance A Three-Server System Had 0.54 Customer Waiting In The Queue For Service While A Two-Server System Had 13.86 Customers In The Queue. In A Two-Server System, A Customer Spends 58.3 Minutes And 50.6 Minutes In The System And In The Queue Respectively As Against 12.1 Minutes And 1.97 Minutes Respectively For A Three-Server System.

(b) A three-server system has a high probability of being free (idle) 5.72%, than either of two-server or four-server systems.

Similarly, United Bank for Africa Plc was used as a representative of the service industry and in the course of this study, the following were observed. The maximum queue length was 10, which sometimes increase to 35 i.e. 21.14 customers in excess of 13.86 customers expected. Total number of arrivals were 719 per week or 17.98 customers per hour. This means inter-arrival time was 3.34 minutes per customer. The arrival process follows a Poisson distribution, while queue discipline was first-come-first-served (FCFS). Queue population was unlimited and its behaviour includes jockeying, reneging and balking. This study revealed that on the average, 624 customers left before service in a month at a rate of 31.2 per day, but adoption of a three-server and four-server systems will reduce this rate to 12.49 and 5.5 customers respectively. The bank adopted a two-server-paying system and service rate was 9.59 customers per server per hour, while average service time was 7.59 minutes. It was equally observed that while some customers had to wait for as much as two hours before service, some spent only two minutes. Introduction of an additional server is capable of reducing loss of customers by 60% (obtained from interview of staff). It was also noticed that friends of cashiers received quicker service, at the expense of others and teller points were opened and closed at will leading to customers confusion and complaints because their waiting time was higher than expected.

Conclusion

This study uncovered the applicability and extent of usage of queuing models in achieving customer satisfaction at the lowest cost. Customers were unhappy due to delay in service delivery, while the bank was reluctant in increasing the number of service units because the theoretical underpinnings of queuing model was not well understood. Otherwise, the service units would have been increased to three to achieve better results at a lower cost of N2,069.50 as against two server units at a cost of N18.414.00. The current two-server system being used was not optimal. The use of four-server system, although eliminates waiting time, but at a higher cost which is not optimal too.

Recommendations

Based on the summary and conclusion of this study, the following recommendations were made for efficiency and quality of service to customers in the United Bank for Africa Pie.

1. The management should adopt a three-server model to reduce total expected costs and increase customer satisfaction.
2. The queue characteristics should be viewed from the stand point of customers as to whether the waiting time is reasonable and acceptable by making queue discipline fair and varying the number of service channels according to queue circumstances.
3. The management should educate their operation managers and other staff on the application of queuing models to operational problems.
4. They should trust its employees, empower them, enrich their jobs by making them multi-skilled through continuous training to enable them eliminate unnecessary counter-check handoffs, while allowing them to complete many processes in the front line.

References

Agbadudu, A.B. (1996), *Elementary Operation Research.* Vol. I: Benin City, A.B. Mudiaga Limited.

AshleyD.W. (2006), "*Introduction to Waiting Line Models*" Available at Google, http/www.com.

Bunday, B.D. (1996), *An Introduction to Queuing Theory:* New York: Halsted Press.

Elegalam, (1978), "*Customer Retention Versus Cost Reduction Technique",* A Paper presented at the Bankers Forum held in Lagos. Gomey, L. (1981), *Queuing Theory: A Problems Solving Approach* New York: Petrocelli Books, Inc.

Gross, D. and Harris C.M. (1998), *Fundamentals of Queuing Theory.* New York: John Wiley and Sons.

Jay, H. and Barry, R. (1993) *Production and Operation Management Strategies and Tactics, Business and Economic,* Rich Wolf, Allyn and Bacon.

Juwah (1986), *"Operational Problems and Customers ' Dissatisfaction",* CBN Bullion.

Kotler, P. (1999), *Marketing Management, The New Millennium Edition.* Michael, K. (2001), *"Service on Call" in Banking Technology* News Service Network, http/www.fineos.com/documentation/features/bankreach.com.

Okeke, A.O. (2003), *Quantitative Methods for Business Decisions* Enugu, Macro Academic Publishers.

Oladapo (1988), "*Relationship between Arrival Rates of Customers* and Bank's Service Rates", Seminar Paper Series, Obafemi Awolowo University, lle-Ife.

Prabhu, N.U. (1994), *Foundations of Queuing Theory:* Ithaco, NY: Khuwar Academic Publishers.

Prieger, J.E. (2001), *Estimation of a Simple Queuing System with Units-in-Service and Complete Data.* An Unpublished Ph.D. Dissertation, Department of Economics, University of California.

Safe Associate Ltd. (2002) Quantitative Methods, Accounting Technician Series.

Smith, W. (1999) "Working Schedules and Managing Queues" Retrieved from www.academic.emporia.edii/smithwil.

Taha, H.A. (1997), *Operations Research: An Introduction.* New Jersey, Prentice Hall Incorporated.

Taha, H.A. (2001), *Operations Research: An Introduction.* New Delhi, Prentice Hall of India.

Uzoagulu, A.E. (1998), Writing Research Project Reports, Enugu, John Jacob's Classic Publishers Ltd.

Yechiali, U. Altmand, E. and Kofma, D. (1995), *"Discrete Time Queues with Delayed Information* Vol. 19:361-376.