ISSN 1115 - 960X

Ilorin Journal of Business and Social Sciences

Volume 9, Nos. 1 & 2

2004

Published by the Faculty of Business and Social Sciences
University of Ilorin, Ilorin.

Ilorin Journal of Business and Social Sciences

Olaniyi T. A
Accounting & Finance Dept;
University of Ilorin
Ilorin

Pg: 139 - 145

Ilorin Journal

of

pages
139 - 145

Business and Social Sciences

Published by
The Faculty of Business and Social Sciences,
University of Ilorin, Ilorin.

SUBSCRIPTION INFORMATION

	NIGERIA		FOREIGN	
	Naira	*Dollars*	*Pounds*	*Deutschmarks*
Individual	N500	$30	£25	35
Library	N600	$35	£30	40
Students	N500	$30	£30	35

Subscription procedures:

All subscribers should send certified bank cheques or bank draft to

The Editor - in - Chief,
Ilorin Journal of Business and Social Sciences
University of Ilorin, P.M.B. 1515, Ilorin, Nigeria.

CONTENTS

Ilorin Journal of Business and Social Sciences Vol. 9, 1 & 2, 2004

An Appraisal of Cost of Queuing in Nigerian Banking Sector: A Case Study of First Bank of Nigeria Plc; Ilorin.

T. A. OLANIYI

Department of Accounting and Finance, University of Ilorin, Nigeria

Abstract

The endless customers waiting for service delivery in Nigeria banking halls is a phenomenon that bother both the management of banking institutions and the customers alike. Thus the need to optimize total operating cost by determining the optimal balance between the cost of making customers to wait for service and the cost of providing additional service. Some studies have claimed that service improvement can be achieved by increasing the number of servers, but to what extent can this be done to minimize overall total cost? This study examines the validity of this claim by comparing the performance of a single channel with multi – channel queuing models in achieving cost reduction and customer satisfaction objectives using the First Bank of Nigeria Plc., Ilorin Branch as a case study. The study reveals a preference for a multi-channel system, determines a limit to its usage and makes recommendations for improvement in service delivery.

Introduction

As the world turns to a global village characterized by intense and ever increasing competition, operation bank managers continue to experience wrenching changes which they must keep up with for survival. Bank customers have become increasingly demanding. Today they require high quality, low price and immediate service delivery and tomorrow they want additional components of value from their chosen banker. Since service delivery in banks is personal, customers are either served immediately or join a queue (waiting line) if the system is busy.

Waiting line is what we encounter everywhere we go, while shopping, checking into hotels, at hospitals and clinics etc. 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 (Yechiali etal 1995).

In situations where facilities are limited and cannot satisfy the demand made upon them, bottlenecks occur which manifest as queue but customers are not interested in waiting in queues (Kelly, 2001). 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).

Objectives of the study

The obvious cost implications of customer 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.

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 First Bank Plc 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 recommendation base on the findings from the study.

Justification for the study

The previous works on the subject matter of this study only identify the need for the application of queuing models to customer waiting problem in banks and its associated costs but fails to determine optimal capacity of server usage. This study therefore aims to fill that gap by determining the maximum number of servers that can be used in order to minimize total expected costs and achieve optimal customer satisfaction.

Literature Review

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 31.8% efficient and idle 68.2% of the time. However, Ashley (2000) 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 – 120 minutes. Baale (2002) 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 centers (average of 64.32 minutes) compared to (average of 22.2 minutes) in rural areas. To buttress these observations, Juwah (1986) found out that customers spent between 55.27 to 64.56 minutes making withdrawal 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 awaiting a service function. It is an everyday occurrences and results when the number of calling units exceed the number of available service centers.

Queues are integral parts of any service system which refers to the whole situation from arrival of inputs or times to their departure.

According to Ashley (2000), the variants of queuing models that can be applied to waiting problems include: a simple system, multi-channel system, constant service and limited population models.

A Simple System (MMI) is a single line and single sever system which consist of items forming a single queue which is served by a single facility while a Multi-channel System is a system where two or more servers or channels are available to handle arriving customers. Here a common line is formed and the customers at the head of the line proceeds to the first free server.

There is also a Constant Service Time Model where the customers service time is constant instead of exponential distributed times and lastly the Limited Population Model where there 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 determines 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:

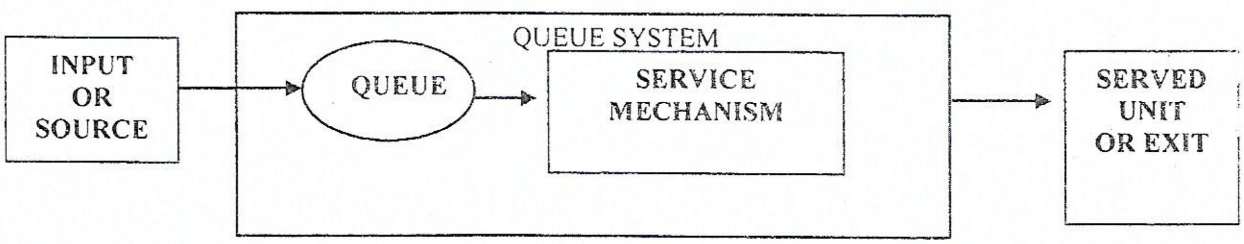
1. **Preemptive Priority:** This is common in emergency situations which allows customers that arrive at any time to replace customers that are being served e.g in patient treatment in hospitals.
2. **Non-Preemptive Priority:** Here items in the queue are arranged so that the item with the highest priority in the system is served first and there is no displacement of items in service. The methods include:

i. **FIFO (First in – First out):** Allows the first item to enter the system (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 on the queue or that which enters the system last is served first.

Queue structure

This is the nature of the queuing system in terms of input, queue and service mechanism. This is illustrated below.



SOURCE: Jay & Barry (1993) pg. 188

Features of a simple queue

Safe (2002) listed the characteristics of a single server system to include:

(a) A single service channel and variables arrivals following a Poisson distribution while service time is exponentially distributed.

(b) (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 Associates (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 serves are assumed to perform at the same rate.

Methodology of the study

Nigerian banking industry most bedevilled with customer waiting problems is studied here for a period of four weeks in First Bank of Nigeria Plc. through observation, interview and questionnaire administration. The variables measured include arrival rate (λ) and service rate (μ). They are analysed for simultaneous efficiency in customer satisfaction and cost minimization through the use of a single channel and multi-channel queuing models which are 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 / teller's service rate were used and were obtained by observation while customer attitude survey was carried out through a total of fifty questionnaires per week administered on two hundred customerss randomly selected for such purpose. These activities were carried out for 4 weeks (December 2002) that hosted four principal festivities: Muslim Sallah, Xmas, and end of a year and beginning of another year). This month is therefore considered the busiest in the last eight years and thereby considered the most appropriate for this study.

The variables are analyzed using the queuing model presented in table 3.1.

Note that, (λ) = arrival rate and (μ) = service rate

Since there are (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 8 hours.

TABLE 3.1: Formula for computing performance measures of queuing system

PERFORMANCE MEASURE		SINGLE SERVER	MULTI-SERVER (2 SERVER)	MULTI SERVER (3 SERVER)	MULTI SERVER (4 SERVER
1.	Probability that there are no customer in the system i.e the servers are idle (PO)	$1 - \lambda/\mu$	$\frac{1}{\Sigma^{m-1}(\lambda/\mu)+(\lambda/\mu)^m \frac{mu}{mu-\lambda}}$	$\frac{1}{\Sigma_{n=0}^{m-1} \frac{\lambda/\mu^n}{n!}+\frac{(\lambda/\mu)^m}{m!}\frac{mu}{mu-\lambda}}$	$\frac{1}{\Sigma_{n=0}^{m-1} \frac{\lambda/\mu^n}{n!}+\frac{(\lambda/\mu)^m}{m!}\frac{mu}{mu-\lambda}}$
2.	Average number of customer in the system (LS)	$\frac{\lambda}{\lambda - \mu}$	$\frac{\lambda/\mu\ (\lambda/\mu)^2(PO)}{m-1)!(m\mu-\lambda)^2}+\frac{\lambda}{\mu}$	$\lambda/\mu\ \frac{(\lambda/\mu)m(PO)}{m\mu-\lambda)}+\frac{\lambda}{\mu}$	$\frac{\lambda/\mu\ (\lambda/\mu)m(PO)}{m-1)!(m\mu-\lambda)}+\frac{\lambda}{\mu}$

3.	Average time a customer spends in the system (WS)	$1/\lambda-\mu$	LS/λ	LS/λ	LS/λ
4.	Average number of customer in the queue (LQ)	$\frac{\lambda^2/\mu}{(\mu-\lambda)}$	LS - λ/μ	LS - λ/μ	LS - λ/μ
5.	Average time a customer spends in the queue (WQ)	$\frac{\lambda}{\mu(\mu-\lambda)}$	LO/λ	Lq/λ	Lq/λ

Source: Jay $ Barry (1993) pg. 192 – 197
Po, Ls, Ws, LQ, And WQ are conventional symbols for measuring the performance they describe.

Data analysis and discussion of results

Table 4.1 shows that a 3-server system is better than a single server, 2 - server or 4-server system in terms of the performance criteria used. For instance, in terms of cost consideration, a 3-server system records the lowest cost of N18, 561.884 compared to a 2- server and 4-server systems that records N24, 016.66 and N30, 947.61 respectively.

The average time a customer spends in the system and in the queue are 11minutes and 1.79 minutes respectively for a 3-server system. These are however lower than that of a 2- server system (53 minutes and 46 minutes). This is further worse than that of a 4-server system that records 11.74 minutes and negligible time spent on the system and in the queue. In terms of average number of customers in the system a 2 server system has 14.5, while a 3 server and a 4 server systems have 3 and 3.2 customers respectively. The probabilities of idleness are 3%, 5.2% and 3.6% respectively for a 2, 3 and 4-servers systems respectively. The average time a customer spends in the queue and in the system for a single server, system are –0.246 minutes and –7.86 minutes respectively while the system has –1.4 and –2.14 customers in the queue and in the system respectively. The system is likely to be idle for –0.875 minutes.

Table 4.1: PRESENTATION OF RESULTS

	PERFORMANCE MEASURE	SINGLE SERVER SYSTEM	2 SERVER SYSTEM	3 SERVER SYSTEM	4 SERVER SYSTEM
1.	Probability of the system being idle	-0.875	3%	5.2%	3.6%
2.	Average number of customers in the system	-2.14	14.5	3	3.2
3.	Average time a customer spends in the system	-7.86	53min.	11min.	11.74min.
4.	Average number of	-1.4	12.6	0.49	-0.00.5

	customer in the queue				
5.	Average time a customer spends in the queue	-0.246 min	46 min.	1.79 min.	
6.	Total Economic Cost	-	₦24,016.66	₦ 18,561.884	₦ 30,947.61

Source: Author's Calculation .

The inappropriateness of a single channel model for solving customer-waiting problems becomes apparent as it shows negative figures for all performance criteria. However, multi-server models were compared and it was seen that:

a. Using a three- server system is better than a two- server system in all ramifications. For instance, a three- server system has 0.49 customer waiting in the queue for service while a two- server system has 12.6 customers on the queue. In a two-server system, a customer spends 53 minutes and 46 minutes in the system and on the queue respectively as against 11 minutes and 1.79 minutes respectively for a three-server system.

b. A three-server system has a high probability of being freer (idle) 5.2% than either of two servers or four-server system

Summarily, First Bank Nigeria 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 9 which sometimes increases to 32. i.e. 19.6 customers in excess of 12.6 customers expected. Total number of arrivals was 3358 per week or 16.53 customers per hour. This mean inter-arrival time was 3.67 minutes per customer. The arrival process follows a Poisson distribution while queue discipline was first-come first served. Queue population was unlimited and its behavior includes jockeying, reneging, balking and collusion. This study reveals that on the average, 567 customers left before service in a month at a rate of 28.35 per day, but adoption of a three-server and four server systems will reduce this rate to 11.35 and 5 customers respectively.

The bank adopted a two - server-paying system and service rate was 8.72 customer per server per hour while average service time (1/m) was 6.9 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 additional server is capable of reducing loss of customer by 60% (obtained from interview of staffs).

It was also noticed that friends of cashiers receive quick service at the expense of others and teller points are opened and closed at will leading to customers confusion and complain 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 are unhappy due to delay in service delivery while the bank is reluctant in increasing the number of service units because the theoretical underpinnings of queuing model is not well understood. Otherwise, the service units would have been increased to three to achieve better results at a lower cost of N18,561.884 as against two service units at a cost of N24,016.66.

A single server is not effective when arrival rate exceeds service rate, while the current two-server system being used is not optimal. The use of four server system although eliminates waiting but at a higher cost which is not optimal too.

Recommendations

Based on the summary and conclusions of this study, the following recommendations are suggested for efficiency improvement and quality of service to customers in First Bank of Nigeria Plc and in the service industry generally:

1. Adoption of three server model to reduce total expected costs;
2. The management should educate their operation managers and other staffs on the application of queuing model to operational problems.
3. It 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;
4. 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;
5. Reengineering the banking operations through IT solutions e.g voice mail and on-line withdrawal system to complement queuing model.
6. Making customers comfortable and unaware of the waiting time by providing electronic notice boards or TV in the waiting room as well as comfortable seats, cooling and toilet facilities and;
7. Improvement of staff-customers relationship and provision of unlabelled paper bags for carrying customers money.

References

Ashley, D. W. (2000) "Introduction to waiting line models" available at Google htt/www.com

Bale, A. F. (2002) "**Managerial Applications of queuing models in the service industry:** A case study of Afribank Nigeria Plc., Ilorin Main Branch", A Research paper submitted to Business Administration Department, Unilorin.

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

Elegalam (1978) "Customer retention versus cost reduction technique" a paper presented at the Bankers forum held at Lagos. Pg. 9-10

Jay, H. 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,** Pg. 14-17.

Michael, K. (2001) "Service on call" in **Banking Technology News Service Network** http:/www.fineos.com/documentation / features / bank teach.com

Nelson, R. T. (1959) "An empirical study of arrival, service time and waiting time distribution of job shop production process", **Report No. 60 Management Research project,** University of Califonia, L. A.

Obayemi, F. E. (1992) A queuing model analysis of a retail organization -a case study of Mr.Biggs, Lagos

Oladapo (1988) "Relationship between arrival rates of customers and bank's service rates", seminar paper series, university of Ife.

Phillip, K. (1999) **Marketing Management,** The New Millenium Edition

Robert, B. T. (1997) "Customer service for the new millennium" **Career Press 800:239-6681 www. innovastioresource.com.**

Safe Associate Ltd. (2002), **Quantitative methods, Accounting Technician** Series pp. 68

Echiali, U. (1992) "elevator polling with globally gated Regime". Google (**ttp:/w.w.w.com.**

Echiali U, Altmand, E., and Kofma, D., (1995) "Discrete Time queues with delayed Information' Vol. 19:361 – 376.