QUEUING MANAGEMENT: IMPLICATION FOR CUSTOMERS SATISFACTION

By

Dr. Mustapha Adeniyi Mudashiru

ABSTRACT

Queuing or waiting line is one of the problems frequently faced on daily basis. It manifests in many forms in the hospitals. While efforts are being made to address other management problems, queuing as an aspect of management problem has not been given the desired attention it deserved perhaps due to lack of proper understanding of what queues means, to incompetent personnel, in adequate infrastructure/facilities, faulty design of appointment schedule and resources constraints. The consequences of such neglect range from frustration, to irritation, loss of goodwill, restlessness and unpleasantness. The general objective of this study therefore is to measure customer's satisfaction by computing the operating characteristics of the waiting line. The specific objectives are to; determine how much time patients spent in the hospital, determine how busy the hospital is, and determine how patients waiting time can be reduced. The analytical tools used in this study are M/M/K queueing model. Data for this study were collected through direct observation of patient's arrival and departure that lasted two weeks. Inter-arrival time of 10 minutes period was used. Finding revealed that on the average, each patient spent 61.87 seconds in the system and 41.67 seconds in the queue 75% of the time the doctor was busy; The patients waiting time can be reduced by about 73.2% and 78.1% in both the queue and in the system respectively. The paper recommends that the present time patients spent both in the queue and in the system should be improved upon, increase the number of doctors by 1. This decision will bring some reliefs to the serving doctors; and the patient waiting time will be further reduced.

INTRODUCTION

Organization, whether they are commercial companies or non-profit making institutions have to meet customers' demands if they are to succeed. As customers need alter, so the business must change by anticipating consumer wants and where appropriate responding to them.

Taylor and Baker (1994) posited that customer satisfaction is widely recognized as the key influence in the formation of customers future purchase intention. Speaking in the same vein, File and Prince (1992) expressed that satisfied customers are likely to tell others about their favourable experiences and thus engaged in positive word of mouth advertising. Dissatisfied customers on the other hand, are likely to switch brands and engaged 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, and Rentz, 1996).

According to Fournier and Mick (1999), customer satisfaction is important, theoretical as well as practical issues for most marketers and consumer researchers. It is of interest to note that the presence of a queue can be good or bad recommendation for public facilities. At times, it is commonly understood that regular

1 c/o Department of Business Administration, University of Ilorin, Ilorin, Nigeria
queues before a private clinic or hospital are signs that patients are served on time, the workers are of good attitude and the medical attention being given is good, thus, to find satisfaction, look for a large queue.

Queues or waiting lines are observed in everyday life in order to bring orderliness into the society in which we live. We observe queue problem in so many places like in the bank, bus stop, supermarkets, hospitals, restaurants, post office, filling stations, airport e.t.c. speaking in a similar vein, Aluko (1987) observed that one of the most frequently encountered problems is waiting line and are with us in all aspects of our existence. Queuing problems are of considerable interest in this world, as a result of its perturbing consistency in both private and public service system. If Queuing is inevitable, the question now is how do we manage it for customer's satisfaction? If many patients arrived the hospital at a given time of the day, then queues will be formed. If the service rate ($\mu$) is less than the arrival rate, the queues elongates, and that if there are $\lambda$ arrivals per hour, the system will be busy $\lambda/\mu$ proportion of time and $w\mu - \lambda$ proportion of time (Wild, 1980).

The main goal of queuing management is to maximize the level of customer 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 customers' 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, conversely, a dissatisfied 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 profits and revenues (Bunday, 1996, Heinete 1994).

To measure customers satisfaction with respect to queue. Various statistics can be used. Gorden and Pressman (1978) classified the statistics into two categories; users oriented and system oriented. A user oriented statistics measures performance related to "what the user experienced." The experience of each user with respect to certain amount of waiting time will not be the same, due to the randomness of the process. A probability distribution (means and standard deviation) can be used to describe the variation in user's waiting times. The following are the major types of user oriented statistics that can be used.

(a) Waiting time- Time spent waiting in the queue for the service
(b) Time spent in the system
(c) If there is balking, reneging or a limited queue, the likelihood (probability) of a user leaving without being served.

From the user's point of view, a system that provides on the average smaller waiting time, less time in the system, or smaller probability of leaving without being served is providing better service or better satisfaction. For system oriented statistics, the characteristics of the system as a whole are described. That is the number of users in the queue. These types of statistics have unique values at each point in time and these values changes with changes in time. In view of Adan and Resing (2001), relevant performance measures in the analysis of queuing models are; the distribution of the waiting time and the sojourn time of a customer. The sojourn time is the waiting time plus the service time. The distribution of the number of customers in the system (including or excluding the one or those in service).

The general objective of this paper therefore is to measure customers' satisfaction by computing the operating characteristics of the waiting line.

Specifically, the objectives of this paper are;
(i) to determine how much time patient spent in waiting line
(ii) to determine how busy is the hospital
(iii) to suggest ways of reducing the patient waiting time

**CONCEPTUAL FRAMEWORK**
Queues are formed when units receiving some types of service cannot be served immediately (Trueman 1977). James and Benson (1988) expressed that the formation of waiting lines (queues) is a phenomenon that occurs whenever the demand for a service exceeds its supply. In the view of Aminu (2008), when limited service facility fail to satisfy the demands for service that are made upon them, bottleneck occurs which generate queues or waiting lines. If long queues develop, it may be an indication that not enough service is being provided. If no queues develop, it may be an indication that too much service is being provided. Either situation can prove costly to the service provider. It is important to mention that queues does not necessary have to be physical in nature, as the queues in front of a doctor, but it may consist of geographically separated unit awaiting services, such as equipment operated at different locations that requires some service.

Oluranti et al (2006) described queuing system as "birth death" system in which the birth occurs when a customer arrives to be served and "death" occurs when the customers depart from service unit. Queuing theory is therefore concerned with the problems such as congestion, delays, waiting lines, design of appointment scheduling e.t.c. it is obvious that the implications of queuing or delaying are very high. Problematic queuing system that are long lines can lead to the customer's perception of excessive, unfair, unexplained waiting time resulting in significant detrimental effects on the customer's overall satisfaction with the service transaction (Wayne and DiSoto 1994).

Queuing is a branch of management science that enables the analyst to describe the behaviour of queuing system (Cook and Russel 1981). According to them, occasion for applying queuing theory are numerous and varied, David and Byer (1986) posited that unlike a single model such as linear programming, queuing theory encompasses a very large group of models with each relating to a different type of queuing solution. They described the queue by computing the operating characteristics of waiting time. Operating characteristics such as the average number of unit waiting to be served and the average time a unit waits to be served. To compute the operating characteristics, the user must specify certain parameter of the queuing system, such as how the units arrived to be served and how the actual service is handled.

**The Element of a System**
Vohra (2007) puts the elements of system as follows; Arrival processes: the arrival from the input population may be classified on different bases as follows:
(a) According to source. The source of customers for a queuing system can be infinite or finite
(b) According to numbers. The customers may arrive for service individually or groups.
(c) According to time. Customers may arrive in the system at known (regular or otherwise) times, or they might arrive in a random way.

**Service System:** There are two aspects of service system, (a) structure of the service system, and, (b) the speed of the service.

**Structure of Service System:** By structure of the service system, we mean how the service facilities exist. There may be
(i) A single service facility. The models that involve a single service models. Single server, single queue model
(ii) Multiple parallel facilities with single queue
(iii) Multiple parallel facilities with multiple queue
(iv) Service facilities in a service
Speed of Service

The speed with which service is provided can be expressed in either of the ways - as service rate and service time.

The service rate describes the number of customers serviced during a particular time period. The service time indicates the amount of time needed to service a customer.

Queue Structure

The important thing to know here is the queue discipline which means the order by which patients are picked up from the waiting line for service. There are a number of possibilities. They are:

(a.) First-come-first-served (FIFO)
(b.) Last-come-first-served (UFO)
(c.) Serve-in random-order (SIRO)
(d.) Priority Service

Source: Vohra (2007)

The M/M/K queuing model being considered in this study assumed a steady state.

Condition for Steady State

According to Cooper (1972), a steady state is a service system characterized by the following:

i. Constant number of servers, average arrival and service rates over a long period of time
ii. The average arrival rate is less than the average service rate times the number of servers
iii. That (i.) and (ii.) hold for a reasonable long period of time.

Transient State

Transient state exists when the behaviour of the system depends on time.

Justification for the study

Many patients complained about the indefinite time they spent waiting in line to obtain cards and to see the doctors. Some patients have resolved to self medication and self treatment because of the stress, frustration and irritation which they have to go through in the hospitals. Relations of some patients have had cause to threaten or to sue some hospitals on the ground that they were responsible for the death of their patients due to delay in attending to them. In order to save the patients from indefinite waiting to see the doctors, and also to save the hospital from possible litigation that may arise from either the patients or their relations, this study is not only necessary but desirable.

METHODOLOGY

The data for this paper were obtained by direct observation, while patients' experiences about the time taken to receive treatment and the quality of the treatment given by the doctor were carried out through structured de-factor interview. The number of patients likely to attend the hospital at a time (i.e. those with cards and those who were about to obtain cards) were used as the sampling frame. The number of patients interviewed numbered 244 represented the sampling size. Purposeful sampling technique of first-come-first-sighted was employed. Inter-arrival time of 10 minutes period was used. The study was carried out for 6 hours each day from 9.00 am to 1.00 pm.

The model for this study is queuing model using multiple servers. Two servers were involved. The servers were the clerk in charge of issuance and retrieval of cards, and the doctors. In this case, there were more
than one service points \((k)\), and there was one waiting line with equal service rates. Patients arrived according to a poisson process at mean rate \(A\). The service time follows an exponential distribution and each of the \(K\) servers worked at an average rate \(\mu\) (with \(k\mu > \lambda\)). The performance measured formula for this queuing system is shown below.

(i.) The probability of having no patient in the system

\[
P_0 = \frac{1}{\sum_{n=0}^{K-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{K!} \left(\frac{\lambda}{K\mu - \lambda}\right)^K} \]

\(\text{...equation (i)}\)

(ii.) Average number of patient in the system

\[
L = \frac{(\lambda/\mu)^k \lambda \mu}{(k-1)! (K\mu - \lambda)^2} \frac{P_0 + \lambda}{\mu} \]

\(\text{...equation (ii)}\)

(iii.) Average queue length i.e. average numbers of patient in the queue

\[
L_q = \frac{(\lambda/\mu)^k \lambda \mu}{(k-1)! (K\mu - \lambda)^2} P_0 \]

\(\text{...equation (iii)}\)

(iv.) Average time in the system

\[
W = \frac{(\lambda/\mu)^k \lambda P_0 + 1}{(k-1)! (K\mu - \lambda)^2} \frac{\mu}{\mu} \]

\(\text{...equation (iv)}\)

(v.) Average time in the queue

\[
W_q = \frac{(\lambda/\mu)^k \mu}{(k-1)! (K\mu - \lambda)^2} P_0 \]

\(\text{...equation (v)}\)

(vi.) Utilization factor

\[
P = \frac{\lambda}{K\mu} \]

\(\text{...equation (vi)}\)

(vii.) Traffic intensity

\[
P = \frac{\lambda}{\mu} \]

\(\text{...equation (vii)}\)

**RESEARCH QUESTIONS**

Queuing requirement of a private clinic hospital were based on the following research questions:

- How do patient arrive in the clinic?
- How much time do patients spend in the clinic?
- Does the patient service time vary with the type of patient?
- How many benches does the clinic have for serving patient?

These questions among others formed the cornerstone of study. They determine the sort of data collected and inferring techniques used (Ijaiya 1997).

The data collected for fourteen (14) days were summarized in the table below:
### TABLE 1: SUMMARY OF DATA COLLECTED AS ARRIVAL AND SERVICE TIME

<table>
<thead>
<tr>
<th>S/N</th>
<th>TIME</th>
<th>TOTAL ARRIVAL</th>
<th>TOTAL NO OF SERVER</th>
<th>TOTAL SERVICE TIME (MINS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>9.00am – 9.10am</td>
<td>26</td>
<td>14</td>
<td>202</td>
</tr>
<tr>
<td>2.</td>
<td>9.11am - 9.20am</td>
<td>23</td>
<td>12</td>
<td>135</td>
</tr>
<tr>
<td>3.</td>
<td>9.21am - 9.30am</td>
<td>20</td>
<td>7</td>
<td>148</td>
</tr>
<tr>
<td>4.</td>
<td>9.31am - 9.40am</td>
<td>16</td>
<td>9</td>
<td>217</td>
</tr>
<tr>
<td>5.</td>
<td>10.41am - 9.50am</td>
<td>24</td>
<td>13</td>
<td>201</td>
</tr>
<tr>
<td>6.</td>
<td>10.51am - 10.00am</td>
<td>22</td>
<td>8</td>
<td>212</td>
</tr>
<tr>
<td>7.</td>
<td>10.01am - 10.10am</td>
<td>17</td>
<td>7</td>
<td>181</td>
</tr>
<tr>
<td>8.</td>
<td>10.11 am - 10.20am</td>
<td>20</td>
<td>11</td>
<td>224</td>
</tr>
<tr>
<td>9.</td>
<td>10.21am – 10.30am</td>
<td>24</td>
<td>6</td>
<td>236</td>
</tr>
<tr>
<td>10.</td>
<td>10.31 am - 10.40am</td>
<td>29</td>
<td>10</td>
<td>218</td>
</tr>
<tr>
<td>11.</td>
<td>10.41am - 10.50am</td>
<td>31</td>
<td>18</td>
<td>206</td>
</tr>
<tr>
<td>12.</td>
<td>10.51am - 11.00am</td>
<td>27</td>
<td>20</td>
<td>191</td>
</tr>
<tr>
<td>13.</td>
<td>11.01am - 11.10am</td>
<td>16</td>
<td>15</td>
<td>261</td>
</tr>
<tr>
<td>14.</td>
<td>11.11am -11.20am</td>
<td>23</td>
<td>9</td>
<td>224</td>
</tr>
<tr>
<td>15.</td>
<td>11.11am - 11.30am</td>
<td>20</td>
<td>8</td>
<td>186</td>
</tr>
<tr>
<td>16.</td>
<td>11.31am - 11.40am</td>
<td>31</td>
<td>13</td>
<td>225</td>
</tr>
<tr>
<td>17.</td>
<td>11.41am - 11.50am</td>
<td>36</td>
<td>10</td>
<td>182</td>
</tr>
<tr>
<td>18.</td>
<td>11.51am - 12.00pm</td>
<td>32</td>
<td>14</td>
<td>193</td>
</tr>
<tr>
<td>19.</td>
<td>12.01pm -12.10pm</td>
<td>19</td>
<td>6</td>
<td>227</td>
</tr>
<tr>
<td>20.</td>
<td>12.11pm - 12.20pm</td>
<td>24</td>
<td>8</td>
<td>212</td>
</tr>
<tr>
<td>21.</td>
<td>12.21pm - 12.30pm</td>
<td>35</td>
<td>11</td>
<td>178</td>
</tr>
<tr>
<td>22.</td>
<td>12.31pm - 12.40pm</td>
<td>21</td>
<td>4</td>
<td>163</td>
</tr>
<tr>
<td>23.</td>
<td>12.41pm - 12.50pm</td>
<td>28</td>
<td>15</td>
<td>176</td>
</tr>
<tr>
<td>24.</td>
<td>12.51pm - 1.00pm</td>
<td>27</td>
<td>6</td>
<td>204</td>
</tr>
<tr>
<td>25.</td>
<td>TOTAL</td>
<td>591</td>
<td>244</td>
<td>4802</td>
</tr>
</tbody>
</table>

Source: Researcher's observation 2009.

Mean arrival rate \((\lambda)\) = \(\frac{\text{Total number of arrival}}{\text{Total number of hours}}\)  
\(= \frac{591}{7} = 7.04 \approx 7\) patients

Mean service rate \((\lambda)\) = \(\frac{\text{Total number of service}}{\text{Total number of time}}\)  
\(= \frac{244}{4802} = 0.05\) = 3 patients

The number of service points \(k = 2\)
TABLE 2: OPERATING CHARACTERISTICS USING M/M/K MODEL

<table>
<thead>
<tr>
<th>OPERATING CHARACTERISTICS</th>
<th>PRESENT</th>
<th>ALTERNATIVE</th>
<th>% CHANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>2</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>7</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>( \mu )</td>
<td>3</td>
<td>5</td>
<td>66.6</td>
</tr>
<tr>
<td>( P )</td>
<td>2.333</td>
<td>1.4</td>
<td>(40)</td>
</tr>
<tr>
<td>( P_u )</td>
<td>0.949</td>
<td>0.466</td>
<td>(50.8)</td>
</tr>
<tr>
<td>( L_q )</td>
<td>3.957</td>
<td>0.173</td>
<td>(96)</td>
</tr>
<tr>
<td>( L_s )</td>
<td>5.860</td>
<td>1.572</td>
<td>(73.2)</td>
</tr>
<tr>
<td>( W_q )</td>
<td>41.67 seconds</td>
<td>1.48 seconds</td>
<td>(96.4)</td>
</tr>
<tr>
<td>( W_s )</td>
<td>61.67 seconds</td>
<td>13.48 seconds</td>
<td>(78.1)</td>
</tr>
</tbody>
</table>

Source: Researcher's Computation M/M/K Queuing Output (2009)

From the table 2 above, "with 2 servers, patients spent 41.67 seconds in the queue and 61.67 in the system. The utilization factor \( P_u \) which represents the rate at which the server were being used was 0.949 representing about 95%. This shows that 95% of the time, the clerk and the doctor were busy.

Discussion of Findings
The result of findings revealed that, while some patients came in by themselves, those with serious health problem were accompanied by either their relatives or relations. Continuous inflow of friends, relations and well-wishers were also noticed. By increasing the number of servers from 2 to \( J \), the utilization factor \( P_u \) will be 0.466 representing 50.8% reduction. The effect of this is a reduction in the patient waiting period in the queue \( W_q \) by 73.2% and 78.1% in the system respectively.

The analysis indicates that the mean arrival rate \( \lambda \), is greater than the mean service rate 1-1. This was responsible for queue. On the average, 7 patients arrived per hour, while 3 patients were served. The result of the analysis puts the traffic intensity at 2.33. Since this is more than 1, it means the flow of patients to the clinic was high. Again, 95% of the time, both the clerk and the doctor were busy and have less time to rest. This Queue discipline of first-come-first-served was observed though there were one or few emergency situations where priority rule was applied. At times there were some pre-arranged schedules.

The finding further revealed that by engaging one additional Doctor, the patient waiting time can be reduced by 73 % in the queue and 78 % in the system.

CONCLUSION
Queue or waiting line is acknowledged by all as a common phenomenon. It is been considered as an efficacious decision support system. It offers a measurable framework, driven by planning and control. When patients wait in queue, there is the danger that waiting time will become excessive leading to the loss of some patients to other hospitals. This may not be too good enough. Allowing them to serve themselves
so easily is a key factor in both keeping and attracting patients. Queuing theory encompasses a very large group of models with each relating to a different type of queuing situation. It is very important to note that these models do not attend to "solve" queuing problem; rather they describe the queue by computing the operating characteristics of waiting line. Description rather than optimization is the objective, of queuing models, and any optimization that takes place must be done by the users varying the system parameters to obtain different sets of operating characteristics. The sets of operating characteristics that closely matches the user needs therefore define the "best" system structure.

One can conclude that queue is inevitable. What is now left for management of any service organization to aim at all times is how to manage it for effective customers' satisfaction.

RECOMMENDATION
In the light of the above discussion and conclusion, the following recommendations are suggested for effective queuing management in Delink Hospital, Oja-Iya, Ilorin.
Sustain the present service rate or improve upon it through a regime of appropriate incentives. Increase the number of Doctors by 1. By this action, the patient waiting time can be further reduced by 73 % in the queue and 78.1 % in the system.
REFERENCES


