A Study of Queuing Model for Banking System

Toshiba Sheikh, Sanjay Kumar Singh and Anil Kumar Kashyap

Chhatrapati Shivaji Institute of Technology
Durg, Chhattisgarh, 491001, India
tsheikh1990@gmail.com1; sanjaysingh@csitdurg.in2; anilkashyap@csitdurg.in3

Abstract
Waiting lines and service efficiency are the important elements for any bank. Queuing theory has been fairly a successful tool in the performance analysis of waiting lines. In this paper, an optimized model is proposed to improve the bank queuing system based on queuing theory. This method can optimize the number of server and improve the service efficiency that could effectively cut down service costs and customer’s waiting time.

Keywords— Queuing system, Multi Server Queuing Model, Service efficiency.

Introduction
Waiting lines or queues is the major source of difficulties to any organisation or service provider institutes like hospitals, banks etc. Queuing theory basically a mathematical approach, used for the analysis of waiting lines. It deals with the problems that involve waiting (or queuing). Queuing theory is used to analysing the congestions and delays of waiting in line. It is used to develop more efficient queuing system that reduce customer waiting time and increase the number of customer that can be served. Customers waiting time depends on the number of customers on queue, the number of servers serving line, and the amount of service time for each individual customers. The time wasted on the queue would have been wisely utilized elsewhere (opportunity cost of time spent in queuing). In a waiting line system, when considering improvements in services to offer an optimal service level, consider the cost of providing a given level of service against the potential costs from having customers waiting. Customer satisfaction is improved by predicting and reducing
waiting time and adjusting staff. Proposes an incremental analysis approach in which the cost of additional servers is compared with the benefits it generates. Servers are added until the increase cost equal the benefits.

**Optimization Method**

Here the queuing system is the Bank Service System. The following assumptions were made for the queuing system at the Punjab National Bank Bilaspur, in accordance with the queuing theory:

- Poisson arrival rate of $\lambda$ customers per unit of time.
- Exponential service times of $\mu$ customer per unit of time.
- Queue discipline is first come first served basis by any of the server.
- The waiting line has two or more identical servers.
- There is no limit to the number of the queue (infinite).
- The average arrival rate is greater than average service rate.

Following notation and terminology are used in the formulation and evaluation of the queuing model:

- $P_n$ = probability of exactly $n$ customers in the system.
- $L_s$ = expected number of customers in the system.
- $L_q$ = expected number of customers in the queue.
- $W_s$ = average time a customer spends in the system.
- $W_q$ = expected waiting time of customers in the queue.
- $Y$ = number of servers.
- $C_s$ = service cost of each server.
- $C_w$ = opportunity cost of waiting by customers.
- $\lambda$ = the mean arrival rate (expected number of arrivals per unit time) of new customers are in systems.
- $\mu$ = the mean service rate for overall systems (expected number of customers completing service per unit time) when $n$ customers are in systems.

**Model Formulation**

All The model adoption in this work is the $(M/M/Y) : (\infty/FCFS)$ Multi Server Queuing Model. This is the extend form single server model where customer in a waiting line can be served by more than one server simultaneously. There are $n$ numbers of customers in the queuing system at any point in time.

If $n < Y$, (number of customers in the system is less than the number of servers), then there will be no queue. However, $Y - n$ number of servers will not be busy. The combined service rate will then be $\mu_n = n\mu$ ; $n < Y$. And, if $n \geq Y$ (number of customers in the system is more than or equal to the number of servers) then all servers will be busy and the maximum number of customers in the queue will be $n - Y$. The combined service rate will be $\mu_n = Y\mu$ ; $n \geq Y$. Following are the properties of the Multi-Server Queuing Model:
The probability of having $n$ customers in the system is given by
\[
\rho = \frac{\lambda}{Y \mu}
\]
\[
P_0 = \left[ \sum_{n=0}^{Y-1} \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n + \frac{1}{Y!} \left( \frac{\lambda}{\mu} \right)^Y \frac{Y\mu}{Y\mu - \lambda} \right]^{-1}
\]
\[
P_n = \begin{cases} 
\left( \frac{\rho^n}{n!} \right) P_0 & n \leq Y \\
\left( \frac{\rho^n}{(Y+1)^{n-Y}} \right) P_0 & n > Y 
\end{cases}
\]

Expected number of the customers waiting on the queue.
\[
L_q = \left[ \frac{1}{(Y-1)!} \left( \frac{\lambda}{\mu} \right)^Y \frac{\mu\lambda}{(Y\mu - \lambda)^2} \right] P_0
\]

Expected number of customers in the system.
\[
L_s = L_q + \frac{\lambda}{\mu}
\]

Expected waiting time of customers in the queue.
\[
W_q = \frac{L_q}{\lambda}
\]

Average time a customer spends in the system.
\[
W_s = \frac{L_s}{\lambda}
\]

Utilization factor i.e. the fraction of time servers are busy.
\[
\rho = \frac{\lambda}{\mu Y}
\]

**Cost Model**

In order to cost optimization and determination of optimal number of servers with the total optimal cost of the system, two opposing costs must be considered. These are: service cost and waiting time cost.

- Expected Service Cost = $YC_s$
- and
- Expected Waiting Time Cost in the system = $(\lambda W_s)C_w$

By summation of the service cost and waiting time cost in the system we get, Expected Total Cost = $YC_s + (\lambda W_s)C_w$
Model Formulation
To evaluate the characteristics of the Multi Server Queuing System at the Punjab National Bank using TORA software. Taking, $\lambda = 75$ customers/hr., $\mu = 50$ customers/hr. on the basis of these data the evaluation is shown in the Table 1.

Table 1. Model Evaluation of Multi-Server Queuing Model at the Punjab National Bank Bilaspur

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\lambda$</th>
<th>$\mu$</th>
<th>System Utilization</th>
<th>$P_o$</th>
<th>$L_s$</th>
<th>$L_q$</th>
<th>$W_s$ (in hours)</th>
<th>$W_q$ (in hours)</th>
<th>Total Cost/hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>75</td>
<td>50</td>
<td>0.50</td>
<td>0.2105</td>
<td>1.7368</td>
<td>0.0236</td>
<td>0.0031</td>
<td>1074.402</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>75</td>
<td>50</td>
<td>0.37</td>
<td>0.2209</td>
<td>1.5448</td>
<td>0.0486</td>
<td>0.0006</td>
<td>471.752</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>75</td>
<td>50</td>
<td>0.30</td>
<td>0.2227</td>
<td>1.5086</td>
<td>0.0086</td>
<td>0.0001</td>
<td>442.579</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>75</td>
<td>50</td>
<td>0.25</td>
<td>0.2230</td>
<td>1.5016</td>
<td>0.0016</td>
<td>0.0000</td>
<td>504.780</td>
</tr>
</tbody>
</table>

Fig.1: Snapshot of TORA Software Calculation

Fig.2: Expected Service Cost Vs Level of Service
A Study of Queuing Model for Banking System

**Fig. 3:** Expected Waiting Cost Vs Level of Service

**Fig. 4:** Expected Waiting Cost Vs Level of Service

**Fig. 5.** Expected number of customers in the system ($L_s$) Vs Probability of the system being idle ($P_0$)
Conclusion
By the analysis it is concluded that as the service level increases at an optimal service level, the waiting time of the customer is reduced, the service efficiency is increased and the customer satisfaction is increased and as well as it is found that at some specific numbers of service windows the total system cost is minimized. By the evaluation it is proved that this optimal model of the queuing is feasible.

References