Analysis of the Sales Checkout Operation in Supermarket Using Queuing Theory

JSKC Priyangika[1#] and TMJA Cooray[2]
[1]*Department of Physical Sciences, Faculty of Applied Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka*
[2]*Department of Mathematics, Faculty of Engineering, University of Moratuwa, Moratuwa, Sri Lanka*
[#]*cham1981@gmail.com*

***Abstract*—** *This paper contains the analysis of Queuing systems for the empirical data of supermarket checkout service unit using queuing theory. One of the expected gains from studying queuing systems is to review the efficiency of the models in terms of utilization and waiting length, hence increasing the number of queues so customers will not have to wait longer when servers are too busy. In other words, trying to estimate the waiting time and length of queue(s) is the aim of this study. Queuing simulation is used to obtain a sample performance result and estimated solutions for multiple queuing models are also of interest. This study requires empirical data which may include variables like, the arrival time in the queue of checkout operating unit (server), departure time, service time, etc. A questionnaire is developed to collect the data for such variables and the reaction of the Supermarket from the customers separately. This model is developed for a sales checkout operation in the supermarket. The model designed for this research is multiple queues multiple-server model. The model contains five servers which are checkout sales counters; attached to each server is a queue. In any service system, a queue forms whenever current demand exceeds the existing capacity to serve. This occurs when the checkout operation unit is too busy to serve the arriving costumers immediately.*

***Keywords*—** *Multiple-Server Model, Queuing Simulation, Service Time*

I. INTRODUCTION

This study is the review of queuing theory and for empirical study the sales checkout service unit of Alankulama supermarket is chosen as an example.Each store is owned and operated separately, but operations are coordinated within the group. All feature Alankulama brand products. There are two AlankulamaSupermarkets in Anuradhapura; the bigger one was chosen to be the research object and to collect data from.

The main purpose of this paper is to review the application of queuing theory and to evaluate the parameters involved in the service unit for the sales checkout operation in Alankulama Supermarket. Therefore, a mathematical model is developed to analyze the performance of the checking out service unit. Two important results need to be known from the data collected in the supermarket by the mathematical model: one is the 'service rate' provided to the customers during the checking out process, and the other is the gaps between the arrival times (inter-arrival time) of each customer per hour. In order to get an overall perspective of the customer's quality of service, the questionnaires which indicate the result in percentages, are also used to get the evaluation from the customers directly.

There are five counters in Alankulama supermarket at one place, which means consisting of five servers with five queues in terms of Queuing Theory. A queue forms whenever current demand exceeds the existing capacity to serve when each counter is so busy that arriving customers cannot receive immediate service facility. So each server process is done as a queuing model in this situation.

The data used in the Queuing model is collected for an arrival time of each customer in two days by the questionnaire form. The observations for number of customers in a queue, their arrival-time and departure-time were taken without distracting the employees. The whole procedure of the service unit each day was observed and recorded using a time-watch during the same time period for each day. In addition, the questionnaires are conducted at the same timings for each day.

The aim of studying queuing system simulation is trying to detect the variability in a quality of service due to queues in sales checkout operating units, find the average queue length before getting served in order to improve the quality of the services where required, and obtain a sample performance result to 2 obtain time-dependent solutions for complex queuing models. The defined model for this kind of situation where a network

of queues is formed is time-dependent and needs to run simulation. The results obtained from Alankulama Supermarket using queuing model suggest that sales checkout operating unit is rather busy each day of a week but the service is satisfactory.

I. METHODOLOGY

A. Queuing Models with Single Stage (facility)

The term queuing system is used to indicate a collection of one or more waiting lines along with a server or collection of servers that provide service to these waiting lines. The example of Alankulama supermarket is taken for queuing system discussed in this chapter include:
1) a single waiting line and multiple servers (Figure 1),
2) multiple waiting lines (arranged by priority) and multiple servers (Figure 2) , and
3) a single waiting line and a single server (Figure 3).
All results are presented in next chapter assuming that First Come First Serve basis.

The supermarkets may consist of multiple units to perform same checkout operation of sales, which are usually set all together besides the entrance of the supermarket. Each unit contains one employee. This kind of a system is called a multiple-server system with single service facility, in other words multiple checkouts counters (service units) with sales checkout as a service available in a system. There are two possible models for multiple-server system:
(i) Single-Queue Multiple-Server model, and
(ii) Multiple-QueueMultiple-Server model.

Using the same concept of model, the sales checkout operating units are all together taken as a series of servers that forms either single queue or multiple queues for sales checkout (single service facility) where the arrival rate of customers in a queuing system and service rate per busy server are constants regardless of the state of the system (busy or idle). For such a model the following assumptions are made:

1)Assumptions :

(i) Arrivals of customers follow a Poisson process.
The number of the customers that come to the Queueof sales checkout server during timeperiod $(t\ , t+\Delta t)$ only depends on the length of the time period 'Δt' but no relationship with the start time 't'.If Δt is small enough, there will be at most one customer arrives in a queue of a server during timeperiod $(t\ , t+\Delta t)$.

Therefore, the number of customers that arrive in an interval $(t\ , t+\Delta t)$ follows a Poisson distribution and the arrivals of them in a queue follows a Poisson process. A Poisson process as a sequence of events 'randomly spaced in time'.

(ii)Inter-arrival times of a Poisson process are exponentially distributed.

(iii) Service times are exponentially distributed.

(iv)Identical service facilities (same sales checkout service on each server)

(v)No customer leaves the queue without being Served

(vi)Infinite number of customers in queuing system of supermarket (i.e. no limit for queue capacity)

(vii) FIFO (First In First Out) or FCFS (First Come First Serve)

Customers arriving from different flows are treated equally by placing into the queues, respecting strictly, their arriving order. Already in the queue are served in the same order they entered, this means first customer that comes in the queue is the first one that goes out.

All customers arriving in the queuing system will be served approximately equally distributed service time and being served in an order of first come first serve, whereas customer choose a queue randomly, or choose or switch to shortest length queue. There is no limit defined for number of customers in a queue or in a system.

B. Basic Queuing Process

Customers requiring service are generated over time by an input source. The required service is then performed for the customers by the service mechanism, after which the customer leaves the queuing system. We can have following two types of models: One model will be as Single-queue Multiple-Servers model (Figure 1) and the second one is Multiple-Queues, Multiple-Servers model (Figure 2) (Sheu,C., Babbar S. (Jun 1996)).

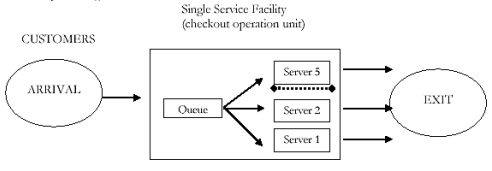

Fig. 1: Single Stage Queuing Model with Single-Queue and Multiple Parallel Servers

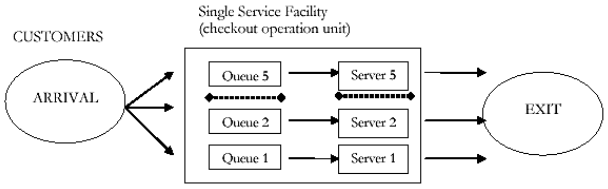

Fig. 2: Single Stage Queuing Model with Multiple Queues and Multiple Parallel Servers

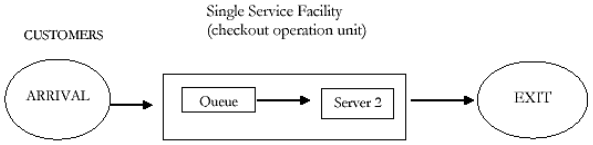

Fig. 3: Single Stage Queuing Model with Single-Queue and Single-Server

The Queuing model is commonly labeled as M/M/c/K, where first M represents Markovianexponential distribution of inter-arrival times, second M represents Markovian exponential distribution of service times, c (a positive integer) represents the number of servers, and K is the specified number of customers in a queuing system. This general model contains only limited number of K customers in the system. However, if there are unlimited number of customers exist, which means K = ∞, then our model will be labeled as M/M/c (Hillier & Lieberman, 2001.)

C. Parameters in Queuing Models (Multiple Servers, Multiple Queues Model)

n = Number of total customers in the system (in queueplus in service)

c = Number of parallel servers (Checkout sales operation units in Alankulama super market)

λ = Arrival rate (1 / (average number of customers arriving in each queue in a system in one hour))

μ = Serving rate (1 / (average number of customers beingserved at a server per hour))

c μ = Serving rate when c > 1 in a system

ρ =System intensity or load, utilization factor (= l/(cμ)) (the expected factor of time the server is busy that is, service capability being utilized on the average arriving customers)

Departure and arrival rate are state dependent and are in steady-state (equilibrium between events) condition.

D. Notations and their Description for Single Queue and Parallel Multiple Servers Model

P_n = probability that there are exactly n customers in the system in steady–state condition

$$\sum_{n=0}^{c-1} P_n + \sum_{n=c}^{\infty} P_n = 1$$

$$P_n = \begin{cases} \frac{\lambda}{n\mu} P_{n-1} = \frac{\lambda^n}{n!\mu^n} P_0 & for\ 1 \le n \le c-1 \end{cases}$$

$$P_n = \frac{1}{c^{n-c} \times c!} \left(\frac{\lambda}{\mu}\right)^n P_0 \quad for \quad n \ge c$$

L_q = average number of the customers in the queue

$$= P_c \frac{\rho}{(1-\rho)^2} \quad where \ P_c = \frac{\left(\frac{\lambda}{\mu}\right)^c P_0}{c!}$$

L_s = average number of the customers in the system = $L_q + \frac{\lambda}{\mu}$

W_q = waiting time in the queue = $\frac{L_q}{\lambda}$

W_s = waiting time in the system = $\frac{L_s}{\lambda}$

There are no predefined formulas for networks of queues,that is for multiple queues (Figure 2). A complexity of the model is the main reason for that. Therefore, we use notations and formulas for single queue with parallel servers. In order to calculate estimates for multiple queues multiple servers' model, we may run simulation.

We are next concerned about how to obtain solution for a queuing model with a network of queues?
Such questions require running Queuing Simulation. Simulation can be used for more refined analysis to represent complex systems.

E. Queuing Simulation

The queuing system is when classified as M/M/c with multiple queues where number of customers in the system and in a queue is infinite, the solution for such models are difficult to compute. When analytical computation of μ is very difficult or almost impossible, a

Monte Carlo simulation is appealed in order to get estimations. A standard Monte Carlo simulation algorithm fix a regenerative state and generate a sample of regenerative cycles, and then use this sample to construct a likelihood estimator of state(Nasroallah, 2004). Although supermarket sales do not have regenerative situation but simulation here is used to generate estimated solutions.

Simulation is the replication of a real world process or system over time. Simulation involves the generation of artificial events or processes for the system and collects the observations to draw any inference about the real system. A discrete-event simulation simulates only events that change the state of a system. Monte Carlo simulation uses the mathematical models to generate random variables for the artificial events and collect observations. (Banks, 2001)

Discrete models deal with system whose behavior changes only at given instants. A typical example occurs in waiting lines where we are interested in estimating such measures as the average waiting time or the length of the waiting line. Such measures occur only when the customer enters or leaves the system. The instants at which changes in the system occurs identify the model's events, e.g. arrival and departure of the customers. The arrival events are separated by the 'interarrival time' (the interval betweensuccessive arrivals), and the departure events are specified by the service time in the facility. The fact that these events occur at discrete points is known as "Discrete-event Simulation." (Taha, 1997)

When the interval between successive arrivals is random then randomness arises in simulations. The time t between customers' arrivals at Alankulama is represented by an exponential distribution; to generate the arrival times of the next customers from this distribution, we have

$$t = -\left(\frac{1}{\mu}\right)\ln(1-R)$$, where R = random number. (1 – R) is a compliment of R, so we can replace (1 – R) with R.

II. ANALYSIS OF CHECKOUT SALES OPERATION SERVICE IN ALANKULAMA

A sales checkout service has 5 waiting lines in a form of parallel cash counters (Figure2).Customers are served on a first-come,first-served (FCFS) basis as a salesman of checkout operation unit becomes free. The data has been collected for only two out of five servers on Wednesday (weekday) by using questionnaires. It was assumed that the customers' crowd is more, on average, on weekday. Although the sales checkout unit has 5 parallel counters out of which 2 were observed (each of them has an individual salesman to deal with the customers in a queue), it is possible that some of the checkout units are idle. The data collected from questionnaires were tabulated in a spreadsheet in order to calculate the required parameters of queuing theory analysis.

Firstly, the confidence intervals are computed to estimate service rate and arrival rate for the customers. Then the later first part of the analysis is done for the model involving one queue and 2 parallel servers (Figure 1), whereas the second part is done by queuing simulation for second model involving 2 queues for each corresponding parallel server (Figure 2).

We can estimate confidence intervals for average service rate and average arrival rate. Assuming service time and arrival time are iid with N(0,1), then the 95% confidence interval for arrival rate can be calculated as follows:

A. Confidence Intervals for weekday:

Table1. Summary of data

	Service time (min/ customer	Arrival time (min/ customer)	Service rate (customers/hour)	Arrival rate (customers/hour)
Mean	01:06	00.37		
Standard deviation	00:06	00.06		
95% confidence Interval	0.9 – 1.3	0.4 – 0.81666	46 - 67	73 - 150

And n = 41 customers

1). Interpretation of Confidence Interval

The confidence intervals show that 73 to 150 customers arrive in 2-server system within an hour whereas46 to 67 customers are served. That means there are still some customers not being served and are waiting for their turn in a queue to be served. This is due to a service time provided by a server to the customers.The service time can vary between 54 sec to 78 sec per customer.

B. Expected Queue Length

We can find the expected length of queue by using empirical data. In survey, the number of customers waiting in a queue was observed (Appendix B). The average of that number in a system is (1+1+3+…+2+0)/41 = 2.07 customers per minute on average waiting in a queue in a system within 25 min of data collection time.

C. Queuing Analysis

On Wednesday (weekday), customers arrive at an average of 98 customers per hour, and an average of 55 customers can be served per hour by a salesperson.

1). Results for Weekday applying Queuing model 1 (Figure1)

The parameters and corresponding characteristics in Queuing Model M/M/2, assuming system is in steady-state condition, are:

n = 41

c = 2

λ = 98 customers / hour

μ = 55 customers per server/ hour

c μ = Serving rate = $2 \times 55 = 110$

$\rho = 0.8909$ = Overall system utilization =89.09%

The probability that all servers are idle (Po) = 0.5769

$L_q = 6.6560$

$W_q = \frac{L_q}{\lambda}$ = 0.0700 hours

2). Interpretation of results for queuing model 1

The performance of the sales checkout service on weekday is sufficiently good. We can see that the probability for servers to be busy is 0.8909, i.e. 89.09%. The average number of customers waiting in a queue is L_q = 6.8560 customers per 2-server. The waiting time in a queue per server is W_q = 4.2 min which is normal time in a busy server. This estimate is not realistic as the model shows that the customers make a single queue and choose an available server. Hence we can consider each server with a queuing model as a single-server single-queue model to get the correct estimate of the length of queue. M/M/1 queue is a useful approximate model when service times have standard deviation approximately equal to their means.

3). Results for Weekday applying Queuing model 3 (Figure.3)

The parameters and corresponding characteristics in Queuing Model M/M/1, assuming system is in steady-state condition, are:

c = 1

λ = 98 customers per hour for 2 servers .That is 49 customers / hour

μ = 55 customers per server/ hour

ρ =0.8909= Overall system utilization =89.09%

The probability that all servers are idle (Po) = 0.1091

W_q = 0.1485 hours

4).Interpretation of results for queuing model 3 (Figure 3)

The performance of the sales checkout service remains same as for 2 servers on weekday. The number of customers in a queue is (7.2758) higher than a queue with two servers. Each customer in a queue has to wait for 8.9 minutes. This means, reducing the number of servers may lead a longer queue.

D. Queuing simulation

It is not possible to obtain solutions for multi-queue models in closed form or by solving a set of equations, but they are readily obtained with simulation methods. The simulation has been run for the same empirical data as for model 1, using software WinQSB for Queuing System Simulation .The mean inter-arrival time and mean service time as taken same for both servers.

1). Results for Weekday Applying Queuing Model 2

07-05-2007	Result	Customer1	Customer2	Overall
1	Total Number of Arrival	139	176	315
2	Total Number of Balking	1	35	36
3	Average Number in the System (L)	29.1828	40.3914	69.5742
4	Maximum Number in the System	51	51	102
5	Current Number in the System	48	51	99
6	Number Finished	90	90	180
7	Average Process Time	1.1000	1.1000	1.1000
8	Std. Dev. of Process Time	0.0012	0.0012	0.0012
9	Average Waiting Time (Wq)	21.1321	28.5634	24.8478
10	Std. Dev. of Waiting Time	11.6340	15.1996	14.0355
11	Average Transfer Time	0	0	0
12	Std. Dev. of Transfer Time	0	0	0
13	Average Flow Time (W)	22.2321	29.6634	25.9478
14	Std. Dev. of Flow Time	11.6340	15.1996	14.0355
15	Maximum Flow Time	39.5823	54.9761	54.9761
	Data Collection: 0 to	100 hours		
	CPU Seconds =	5.4490		

Figure 4. Result from simulation

2). *Interpretation of Queuing Simulation results for model2*

A simulation process has clearly shown the performance of the sales checkout service of two servers including their corresponding queues. The simulation was run for 100 hours. The servers are found to be very busy (99%). The average number of customers waiting in a queue in overall two servers on weekday is L_q = 67.5812 whereas

the waiting time in a queue in overall two servers is approximately W_q = 25.0821 min which is normal time in a very busy server. Such a longer queue can be reduced in size by a decrease in service time or server utilization. Although inter-arrival time and mean service time is same for both servers but there is a small difference in the value of L_q and W_q. This is possible when system hasmultiple queues and queues have jockey behavior. In other words, customers tend to switch to a shorterqueue to reduce the waiting time.

3). Comparison of the Results for Queuing Model 1 and Model 2

The actual structure of our survey example Alankulama super market has queuing model 2 (Figure 2). A queuing model with single queue and multiple parallel servers (Figure 1) does not clearly evaluate performance for each server. For instance, the utilization factor for both servers varies in each analysis, i.e. for model 1 its 89% whereas for model 2 its 99%. A simulation process shows the performance of each server with their corresponding queues (Figure 2). For instance, in server 2 each customer has to wait for 15.67 minutes in case of 40 customers in a queue and in server 1 each customer has to wait for 21.87 minutes in case of 31 customers waiting in a queue for being served.

IV.DISCUSSION AND CONCLUSION

This study reviews a queuing model for multiple servers. The average queue length can be estimated simply from raw data from questionnaires by using the collected number of customers waiting in a queue each minute. We can compare this average with that of queuing model. Three different models are used to estimate a queue length: a single-queue multi-server model, single-queue single-server and multiple queue multi-server model. In case of more than one queue (multiple queue), customers in any queue switch to shorter queue (jockey behavior of queue). Therefore, there are no analytical solutions available for multiple queues and hence queuing simulation is run to find the estimates for queue length and waiting time.

The empirical analysis of queuing system of Alankulama supermarket is that they may not be very efficient in terms of resources utilization. Queues form and customers wait even though servers may be idle much of the time. The fault is not in the model or underlying assumptions. It is a direct consequence of the variability of the arrival and service processes. If variability could be eliminated, system could be designed economically so that there would be little or no waiting, and hence no need for queuing models.

With the increasing number of customers coming to Alankulama for shopping, either for usual grocery or for some house wares, there is a trained employee serving at each service unit. Saleschеckout service has sufficient number of employees (servers) which is helpful during the peak hours of weekdays. Other than these hours, there is a possibility of short Queues in a model and hence no need to open all checkouts counters for each hour. Increasing more than sufficient number of servers may not be the solution to increase the efficiency of the service by each service unit.

When servers are analyzed with one queue for two parallel servers, the results are estimated as per server whereas when each server is analyzed with its individual queue, the results computed from simulation are for each server individually.

REFERENCES

Abolnikov, L., Dshalalow, J. E., Dukhovny, A. M. (1990), “On Some Queue Length Controlled Stochastic Processes,” Journal of Applied Mathematics and Stochastic Analysis, Vol. 3, No. 4

Adan, I.J.B.F., Boxma1, O.J., Resing, J.A.C. (2000), “Queuing models with multiple waiting lines,”Department of Mathematics and Computer Science, Eindhoven University of Technology,

Banks, J., Carson, J. S., Nelson, B. L., Nicol, D. M. (2001), Discrete-Event System Simulation,
Prentice Hall international series, 3rd edition, p24–37

Bhatti, S. A., Bhatti, N. A. (1998), *Operations Research – an Introduction*, Department of ComputerScience, Quad-e-Azam University, p.315–356

Hillier, F. S., Lieberman, G. J. (2001), *Introduction to Operations Research*, McGraw-Hill higher education,7th edition, p834–8

Jensen, Paul A. (2004), “Queuing models,” Operations Research Models and Methods, www.me.utexas.edu/~jensen/ORMM/models/unit/queue/index.html

Keith G. Calkins (May 2005), “Queuing theory and Poisson distribution,” Statistical Probabilities and Distributions, Ch. 10,

www.Andrews.edu/~calkins/math/webtexts/prod10.html

Nasroallah, A. (2004), "Monte Carlo Simulation of Markov Chain Steady-state Distribution," ExtractaMathematicae, Vol. 19, No. 2, p279-288

Sheu, C., Babbar S. (Jun 1996), "A managerial assessment of the waiting-time performance for alternative service process designs," Omega, Int. J. Mgmt Sci. Vol. 24, No. 6, pp. 689-703

Taha, Hamdy A. (1997), Operations Research an Introduction, PHIPE Prentice, 6th edition, p607–643

Tsuei, Thin-Fong; Yamamoto, W., "A Processing queuing simulation model for multiprocessor system performance analysis," Sun Microsystems, Inc

Troitzsch, Klaus G., Gilbert, Nigel (Sep 2006), "Queuing Models and Discrete Event Simulation,"

BIOGRAPHY OF AUTHORS

TMJA Cooray is a Senior Lecturer in Mathematics at the Department of Mathematics Faculty of Engineering , University of Moratuwa with 30 years of teaching experinece. He was also a visiting lecture at the Sir John Kotlawala Defence University, Ratmalana, Sri Lanka.

J.S.K.C. Priyangika is a Lecturer in Mathematics, Department of Physical Sciences , Faculty of Applied Sciences, Rajarata University of Sri Lanka, Mihinthale. Her field of interest : statistics & Operational Research.