ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 6, November 2013

Simulation of Queuing Models

Arnika Tripathi

Abstract*— The queue is where customers wait before being served. A queue is characterized by the maximum permissible number of customers that it can contain. Queues are called infinite or finite, according to whether this number is infinite or finite. The assumption of an infinite queue is the standard one for most queuing models, even for situations where there actually is a (relatively large) finite upper bound on the permissible number of customers, because dealing with such an upper bound would be a complicating factor in the analysis. However, for queueing systems where this upper bound is small enough that it actually would be reached with some frequency, it becomes necessary to assume a finite queue. The queuing system is a typical problem of discrete event system, and the computer simulation is a quite effective way for solving the queuing problem and analyzing the performances of the queuing system. So in this paper we analyze queuing models using queuing model simulator.

***Index Terms*— server, Customers, queuing model, Service time, Arrival time.**

I. INTRODUCTION

A sales checkout service has 5 waiting lines in a form of parallel cash counters (see fig.1) Customers are served on a first-come, first-served (FIFO) 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 (fig.1), whereas the second part is done by queuing simulation for second model involving 2 queues for each corresponding parallel server (fig.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:

A) Confidence Intervals

[(mean arrival time +1.96 × SE (mean arrival time))$^{-1}$,(mean arrival time – 1.96 × SE mean arrival time))$^{-1}$]
Where SE (mean arrival time) = SD(mean arrival time)
Similarly, 95% confidence interval for service rate can be:
[(mean service time +1.96 × SE (mean service time))$^{-1}$, (mean service time – 1.96 × SE mean service time))$^{-1}$]
Where SE(mean service time) = SD(mean service time)

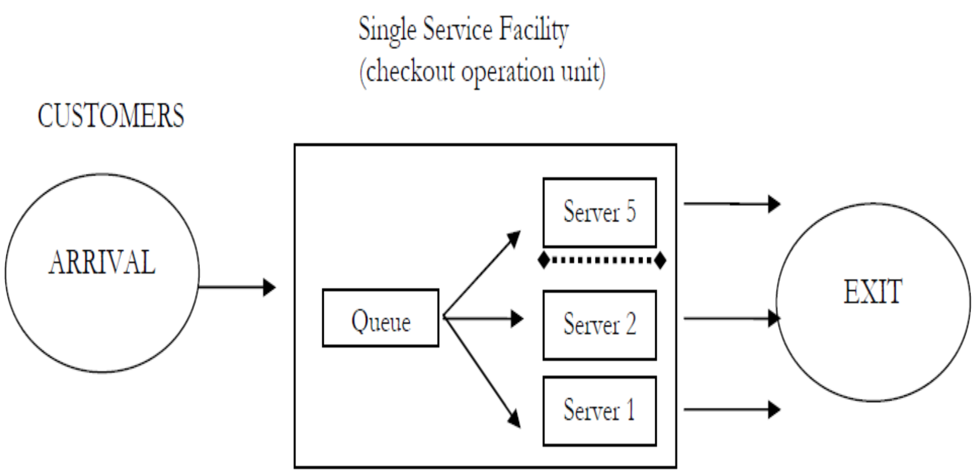

Fig.1 Single Stage Queuing Model with Single-Queue and Multiple Parallel Servers

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 6, November 2013

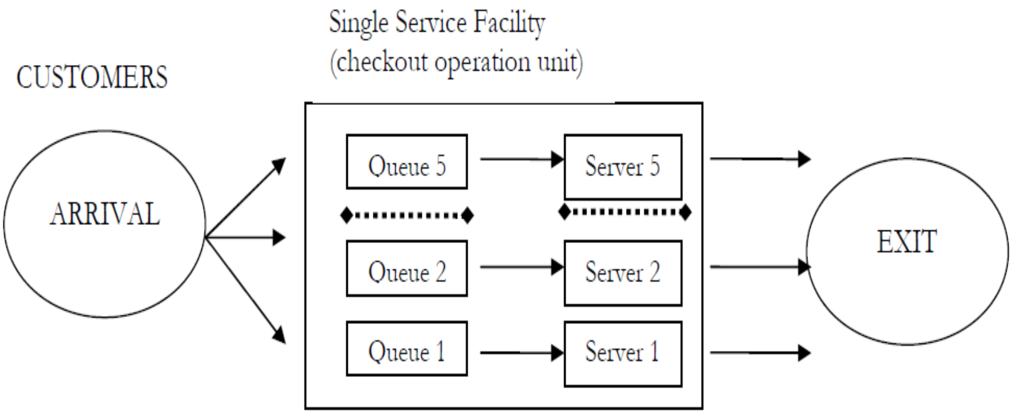

Fig.2 Single Stage Queuing Model with Multiple Queues and Multiple Parallel Servers

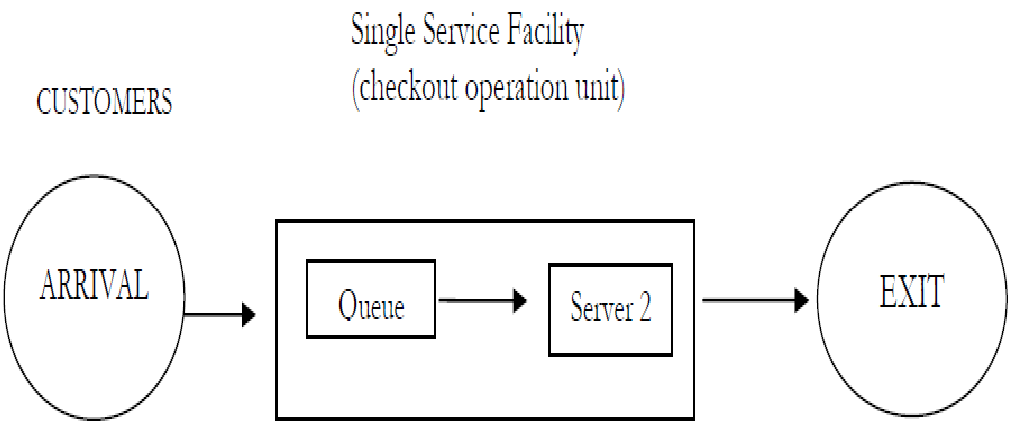

Fig.3 Single Stage Queuing Model with Single-Queue and Single-Server

II. CONFIDENCE INTERVALS FOR WEEKDAY

We have,
Mean (service time) = 01:06 minutes per customer (read clock as min:sec)
SD (service time) = 00:06 min
Mean (arrival time) = 00:37 min per customer
SD (arrival time) = 00:06 min
And n = 41 customers

95% Confidence Intervals for Service Time:
Mean(service time) - 1.96 (SE(service time)) = 54 sec/customer
Mean(service time) + 1.96 (SE(service time)) = 78 sec/customer
SE = SD/sqrt(n)

95% Confidence Intervals for Service Rate:
1/[Mean(service time) + 1.96 (SE(service time))] = 0.01282 = 46 customers/sec
1/[Mean(service time) - 1.96 (SE(service time))] = 0.01852 = 67 customers/sec**
** (0.01852 sec *60 *60)

95% Confidence Intervals for Arrival Time:
Mean(arrival time) – 1.96 (SE(arrival time)) = 24 sec /customer
Mean(arrival time) + 1.96 (SE(arrival time)) = 49 sec /customer

95% Confidence Intervals for Arrival Rate:
1/[Mean(arrival time) + 1.96 (SE(arrival time))] = 0.02041 = 73 customers/sec**
1/[Mean(arrival time) - 1.96 (SE(arrival time))] = 0.04167 = 150 customers/sec
** (0.02041 sec *60 *60)

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 6, November 2013

III. INTERPRETATION OF CONFIDENCE INTERVALS

The confidence intervals show that 73 to 150 customers arrive in 2-server system within an hour whereas 46 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.

IV. 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.

V.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.

A) Results for Weekday applying Queuing model 1(fig. 1)

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

c number of servers = 2
λ arrival rate = 98 customers per hour
μ serving rate = 55 customers per server per hour
cμ (2) (55) = 110 (service rate for 2 servers)
ρ = λ/(cμ) = 98 / 110 = 0.8909
γ = λ/μ = 1.7818
Overall system utilization = ρ = 89.09 %
The probability that all servers are idle (P_o) = 0.5769
Average number of customers in the queue (L_q) is

$$\frac{\gamma^c \rho}{(c)!\,(1-\rho)^2} \times P_0 = 6.8560$$

Average time customer spends in the queue (W_q) = L_q/λ = 0.0700 hours

B) 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.

C) Results for Weekday applying Queuing model 3 (fig. 3)

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

c number of servers = 1
λ arrival rate = 98 customers per hour for 2 servers i.e. 49 customers
μ serving rate = 55 customers per server per hour
ρ = λ/(cμ) = (98 ÷ 2)/ 55 = 0.8909
γ = λ/μ = 0.8909 (=ρ in case of c = 1)
Overall system utilization = ρ = 89.09 %
The probability that all servers are idle (P_o) = 0.1091
Average number of customers in the queue (L_q) is

$$\frac{\gamma^c \rho}{(c)!\,(1-\rho)^2} \times P_0 = 7.2758$$

Average time customer spends in the queue (Wq) = Lq/λ = 0.1485 hours

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 6, November 2013

D) Interpretation of results for queuing model 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.

VI. 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 interarrival time and mean service time as taken same for both servers. Results for Weekday applying Queuing model 2.

Server 1

Mean interarrival time = 0.6333 min
Mean Serving time = 1.1000 min
Server utilization = ρ = 99.00 %
Number customers served = 93 customers
Average number of customers in the queue (L_q) = 28.1820 customers
Average time customer spends in the queue (W_q) = 21.3131 min

Server 2

Mean interarrival time = 0.6333 min
Mean Serving time = 1.1000 min
Server utilization = ρ = 99.00 %
Number customers served = 77 customers
Average number of customers in the queue (L_q) = 39.3991 customers
Average time customer spends in the queue (W_q) = 28.8511 min

Overall for two servers

Mean Serving time = 1.1000 min
Server utilization = ρ = 99.00 %
Average number of customers in the queue (L_q) = 67.5812 customers
Average time customer spends in the queue (W_q) = 25.0821 min

A) Interpretation of Queuing Simulation Results For Model 2

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 interarrival 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 has multiple queues and queues have jockey behavior. In other words, customers tend to switch to a shorter queue to reduce the waiting time.

B) Comparison of The Results For Queuing Model 1 And Model 2

The actual structure of our survey example ICA has queuing model 2 . A queuing model with single queue and multiple parallel servers 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 (fig 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.

VII. CONCLUSION

This section 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

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 6, November 2013
system of ICA 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 sales operation service for shopping, either for usual grocery or for some house wares, there is a trained employee serving at each service unit. Sales checkout 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

[1] "Kuehn, P.J., 1979, "Approximate Analysis of General Queuing Networks by Decomposition", IEEE Transactions on Communications COM-27, 113-126.

[2] Vasumathi.A, Dhanavanthan P, "Application of Simulation Technique in Queuing Model for ATM Facility", International journal of applied engineering research, dindigul, Volume 1, No 3, 2010

[3] G. S. Fishman, "Estimation in Multiserver Queuing Simulations," Oper. Res. 22, pp 72, 1974.

[4] M. A. Crane and D. L. Iglehart, "Simulating Stable Stochastic Systems, I: General Multiserver Queues," J. ACM 21, pp 103, 1974.

[5] Atlason J, Epelman M, Henderson SG. Call centre staffing with simulation and cutting plane methods. Annals of Operations Research; 127(1-4): 333-358, 2004.