How to Cite:
Yadav, C., Sharma, R. C., & Shanker, U. (2022). Analysis of waiting and service cost for a multi-server queuing model in a tertiary care hospital. *International Journal of Health Sciences*, *6*(S8), 5140–5148. https://doi.org/10.53730/ijhs.v6nS8.13393

Analysis of waiting and service cost for a multi-server queuing model in a tertiary care hospital

Chandramani Yadav
Research Scholar, OPJS University, India
Corresponding author email: chandramaniyadav9@gmail.com

Dr. Ratnesh Chandra Sharma
Professor, OPJS University, Rajasthan, India

Dr. Uma Shanker
Associate Professor, OPJS University, Rajasthan, India

***Abstract*---**A common situation that occurs in everyday life is that of queuing or waiting in line. The problem of queuing in relation to the time spent by patients to access clinical services is increasingly becoming a major source of concern to most health –care providers. As the patients wait too long for service could result to cost to them which is called as waiting cost. Providing good service capacity to operate a system involves excessive cost. But not providing enough service capacity results in excessive waiting time and cost. In this study, the queuing characteristics at the tertiary care hospital of Firozabad was analyzed using a multi-Server queuing Model and the Waiting and service Costs determined with a view to determining the optimal service level. Data for this study was collected at the tertiary care hospital for four weeks through observations, interviews and by administering questionnaire. The data was analyzed using TORA optimization Software as well as using descriptive analysis. The results of the analysis showed that average queue length, waiting time of patients as well as over utilization of doctors at the Hospital could be reduced at an optimal server level of 24 doctors and at a minimum total cost as against the present server level of 27 doctors with high Total Cost which include waiting and service costs. This model can also be used by decision makers and other policy makers to solve other multi-Server queuing problems.

***Keywords*---**waiting cost, service cost, queuing model, multi-server, hospital queuing.

Introduction

Queues or waiting lines or queuing theory, was first introduced by A.K. Erlang a Danish Engineer in 1913 in the context of telephone facilities. He was experimenting with the fluctuating demand for telephone facilities and its effect on automatic dialing equipment at the Copenhagen telephone System. Since World War II this theory has been applied to many business and human service fields. Literature on queuing indicates that waiting in line or queue causes inconvenience to economic costs to individuals and organizations. Healthcare/Emergency Services, airline companies, banks, manufacturing firms etc., try to minimize the total waiting cost, and the cost of providing service to their customers.

In this paper the optimum total System cost is analyzed using waiting and Service cost. These two basic costs mentioned are costs associated with patients or customers having to wait for service (Wait Cost) which include loss of business as some patients might not be willing to wait for service and may decide to go to the competing organizations, cost due to delay in care or the value of the patient's time (opportunity cost of the time spent in queuing) and decreased patients satisfaction and quality of care. While Service Cost is the cost of providing service. These includes salaries paid to employees, salaries paid to employees or servers while they wait for service from other servers [18]. Cost of waiting space, facilities, equipment, and supplies. Using the estimation of waiting cost allows decision makers to have the capability of determining the optimal number of servers by minimizing the total cost including the service cost and the waiting cost. The cost of waiting for every individual differs depending on what the individual earns every hour. Some might have their cost of waiting in multiples of other people's value.

Reviews research on models for evaluating the impact of bed assignment policies on utilization, waiting time, and the probability of turning away patients.[11] reviewed the use of queuing theory in pharmacy application with particular attention to improving customer satisfaction. Customer satisfaction is improved by predicting and reducing waiting times and adjusting staffing. [20] Proposes an incremental analysis approach in which the cost of an additional bed is compared with the benefits it generates. Beds are added until the increase cost equal the benefits. [16] Considered a pharmacy queuing system with pre-emptive service priority discipline where the arrival of a prescription order suspends the processing of lower priority prescriptions. Different costs are assigned to wait-times for prescriptions of different priorities. [4] Chose the number of messengers required to transport patients or specimens in a hospital by assigning costs to the messenger and to the time during which a request is in queue. The author also calculated the number of servers required so that a given percentage of requests do not exceed a given wait time and the average number of patients in queue do not exceed a given threshold. [7] Incorporated advertising into their model to control the demand for laboratory services. The model assumes that clients would leave without service if they wait above a certain amount of time. In this study, Waiting and Service Costs at the clinic using a multi-Server queuing model with a view determining the optimal service level was studied.

Materials and Methods

The M/M/S Model

The model adopted in this work is the (M/M/S) :(∞ /FCFS) -Multi-server Queuing Model. For this queuing system, it is assumed that the arrivals follow a Poisson probability distribution at an average of λ customers (patients) per unit of time. It is also assumed that they are served on a first- come, first-served basis by any of the servers (in these case doctors). The service times are distributed exponentially, with an average of μ customers (patients) per unit of time and number of servers S. If there are n customers in the queuing system at any point in time, then the following two cases may arise:

1. If n<S, (number of customers in the system are less than the number of servers), then there will be no queue. However, (S–n) number of servers will not be busy. The combined servicerate will then be $\mu_n = n\mu$; n<S
2. If n≥S, (number of customers in the system are 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 – s). The combined service rate will be; $\mu_n = S\mu$; n≥s

From the model the probability of having n customers in thesystem is given by

$$p_n = \begin{cases} \left(\rho^n/_{n!}\right) p_0 & n \le S \\ \rho^n/_{(S!\, S^{n-s})}\, p_0 & n > S;\ \rho = \dfrac{\lambda}{S\mu} \end{cases}$$

$$p_0 = \left[\sum_{n=0}^{S-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{S!}\left(\frac{\lambda}{\mu}\right)^S \frac{S\mu}{S\mu-\lambda}\right]^{-1} \quad \text{.....(1)}$$

We now proceed to compute the performance measures of the queuing system.

The expected number of the customer (patients) waiting on the queue (length of line) is given as:

$$L_q = \left[\frac{1}{(S-1)!}\left(\frac{\lambda}{\mu}\right)^S \frac{\mu\lambda}{(\mu S-\lambda)^2}\right] p_0 \quad \text{..... (2)}$$

Expected number of customers (patients) in the system

$$L_S = L_q + \frac{\lambda}{\mu} \quad \text{..... (3)}$$

Expected waiting time of customer (patients)in the queue

$$W_q = \frac{\lambda}{\mu} \quad \text{..... (4)}$$

Expected time a customer (patient) spends in the system:

$$W_s = \frac{L_S}{\lambda} \quad \text{..... (5)}$$

Utilization factor i.e. the fraction of time servers (doctors) are busy

$$\rho = \frac{\lambda}{\mu S} \quad \text{..... (6)}$$

Where:

λ = the arrival rate of patients per unit time,
μ = the service rate per unit time,
s = the number of servers,
p_0 = the probability that there are no customers (patients) in the system,
L_q = Expected number of customers in the queue,
L_S = Expected number of customers in the system,
W_q = Expected time a customer (patient) spends in the queue,
W_S = Expected time a customer (patient) spends in the System.

Introduction of Costs in to the Model

To evaluate and determine the optimum number of servers in the system, two opposing costs must be considered in making these decisions:

(i) Service costs
(ii) Waiting time costs of customers. Economic analysis of these costs helps the management to make a trade-off between the increased costs of providing better service and the decreased waiting time costs of customers derived from providing that service.

Expected Service Cost $E(SC) = SC_S$ ..... (7)

Where
S= number of servers,
C_S= service cost of each server

Expected Waiting Costs in the System
$E(WC) = (\lambda . W_S)C_w$ ..... (8)

Where λ =number of arrivals,
Ws= Average time an arrivalspends in the system
C_w= Opportunity cost of waiting by customers (patients) Adding (7) and (8) we have,

Expected Total Costs $E(TC) = E(SC) + E(WC)$..... (9)
Expected Total Costs $E(TC) = SC_S + (\lambda . W_S)\ C_w$.....(10)

Data Collection

Data for this study were collected from tertiary care hospital of Firozabad. The methods employed during data collection were direct observation and personal interview and questionnaire administering by the researcher. Data were collected for four weeks (Monday to Saturday). The following assumptions were made for queuing system at the tertiary care hospital which is in accordance with the queue theory. They are:

1. Arrivals follows Poisson probability distribution at average rate of λ customers (patients) per unit of time.
2. The queue discipline is First-Come, First-Served (FCFS) basis by any of the servers. There is no priority classification for any arrival.
3. Service times are distributed exponentially, with an average of μ patients per unit of time.
4. There is no limit to the number of the queue/patients (infinite).
5. The service providers are working at their full capacity.
6. The average arrival rate is greater than average service rate.
7. Servers here represent only doctors no other medical personnel.
8. Service rate is independent of line length; service providers do not go faster because the line is longer.
9. The Balking and Reneging behavior patients are not included in this study.

Results and Discussions

Analysis of Data

TORA Optimization software (Version 2.0 Feb. 2006) was used by us to compute the performance measures of the multi-server queuing system at the tertiary care hospital of Firozabad using arrival rate λ =175 patients/hr., Service rate μ= 8 patients/hr. and number of servers (S) = 27

Table1. Queuing Characteristics of Multi-server Queuing Model of the tertiary care hospital of Firozabad

Scenario	S	Lambda	Mu	L'da eff	p_o	Ls	Lq	Ws	Wq
1	22	175	8	175	0	191.3905	169.516	1.09366	0.9686
2	23	175	8	175	0	36.3405	14.465	0.20766	0.0826
3	24	175	8	175	0	27.65525	5.78	0.15803	0.033
4	25	175	8	175	0	24.78875	2.913	0.14165	0.0166
5	26	175	8	175	0	23.47975	1.604	0.13417	0.0091
6	27	175	8	175	0	22.7955	0.92	0.13026	0.0052
7	28	175	8	175	0	22.41225	0.537	0.12807	0.003
8	29	175	8	175	0	22.19	0.315	0.1268	0.0018
9	30	175	8	175	0	22.0605	0.184	0.12606	0.001
10	31	175	8	175	0	21.98175	0.107	0.12561	0.0006
11	32	175	8	175	0	21.93625	0.061	0.12535	0.0003

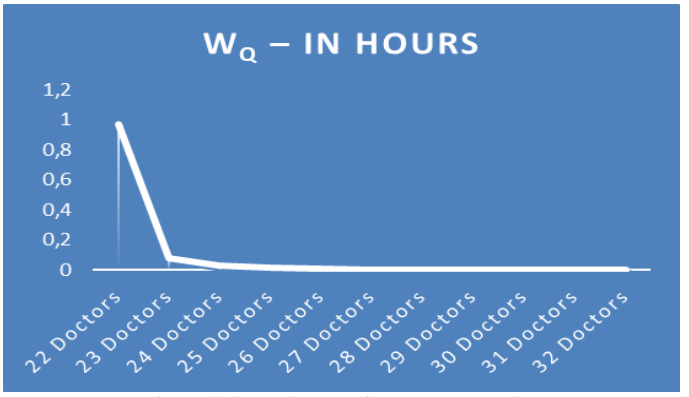

Fig. 1- Expected Waiting time of Customer in queue against

Fig. 2- Expected time of Customer spends in system against number of Server
number of Server

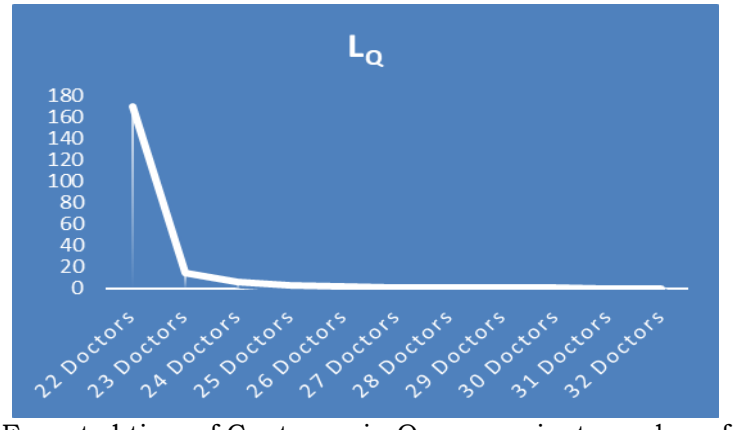

Fig. 3- Expected time of Customer in Queue against number of Server

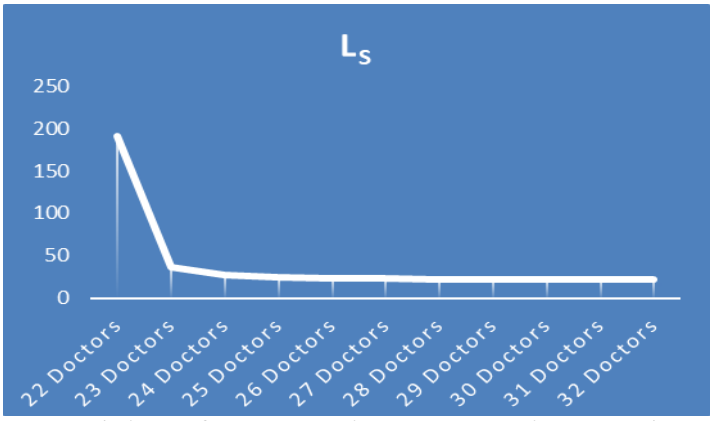

Fig. 4- Expected time of Customer in system against number of Server

Table 2: Summary analysis of the Multi -Server queuing Model of the tertiary care Hospital

Performance Measure	Arrival rate (λ)	Service rate(μ)	System Utilisation	L_S	L_q	W_S – in hours	W_S – in hours	Po	Total System Cost/hr
22 Doctors	175	8	99.43%	191.3905	169.516	1.09366	0.9686	0	19997.34
23 Doctors	175	8	95.11%	36.3405	14.465	0.20766	0.0826	0	9443.835
24 Doctors	175	8	91.15%	27.65525	5.78	0.15803	0.033	0	9135.868
25 Doctors	175	8	87.50%	24.78875	2.913	0.14165	0.0166	0	9235.213
26 Doctors	175	8	84.14%	23.47975	1.604	0.13417	0.0091	0	9443.583
27 Doctors	175	8	81.02%	22.7955	0.92	0.13026	0.0052	0	9695.685
28 Doctors	175	8	78.13%	22.41225	0.537	0.12807	0.003	0	9968.858
29 Doctors	175	8	75.43%	22.19	0.315	0.1268	0.0018	0	10253.3
30 Doctors	175	8	72.92%	22.0605	0.184	0.12606	0.001	0	10544.24
31 Doctors	175	8	70.56%	21.98175	0.107	0.12561	0.0006	0	10838.72
32 Doctors	175	8	68.36%	21.93625	0.061	0.12535	0.0003	0	11135.54

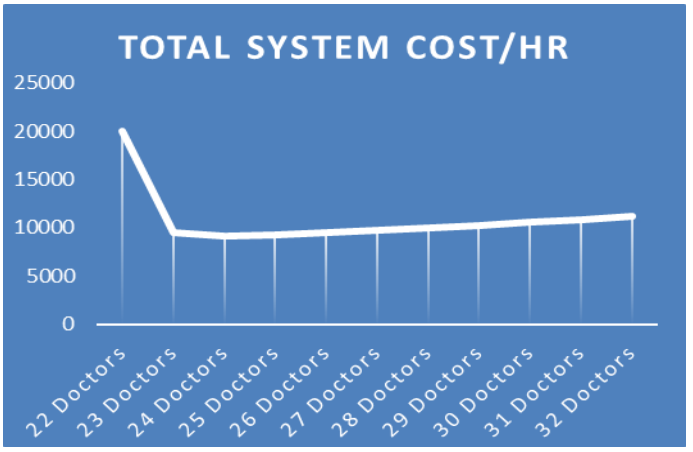

Fig. 5- Total System cost per hour against number of Servers

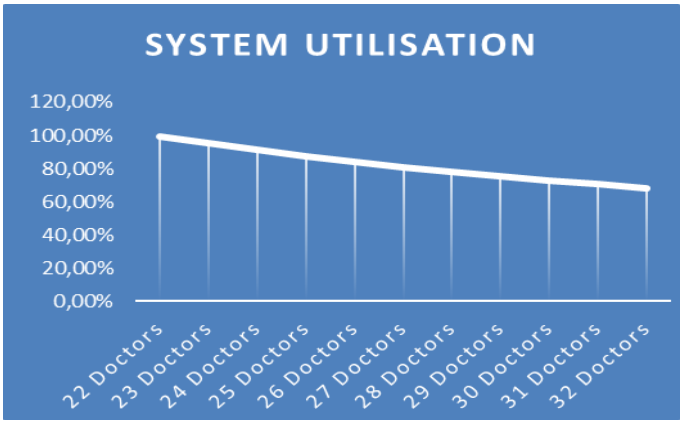

Fig. 6- System utilization against number of Servers

Discussion of Result

The graphs (Fig. 5) show that optimal server level at the Clinic is achieved when the number of servers (doctors) is 24 with a minimum total cost of Rs 9135.868 per hr as against the present server level of 27 doctors at the Clinic which have high total cost of Rs. 9695.685 per hr. It should also be noted that patients' average wait time and congestion in the system is also less at this optimal server level.

Conclusion

The queuing characteristics at the tertiary care hospital of Firozabad was analyzed using a Multi-Server queuing Model and the Waiting and service Costs determined with a view to determining the optimal service level. The results of the analysis showed that average queue length, waiting time of patients as well as underutilization of doctors could be increased when the service capacity level of doctors at the Clinic is decreased from 28 to 24 at the minimum total costs which include waiting and service costs. Service cost gets decrease as a hospital attempts to shrink its level of service. This could be done by shrinking the service facilities or using models that consider cost optimization.

References

1. Amakon, U.S ".How Large is the Opportunity of Queuing in Service Centres? Evidence from Eastern Nigeria". Department of Economics,Nnamdi Azikiwe University, Awka.2008.
2. Bastani, P " A Queuing Model of Hospital Congestion" An Unpublished M.Sc Theses, Department of Mathematics, Simon Fraser University Burnaby,B.C Canada. 2009.
3. Elegalam "Customer Retention Versus Cost Reduction technique" A Paper Presented at the Bankers Forum held at Lagos, pg.9-10.1978.
4. Gupta, I., Zoreda, J. and Kramer, N. "Hospital Manpower Planning by Use of Queuing Theory". Health Services Research, 6, 76-82.1971
5. Kandemir-Caues, C., Cauas, L. ", An Application of Queuing Theory to the Relationship between Insulin Level and Number of Insulin Receptors".

Turkish Journal of Biochemistry, 32 (1): 32-38, 2007.
6. kembe, M.M., Onah, E.S., lorkegh, S, "A Study of Waiting and Service cost of a Multi-Server queuing Model in a Specialist Hospital" international journal of scientific & technology research volume 1, issue 8, september 2012 issn 2277-8616.
7. Khan, M.R. and Callahan, B.B. "Planning Laboratory Staffing with a Queuing Model". European Journal of Operational Research, 67, 1993.
8. Kostas, U.N. "Introduction to Theory of Statistics". Mc-Gram Hill, Tokyo. 1983.
9. McClain, J.O. ".Bed Planning Using Queuing Theory Models of Hospital Occupancy: a Sensitivity Analysis". Inquiry, 13,167-176,http://www.ncbi.nlm.nih.gov/ 1976.
10. Ndukwe, H.C, Omale, S and Opanuga O.O ' Reducing Queues in Nigerian Hospital Pharmacy".African Journal of pharmacy and pharmacology Vol.5(8).pp.1020-1026. 2011.
11. Nosek, R.A. and Wilson, J.P." Queuing Theory and Customer Satisfaction: A Review of Terminology, Trends and Applications to Pharmacy Practice. Hospital Pharmacy", 36,275-276, http://www.drum.lib.umd.edu. 2001.
12. Obamiro, J.K. 'Queuing theory and Patient Satisfaction: An overview of terminology & application in Ante-Natal care unit.http://www.upg-bulletin-se.ro
13. Olaniyi, T.A." An Appraisal of Cost of Queuing in Nigerian Banking Sector: A Case Study of First Bank of Nigeria Plc, Ilorin".Journal of Business & Social Sciences. Vol.9, Nos,1&2, pages 139-145. 2004.
14. Rosenquist, C.J." Queuing analysis: A Useful planning and Management techniques for radiology". Journal of Medical Systems, 11,413-4, 1987.
15. Sharma, J.K. "A text book on Operations Research; Theory Applications". 4th Edition. Macmillan publishers, India. 2009.
16. Shimshak, D.G., Gropp, D.D. and Burden, H.D. " A Priority Queuing Model of a Hospital Pharmacy Unit." European journal of operational Research .7, 350-354. 1981.
17. Stakutis C, Boyle T " Your Health, your Way: Human-enabled Health Care." CA Emerging Technologies, pp. 1-10.2009.
18. Vikas, S.. " Use of Queuing Models in Healthcare: Department of Health Policy and Management", University of Arkansas for Medical science Available at: http://works.bepress.com/vikas-singh/13.2006.
19. Young, J.P. ."Estimating bed Requirements in a Queuing Theory Approach to the Control of Hospital Inpatient Census," John Hopkins University, Baltimore, 98-108. http://online library.wiley.com.1962b.
20. Young, J.P. "The Basic Model, in a Queuing Theory Approach the Control of Hospital Inpatient Census", John Hopkins University, Baltimore, 74-79.http/online library. Wiley.com.1962a.