

THE OPTIMAL COST OF AN OUTPATIENT DEPARTMENT BY USING SINGLE AND MULTIPLE SERVER SYSTEM

Dr. Win Lei Lei Aung

Associate Professor

Faculty of Computing, University of Information Technology, Hlaing , Yangon

wllaung@gmail.com

Abstract— Hospital management tradeoffs always exist between the costs in providing better service and waiting time of patients in any hospital. The objective of this study was to investigate whether increasing the cost for better service decreases the cost of patients waiting time or not, by using the technique of single and multiple servers which is based on the theory of Markovian queuing system. For this study, four weeks data has been taken from a public hospital. The results shows that as the service capacity level of doctors at the hospital increases from three to four servers then minimum total costs (include waiting and service costs) and the waiting time of patients as well as overutilization of doctors can be reduced. This paper also suggests that increasing the service units up to four servers will achieve lower cost as against two or more service units.

Keywords: *Outpatient Department (OPD), Performance measures, the service cost, the waiting cost*

I. INTRODUCTION

Outpatient department (OPD) is one of the most important parts of hospital management and is visited by large section of community. This is the first point of contact between patient and hospital staff. The problems faced by the patients in that department are overcrowding, delay in consultation, lack of proper guidance etc. which lead to patients becoming dissatisfied. Every patient in each hospital is in search of hassle free and quick services in this fast growing world which is only possible with optimum utility of the resources through multitasking in a single server system in the OPD for better services. As patients flow increases because of that demands for hospitals and, better and quick services increases in [1] and [2].

In any hospital, patients come to the Outpatient Department without prior appointment and patients have to wait to receive medical service that may be

waiting before, during or after being served. Generally queues are formed when the demand for a service exceeds its supply. For many patients or customers, waiting in lines or queuing is annoying or negative experience. A few of the factors that are responsible for long waiting lines or delays in providing service are: lack of passion and commitment to work on the part of the hospital staff, overloading of available doctors, doctors attending to patients in more than one clinic etc.

A good patient flow means that the patient queuing is minimized while a poor patient flow means patients suffer considerable queuing delays [7]. Considering these points mentioned above, our present study proposes to evaluate the patients waiting problems in terms of the performance measure and also to determine the level of service that minimizes the total cost of the expected cost of service and the expected cost of waiting.

II. MATERIALS AND METHODS

The study area is Yangon, which is a fast developing city in the MYANMAR. In every hospital, Yangon is actually facing a big issue of patient waiting time problem at every day. There is heavy inflow of sick patients in this region from neighboring rural areas or from smaller towns because of the availability of advanced health facilities. For our study we selected one of the leading public hospitals of the region, North Oaklapa General Hospital in Yangon where it was observed that there was a heavy flow of patients thought the week. Data was collected over a period of four Weeks from 15.10.2019 to 13.11.2019 as shown in Table 1. Table 1 shows the data collected for one week. These data represents the number of patients at OPD. Data was collected from the Patients who visited Out- Patient Department during day shift by using direct observation method.

In this observational study, the traffic intensity of the out patients at registration counters such as the arrival rate (λ), service rate (μ) and number of

servers was measured at hourly interval. The standard simple queuing model assumes that

1. Arrivals have the Poisson distribution
2. Service times have the exponential distribution
3. Arrivals and service times are all independent. (Independence means, for example, that: arrivals don't come in groups, and the server does not work faster when the line is longer.)

Table 1 measured data for one week

Time (hr)	9:00 AM to 10:00 OAM	10:00 AM to 11:00 AM	11:00 AM to 12:00 O AM	1:00 PM to 2:00 PM	2:00 PM to 3:00 PM	3:00P M to 4:00P M
days						
15.10.19	60	70	80	65	50	45
16.10.19	43	63	70	100	45	55
17.10.19	33	56	50	120	55	65
18.10.19	23	88	45	80	65	77
19.10.19	12	102	35	90	78	78
20.10.19	9	133	25	105	73	82
21.10.19	8	117	40	120	88	63

In this M/M/m queuing model, the 1st M indicates the inter-arrival time distribution arrivals follows Poisson distribution with parameter λ , and 2nd M indicates service time distribution [5]. Here, it is exponential distribution with parameter μ and 3rd m indicates number of servers available those are in parallel. The system has multiple servers and uses the FIFO service discipline. The waiting line is of infinite size. In the case M/M/1 queuing model, there is only one single server. It means that the system can process a single request at a time. But in this M/M/m queuing model, there are m numbers of servers that are connected in parallel [5]. Queuing model enable finding an appropriate balance between the cost of service and the amount of waiting. The ultimate goal of queuing analysis is to minimize two costs, which is service capacity cost and customer waiting cost.

These are analyzed for simultaneous efficiency in patient satisfaction and cost minimization through the use of a single server M/M/1(∞ : FCFS) and multi-server queuing models, which are then compared for a number of queue performances such as; the average time each patient spends in the queue and in the system, average number of patients in the

queue and in the system and the probability of the system being idle. In M/M/1(∞ : FCFS) queuing model, the arrival of patients in a fixed time interval belongs to Poisson probability distribution at an average rate of λ patients per unit time. It is also assumed that the service time was exponentially distributed, with an average rate of μ patients per unit of time. The hypothetical structure of Single-server queuing model is shown in Figure 1. The hypothetical structure of Multi-server queuing system is as shown in Figure 2.

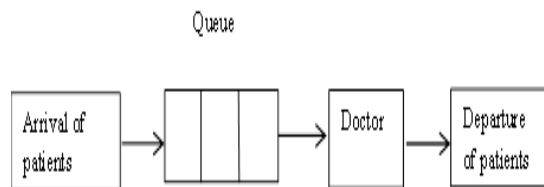


Figure 1 Single server queuing system.

The performance measures of single server system

Average server utilization factor in the system

$$\rho = \frac{\lambda}{\mu} \quad (1)$$

Probability that no patient in the system

$$P_0 = 1 - \rho \quad (2)$$

Average number of patients in the system

$$L_s = \frac{\rho}{1-\rho} \quad (3)$$

Average time a patient spends in the system

$$W_s = \frac{L_s}{\lambda} \quad (4)$$

Average number of patients in the queue

$$L_q = L_s - \rho \quad (5)$$

Average time a patient spends in the queue

$$W_q = \frac{L_q}{\lambda} \quad (6)$$

For the multi-server queuing model, the M/M/c (∞ : FCFS) model has been adopted. The basic hypothetical structure of multi-server queuing model is shown in Figure 2.

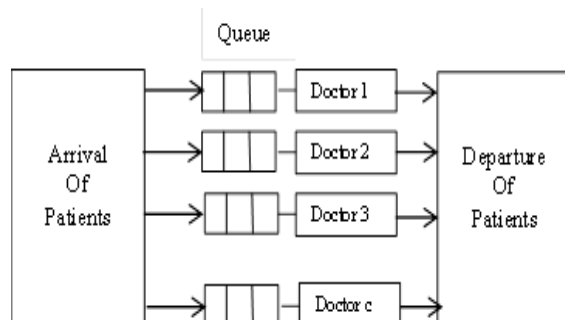


Figure 2 Multi server queuing system.

The performance measures of multi-server system
Average server utilization factor in the system

$$\rho = \frac{\lambda}{c\mu} \quad (7)$$

Probability that no patient in the system

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1} \quad (8)$$

Average number of patients in the queue

$$L_q = \frac{\rho(c\rho)^c}{c!(1-\rho)^2} P_0 \quad (9)$$

Average number of patients in the system

$$L_s = L_q + \rho \quad (10)$$

Average time a patient spends in the system

$$W_s = \frac{L_s}{\lambda} \quad (11)$$

Average time a patient spends in the queue

$$W_q = \frac{L_q}{\lambda} \quad (12)$$

In this queuing system, the arrival of patients is assumed to follow a Poisson process, and service times are assumed to have an exponential distribution. Let the number of servers be c , providing service independently of each other. It is also assumed that the arriving patients form a single queue and the one at the head of the waiting line enters into service as soon as a server is free. No server stays idle as long as there are patients to serve. If there are n patients in the queuing system, then two possibilities may arise:

Case.1: $k \leq c$. In this case no patient has to wait for service. However, $(c-n)$ patients will be in queue and the rate of servicing will be $k\mu$.

Case.2: $k > c$. In this case, all the doctors will be busy and the maximum number of patients ρ will be $(k-c)$ in the queue and the rate of service will be $c\mu$.

The Variables are analyzed by using the performance measures of the Queuing Models [M/M/1(∞ : FCFS) and M/M/c (∞ : FCFS)] as presented in equation (1) to (12). There are (5 x 4 weeks) working days in a month used in this study while the working hours per day are 24 hours for casualty service and 5 hours for outpatient service. Queuing models can be used to determine the operating performance of a waiting-line system.

In the economic analysis of waiting lines, we seek to use the information provided by the queuing model to develop a cost model for the waiting line under study. Then we can use the model to help the hospital management to make a trade-off between the increased costs of providing better service and the decreased waiting time costs of patients derived from providing that service. To determine the level

of service that minimizes the total cost of the expected cost of service and the expected cost of waiting, we utilize the cost analyzing model.

In cost model, we consider the cost of patient time, both waiting time and servicing time, and the cost of operating the system.

Let C_w = the waiting cost per unit per patient and

C_s = Cost of providing service per doctors per unit of time

Therefore, the total cost per minute is

Total cost = $C_w L_s + C_s c$, where L_s is the average number of patients in the system and c is the number of servers/doctors.

III. RESULTS

Using equation (1) to equation (12), we get the performance measures of single doctor system and multi doctor system as shown in following Table 2 and Table 3. Table 2 shows the average server utilization in the department, the probability of no patient in the department, the average number of patient in the outpatient department and the average waiting time in this department.

Table 2 Summarized data ρ , P_0 , L_s and W_s for single server and multi-server system

Doctors	Average patients Utilization in the system (ρ)	Probability that there are no patient in the system (P_0)	Average number of patients in the system (L_s)	Average time a patient spends in the system (W_s)
1	60%	0.40000	1.50000	0.50000
2	30%	0.53850	0.65935	0.11978
3	20%	0.54795	0.60616	0.20616
4	15%	0.54873	0.60062	0.20021
5	12%	0.54881	0.60006	0.20002
6	10%	0.54881	0.60001	0.20000
7	8.6%	0.54881	0.60000	0.20000

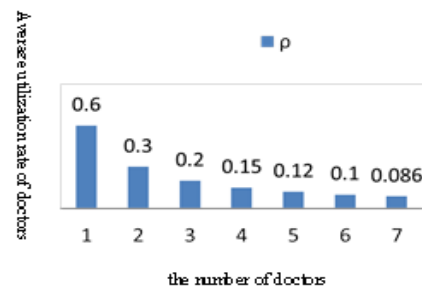


Figure 3 Utilization rate against the number of doctor.

In Figure 3, the average utilization rate of doctor decreases with increasing the number of doctor. Table 3 also summarizes the cost for the single server system and multiple-server system. We get the optimal cost from the data of Table 2.

Table 3 Summarized data L_q , W_q and the total cost for single server and multi-server system

Patients	Average number of patients in the system (L_q)	Average time a patient spends in the system (W_q)	The total cost
1	0.90000	0.30000	2250 kyats
2	0.05935	0.01978	989 kyats
3	0.00616	0.00206	909 kyats
4	0.00062	0.00021	900 kyats
5	0.00006	0.00002	900 kyats
6	0.00001	0.00000	900 kyats
7	0.00000	0.00000	900 kyats

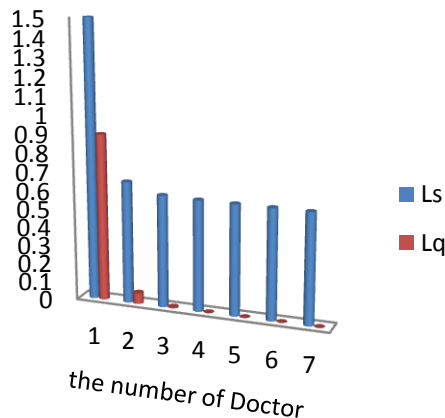


Figure 4 Average numbers of patients in the system and in the queue.

Figure 4 shows increased the number of doctor decreased the average number of patient in the queue. It saves time. For better graphical representations of the above summarized table, Figures 3-7 are shown. In Figure 3 we depict the average server utilization in the system against the number of doctors. As observed, server utilization rate decreases with increasing number of doctors. It is also noted in Figure 6 that the probability that there are zero patients rise upward as number of doctors increase and the expected length of the queue (L_q) and the system (L_s) in Figure 4 decline and rise upward respectively.

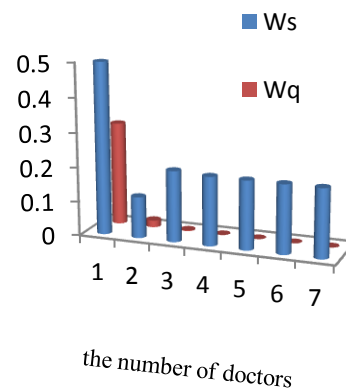


Figure 5 Average time a patient spends in the system and in the queue.

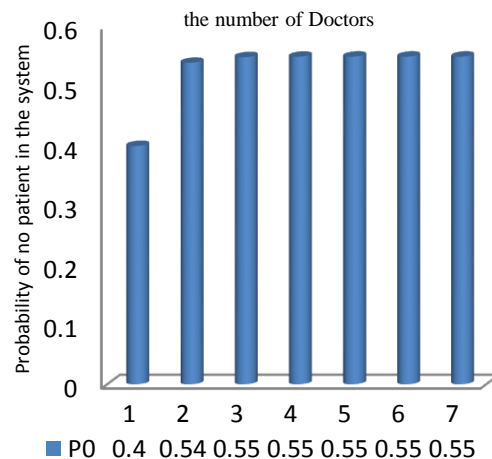


Figure 6 Probability of no patient in the system with the increased number of doctors.

Also the expected waiting time in the queue (W_q) and the expected time in the system (W_s) decrease in Figure 5. Figure 7 shows that the expected total cost and service cost fall downward and then rise upward. However, from the figures it is evident that the patients waiting time is optimum when the server is 4 as compared to server 2 and 3. It is also noted that patient's congestion and expected wait time is less than the optimum level. In optimizing of queue, there are optimizing over the number of servers, optimizing over the mean service rate and optimizing over the mean arrival rate [14]. In this paper, we get the optimal level from optimizing over the number of servers (doctors).

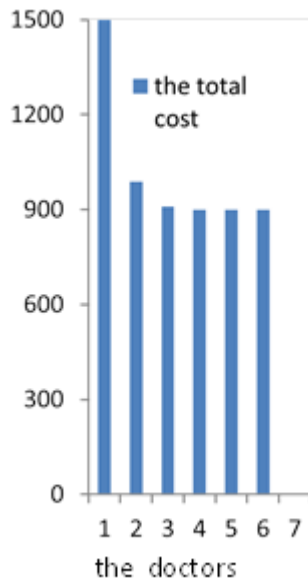


Figure 7 the optimal cost against the number of the doctors.

IV. DISCUSSION

Table 2 shows that a 4-server system is better than a single server, 2-server or 3-server system in terms of the performance criteria used. For instance, in terms of cost considerations, assume that the waiting cost is 1500 kyats per patient, for 4-server system records the lowest cost of 900 kyats compared to a 2-server and 3-server system that records 989kyats and 909 kyats respectively. These costs included the entire cost done by the hospital. The average time a patient spends in the system and in the queue are 0.2 minutes and 0.00021 minutes respectively for a 4-server system.

The probabilities of idleness are 30.0% and 20.0% respectively for 2 and 3 server systems respectively. The average time a patient spends in the queue and in the system for a single server system is 0.3 minutes and 0.5 minutes respectively while the system has 4.5 and 1.5 patients in the queue and in the system respectively. The system is likely to be idle for 0.4 minutes.

In [8], a survey comparing the performance of a single channel with multi-channel queuing models in achieving cost reduction and patient satisfaction using a hospital study, it was concluded that the 3-server system is better than a single server, 2-server or 4-server system. The average time a patient spends in the system and in the queue are 11 minutes and 1.79 minute.

However, another study revealed a positive correlation between arrival rates of customers and

bank's service rates where it was concluded that the potential utilization of the banks service facility was 3.18% efficient and idle 68.2% of the time [9]. A One week survey revealed that 59.2% of the 390 persons making withdrawals from their accounts spent between 30 to 60 minutes while 7% spent between 90 and 120 minutes [10]. Further it was observed in another survey that although the mean time spent was 53 minutes by customers, they prefer to spend a maximum of 20 minutes only [11]. Their study revealed worse service delays in urban center's (average of 64.32 minutes) compared to (average of 22.2 minutes) in rural areas.

Moreover those customers spend between 55.27 to 64.56 minutes making withdrawal from their accounts Efforts in this study are directed towards application of queuing models in capacity planning to reduce patient waiting time and total operating costs [12].

V. Conclusion

The results of the analysis showed that that as the service capacity level of doctors at the hospital increases from three to four servers then minimum total costs (include waiting and service costs) and the waiting time of patients as well as overutilization of doctors can be reduced. The study also suggests that, to optimize the processing time for the patients it is necessary to rationalize the utilization of the servers for effective utilization of human resource. Otherwise, the service units may be increased to four to achieve better results at a lower cost as against two or three service units. A single server is not effective as much as compared to multiple servers. Whereas, five servers eliminates waiting cost but at a higher cost which is not optimal too.

ACKNOWLEDGMENTS

Firstly, I would like to thank Dr. Pyke Tin who shares ideas and helpful suggestion. I am also grateful to Dr. Swe Swe Kyaw , head of Faculty of Computing , who motivates me to do this. And then, I appreciate to my mother for her patient, understanding and encouragement during my work that has to successful finish. Finally, I am also thankful for all my teachers and colleagues since my childhood and I sincerely thank all concerned persons of North Oaklapa General Hospital for their help and cooperation during the data collection

process.

REFERENCES

- [1] Srinivasan, A. V. (2000). Managing a Modern Hospital, Response Books. A Division of Sage Publishers India (P) Ltd., New Delhi, 53-69.
- [2] Llewelyn, R. and H. M. C. Davies.(1996). Macaulay Hospital Planning and Administration, WHO, Geneva, 105-11.
- [3] Hiller, S.F. and Lieberman, J. G., Introduction to Operations Research, Boston: Mcgraw Hill, Eight Edition.pp.834-838, 2005.
- [4] Singh V., "Use of Queuing Models in Health Care", Department of Health Policy and Management, University of Arkansas for Medical Sciences, 2007.
- [5] Raj Jian, "The Art Of Computer Systems Performance Analysis", John Wily & Sons Inc, computer press award winner, pp 592-604, 1992.
- [6] Trivedi, Kishor S, "Probability and Statistics with Reliability, Queuing and Computer Application", John Wiley & Sons Inc, 2016.
- [7] Hall, W. R. (2006). Patient Flow: The New Queuing Theory for Healthcare, OR/Ms, Today 2006.
- [8] Brahma, P. K. An Appraisal of cost of queuing in healthcare sector a case study of IMS and SUM hospital, International Journal of Multidisciplinary Research, Vol. 2(4).
- [9] Oladapo. (1998). Relationship between Arrival Rates of Customers and Bank's Service Rates, Seminar Paper Series, Obafemi Awolowo University, Ile-Ife.
- [10] Elegalam. (1978). Customer Retention Versus Cost Reduction Technique, A Paper presented at the Bankers forum held at Lagos.
- [11] Baale, A. F. (2002). Managerial Applications of Queuing Models in the Service Industry: A Int. j. sci. footpr. Dhar et al. (2014) Case Study of Afribank Nigeria Plc.,Ilorin Main Branch. A Research Paper submitted to Business Administration Department, University of Ilorin, Ilorin.
- [12] Juwah. Operational Problems and Customers' Dissatisfaction, CBN Bullion, 1986.
- [13] S Kalavathy, "Operations Research", Vikas Publishing House Pvt Ltd,pp411-440, 2000.
- [14] Joseph G.Ecker and Michael Kupferschmid, "Introduction to Operations Research", John Wily & Sons, Inc, pp 409-413, 2014.