International Journal of Advanced Statistics and Probability, 6 (2) (2018) 50-54

International Journal of Advanced Statistics and Probability

Website: www.sciencepubco.com/index.php/IJASP
doi: 10.14419/ijasp.v6i2.12332
Research paper

Minimal cost service rate in priority queuing models for emergency cases in hospitals

Nse. S. Udoh [1] *, Idorenyin A. Etukudo [2]

[1] Department of Mathematics and Statistics, University of Uyo, Nigeria
[2] Department of Mathematics and Statistics, Akwa Ibom State University, Nigeria
**Corresponding author E-mail: nsesudoh@uniuyo.edu.ng*

Abstract

Performance measures and waiting time cost for higher priority patients with severe cases over lower priority patients with stable cases using preemptive priority queuing model were obtained. Also, a total expected waiting time cost per unit time for service and the expected service cost per unit time for priority queuing models: M/M/2: ∞/NPP and M/M/2: ∞/PP were respectively formulated and optimized to obtain optimum cost service rate that minimizes the total cost. The results were applied to obtain optimum service rate that minimizes the total cost of providing and waiting for service at the emergency consulting unit of hospital.

Keywords: *Priority Queuing Models; Preemptive Model; No Preemptive Model. Waiting Time Cost; Optimum Cost Service Rate; Emergency Case.*

1. Introduction

Two priority queuing models; the preemptive priorities (PP) and the Nonpreemptive priorities (NPP) have been considered. In PP, a lower-priority patient with stable case being served can be ejected back into the queue or interrupted whenever a higher-priority patient with serious/critical case, where prompt treatment is vital enters the queuing system. In NPP, a patient being served cannot be ejected back into the queue even if a higher-priority customer enters the queuing system despite the severity of the case.
In order to solve this queuing problem, service facility must be organized so that an optimum balance is obtained between the cost of waiting time and the cost of idle equipment, Gupta and Hira (2009). The cost of waiting-patients generally includes either direct cost of idle doctors and other medical personnel or indirect cost of loss of goodwill due to dissatisfactory service. However, the cost of idle service facilities is the payment to be made for the period.
Hagen et al (2013) examined several queuing models for intensive care units and the effects on waiting times, utilization, return rates, mortalities and number of patients served. Then, Li, et al (2013) designed a hybridized-queuing model for effective packets scheduling in inter-vehicular communication with a view to obtaining the difference between delay-sensitivity and packet length of services. A study of five schemes proved that the nonpreemptive short-packet-first scheme results in the minimal overall delay. The model was superior in terms of performance indices in packet delivery ratio and throughput. Also, Ke, Li and Ni (2012) used the priority queuing model to study the performance of various strategies based on delay sensitivity and packets length. The non-preemptive short-packets-first strategy was proved to result in the minimal overall delay with different strategies for delay-sensitive and non-delay sensitive services. Consequently, an optimal priority-queuing model for the scheduling of multiple internet services was designed based on the above conclusions. Siddarthan, Jones and Johnson (2006) investigated the increased waiting time costs imposed on society due to inappropriate use of the emergency department by patients seeking non-preemptive or primary care. It proposed a simple economic model to illustrate the effect of this misuse at a public or not-for-profit hospital. The result showed that non-emergency patients contribute to lengthy delays in the emergency department for all classes of patients. It therefore proposed a priority queuing model to reduce average waiting times.
Several other works on the analysis of emergency waiting time and queuing systems with priority service discipline abound in the literature; see, for example, Udoh (2010), Bedford and Zeephongsekkul (2003), Blake and Carter (1996), among others. The objective of this work is to formulate and optimize the total cost of waiting for service and service cost per unit time to determine the particular level of service which minimizes the total cost of providing service and waiting for that service. We also seek to obtain the associated waiting time costs for the higher priority patients and the lower priority patients.

2. Problem formulation

The waiting time and its associated losses can be decreased by increasing the investment in service facilities such as Doctors and other medical facilities. It is desirable to obtain the minimum sum of these two cost: cost of investment and operation, and cost due to waiting by patients for service. The optimum balance of costs can be obtained by scheduling the flow of patients requiring service and/or providing proper number of doctors and facilities. Both service facilities and the flow of the patients can be controlled by effective schedule of consulting time as well as provide necessary number of doctors and facilities to minimize the overall cost. According to Taha (2007), the estimated waiting time and idle time cost can be obtained as follows;

Let C_w = expected waiting cost/patient/unit time

L_s = expected number of patients in the system

C_t = cost of treating one patient/unit time (that is, cost per unit time

Of having a doctor available

μ = service rate or average number of patients treated/unit time

$\boldsymbol{\lambda}$ = average number of patients' arriving/unit time in the queue

The expected waiting cost per unit time;

$$W_c = C_w L_s = \frac{\lambda}{\mu - \lambda} \tag{1}$$

The expected service cost per unit time;

$$W_S = C_t \mu_{Sc} \tag{2}$$

In addition, total cost

$$C = C_w \frac{\lambda}{\mu - \lambda} + C_t \mu \tag{3}$$

The minimization condition is $\frac{\partial C}{\partial \mu} = 0$

$$\therefore \hat{\mu} = \lambda \pm \sqrt{\frac{\lambda C_w}{C_t}} \tag{4}$$

where $\hat{\mu}$ is the minimum cost service rate.

3. Performance measures and estimated waiting time cost for M/M/2: ∞/FCFS (in the highest priority class)

Notations

Let Pn = probability of exactly n customers in the queuing system

L = expected number of customers in queuing system

Lq = expected queue length (excluding customers being served

W = waiting time in the system (includes service time) for each individual customer; $W = E(w)$

Wq = waiting time in queue (excludes service time) for each individual customer; $W_q = E(w_q)$

N_t = number of servers in the queuing system at time t (t ≥ 0)

$P_n(t)$ = probability of exactly n customers in queuing system at time t

S (=2) = number of servers (parallel service channels) in the queuing system

λ_n = mean arrival rate (expected number of arrival per unit time of new customers completing service per unit time when n customers are in system)

$\rho = \lambda/\mu$ = traffic intensity (utilization factor for the service facility)

C_q = expected waiting cost in queue

C_s = expected waiting cost in the system

Assumed $\lambda = \lambda_n$, then in a steady-state queuing process $L = \lambda W$ From little's law, Hillier and Lieberman (1995);

$$L_q = \lambda W_q \ \ and \ \ W_q = W - \frac{1}{\mu} \tag{5}$$

$$\therefore W_q = E(w) = \frac{\lambda}{\mu(\mu - \lambda)}$$

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} \tag{6}$$

When s > 1
The utilization factor is

$$U_k = \begin{cases} \left(\frac{\lambda}{\mu}\right)^n & for\ n = 1,2,...s \\ \frac{\left(1/\mu\right)^s}{s!}\left(\frac{\lambda}{s\mu}\right)^{n-s} = \frac{\left(\lambda/\mu\right)^n}{s!s^{n-1}} & for\ n = s, s_{n+1} \end{cases}$$

If $N \to \infty$ and $\lambda < s\mu$, so that $\rho = \left(\lambda/s\mu\right) < 1$, then

P_o =

$$= \frac{1}{1 + \sum_{n=1}^{s-1} \frac{\left(\lambda/\mu\right)^n}{n!} + \frac{\left(\lambda/\mu\right)^s}{s!} \sum_{n=1}^{\infty} \left(\lambda/s\mu\right)^{n-s}}$$

$$= \frac{1}{\sum_{n=0}^{s-1} \frac{\left(\lambda/\mu\right)^n}{n!} + \frac{\left(\lambda/\mu\right)^s}{s!} \frac{1}{\left(1 - \lambda/s\mu\right)}}$$

$$= \left[\sum_{n=0}^{s=1} \frac{1}{n!} \left(\lambda/\mu\right)^n + \frac{1}{s!} \left(\lambda/\mu\right)^s \frac{s\mu}{s\mu - \lambda} \right]^{-1} \tag{7}$$

and P_n =

$$\left\{ \begin{array}{ll} \frac{\left(\lambda/\mu\right)^n}{n!} P_o & if \quad 0 \le n \le s \\ \frac{\left(\lambda/\mu\right)^n}{s!s^{n-s}} P_o & if \quad n \ge s \end{array} \right\} \ldots \tag{9}$$

Furthermore, L_q = $Lq = \sum_{n=s}^{\infty}(n-s)P_n$

$$= \sum_{j=o}^{\infty} j P_{s+j} = \sum_{j=0}^{\infty} j \frac{\left(\lambda/\mu\right)^s}{s!} \rho^j P_0 = P_0 \frac{\left(\lambda/\mu\right)^s}{s!} \rho \frac{d}{d\rho} \left(\sum_{j=o}^{\infty} \rho^j \right)$$

$$= P_0 \frac{\left(\lambda/\mu\right)^s}{s!} \rho \frac{d\left(\frac{1}{1-\rho}\right)}{d\rho} = P_0 \frac{\left(\lambda/\mu\right)^s \rho}{s!(1-\rho)^2}$$

$$W_q = \frac{Lq}{\lambda} \tag{9}$$

$$C_q = W_q C_w \quad (10)$$

$$W_s = W_q + {}^1/_\lambda = \frac{L_q}{\lambda} + {}^1/_\mu \quad (11)$$

$$C_s = W_s C_w \quad (12)$$

$$L_s = \lambda\,(W_q + {}^1/_\mu) = L_q + {}^\lambda/_\mu \quad (13)$$

$$P\{\omega > t\} = e^{-\mu t}\left[1 + \frac{P_0 \left(\frac{\lambda}{\mu}\right)^s}{S!(1-\rho)}\left(\frac{1 - e^{-\mu t\left(s-1-{}^\lambda/_\mu\right)}}{s-1-{}^\lambda/_\mu}\right)\right] \text{ and } P\{\omega_q > t\} = \left(1 - P\{\omega_q = 0\}\right)e^{-\mu s(1-\rho)t}, \text{ where } P\{\omega_q = 0\} = \sum_{n=0}^{s-1} P_n$$

3.1. Waiting time measure and waiting time cost for M/M/2: ∞/NPP

Let S (=2) = number of servers, μ = mean service rate per busy server, k = the number of priority classes;

$$\lambda = \sum_{i=1}^{N} \lambda_i \quad (14)$$

And

$$r = {}^\lambda/_\mu$$

This result assumes that $\sum_{i=1}^{N} \lambda_i < s\mu$; see Hillier and Lieberman (1995). Let w_{sk} be the steady state expected waiting time in the system (including service time) for a member of priority class k.

Then, wsk = $\frac{1}{AB_{k-1}B_K} + \frac{1}{\mu}$, k = 1, 2,… N (15)

where

$$A = s!\frac{s\mu - \lambda}{r^s}\sum_{j=0}^{s-1}\frac{r^j}{j!} + s\mu;\ j = 0,1,2,...,s-1 \quad (16)$$

$B_0 = 1$

$$B_k = 1 - \frac{\sum_{i=1}^{k}\lambda}{s\mu} \quad \text{for k = 1, 2,… N} \quad (17)$$

$$L_{sk} = \lambda_k w_{sk} = \lambda_k\left(\frac{1}{AB_{k-1}B_k} + \frac{1}{\mu}\right);$$

And the estimated waiting time cost for this class is given by

$$W_c = W_q\, C_w \quad (18)$$

Where

$$W_q = \frac{1}{AB_{k-1}B_k} + \frac{1}{\mu} \ \ and \ \ L_q = \lambda_k W_q = \frac{\lambda}{AB_{k-1}B_k}$$

3.1.1. Minimum cost service rate (μ^*) for the NPP queuing model

The waiting time cost

Wc = CwLs = Cw

$$\left(\lambda_k W_{sk} + \frac{1}{\mu}\right) = C_w\left[\lambda_k\left(\frac{1}{AB_{k-1}B_k} + \frac{1}{\mu}\right)\right]$$

And the expected service cost S_c per unit time is; $S_c = C_t\mu$,
The total cost;

$$C_{NPP} = W_c + S_c = C_w\left[\lambda_k\left(\frac{1}{AB_{k-1}B_k} + \frac{1}{\mu}\right)\right] + C_t\,\mu \quad (19)$$

$$C_{NPP} = C_w\lambda_k\left[\frac{1}{\left(s!\frac{s\mu-\lambda}{r^s}\sum_{j=0}^{s-1}\frac{r^j}{j!} + s\mu\right)\left(1-\frac{\sum_{i=1}^{k-1}\lambda_i}{s\mu}\right)\left(1-\frac{\sum_{i=1}^{k}\lambda_i}{s\mu}\right)} + \frac{1}{\mu}\right] + C_t\mu$$

Let $C_w\lambda = A$; s! =B; $\sum_{j=0}^{s-1}\frac{r^j}{j!} = D$; $\sum_{i=1}^{k-1}\lambda_i = E$; $\sum_{i=1}^{k}\lambda_i = F$

$$C_{NPP} = A\left[\frac{1}{\left(BD\left(\frac{s\mu-\lambda}{r^s}\right) + s\mu\right)\left(1-\frac{E}{s\mu}\right)\left(1-\frac{F}{s\mu}\right)} + \frac{1}{\mu}\right] + C_t\mu$$

$$= A\left[\frac{1}{\frac{BDs\mu}{r^s} + s\mu + \frac{EF}{s\mu} - \left(F + E + \frac{BD\lambda}{r^s}\right)} + \frac{1}{\mu}\right] + C_t\mu$$

Let also $H = \left(F + E + \frac{BD\lambda}{r^s}\right)$; $G = \left(\frac{BD}{r^2}\right)$ and $P = (EF)$

$$C_{NPP} = A\left[\frac{1}{Gs\mu + s\mu + \frac{P}{s\mu} - H} + \frac{1}{\mu}\right] + C_t\mu = A\left[\frac{1}{\left(\frac{G(s\mu)^2 + (s\mu)^2 + P - Hs\mu}{s\mu}\right)} + \frac{1}{\mu}\right] + C_t\mu$$

$$= \frac{As\mu}{(s\mu)^2[G+1] - Hs\mu + P} + \frac{A}{\mu} + C_t\mu \quad (20)$$

Theorem 1: Suppose $\tau(\mu)$ is the total expected waiting cost per unit in a queue, we assumed that the expected service cost per unit time in a queue is continuous in the interval $\chi_1 \le \mu \le \chi_2$ and is differentiable with relative minimum cost at μ^*. Then $\chi_1 \le \mu^* \le \chi_2$ if $\tau'(\mu^*) = 0$.
Proof: consider any interval $\boldsymbol{\chi_1}$ and $\boldsymbol{\chi_2}$ in the queue and k an increment, we have

$$\tau'(\mu^*) = \lim_{k\to 0}\left(\frac{\tau(\mu^*+k) - \tau(\mu^*)}{k}\right) \quad (21)$$

If (21) exists as a definite number, $\tau'(\mu^*) \le \tau(\mu^* + k)\ for\ all$ k sufficiently close to zero,

Then,

$$\frac{\tau(\mu^*+k) - \tau(\mu^*)}{k} \ge 0 \text{ if k > 0 and } \frac{\tau(\mu^*+k) - \tau(\mu^*)}{k} \le 0 \text{ if k < 0} \quad (22)$$

Thus $\tau'(\mu^*)$ in (21) gives the limit as k→0 through positive values as $\tau'(\mu^*) \ge 0$ and through negative values as $\tau'(\mu^*) \le 0$ (23)

Therefore, to satisfy both inequalities in (23), we have $\boldsymbol{\tau'(\mu^*)} = 0$
From theorem 1;

$$\frac{\partial C_{NPP}}{\partial \mu} = \frac{PAs - QAs^3\mu^2}{\left(Q(s\mu)^2 - Hs\mu + P\right)^2} - \frac{A}{\mu^2} + C_t = 0$$

$$\frac{(PAs - QAs^3\mu^2)\mu^2 - A\left((Qs^2\mu^2 - Hs\mu + P)^2\right) + C_t\left((Qs^2\mu^2 - Hs\mu + P)^2\right)\mu^2}{(Qs^2\mu^2 - Hs\mu + P)^2\mu^2} = 0$$

$$(PAs - QAs^3\mu^2)\mu^2 - A\left((Qs^2\mu^2 - Hs\mu + P)^2\right) + C_t\left((Qs^2\mu^2 - Hs\mu + P)^2\right)\mu^2 = 0$$

$$PAs\mu^2 - QAs^3\mu^4 - A(Q^2s^4\mu^2 - 2HQs^3\mu^3 + 2PQs^2\mu^2 - 2HPs\mu + H^2s^2\mu^2 + P^2)$$

$$+ C_t(Q^2s^4\mu^6 - 2HQs^3\mu^5 + 2PQs^2\mu^4 - 2HPs\mu^3 + H^2s^2\mu^4 + P^2\mu^2) = 0$$

$$(Q^2s^4C_t)\mu^6 - (2HQs^3C_t)\mu^5 + (2PQs^2C_t - AQs^3 - AQ^2s^4 + H^2s^2 + P^2C_t)\mu^4$$

$$+ (2AHQs^3 - 2HPsC_t)\mu^3 + (APs - 2APQs^2 - AH^2s^2)\mu^2 + (2AHPs)\mu - AP^2 = 0 \quad (25)$$

The resulting polynomial in (25) is a function of μ which can be solved for optimum service rate μ^* that minimizes the total cost of providing service and waiting for that service in a NPP queuing model by substituting the values of the constants A, H, Q, P, s, and C_t for a given queuing process.

3.2. Waiting time measure and waiting time cost for preemptive priority model: MM/2: ∞/PP

If preemption is allowed under the condition in NPP, the total expected waiting time in the system (including the total service time) will change to:

$$w_{sk} = \frac{1/\mu}{w_{sk} = B_{k-1}B_k} \text{ for } k = 1, 2, \ldots, N \text{ and } s = 1 \quad (26)$$

$L_k = \lambda_k w_{sk}$ for k =1, 2… N.
When $s > 1, w_k$ can be calculated by an iterative procedure.
To determine the expected waiting time in the queue (excluding service time) for priority class k, we have;

$$W_q = W_k - 1/\mu \quad (27)$$

Also, if $s > 1$, W_k can be calculated by an iterative procedure. The associated expected waiting time cost is; $w_q C_w$.

3.2.1. Minimum cost service rate (μ^*) for PP queuing model

The waiting time cost is;

$$w_c = C_w L_s = C_w\left(\lambda\left(W_q + \frac{1}{\mu}\right)\right) = C_w\lambda\left[\left(\frac{\lambda}{\mu(\mu-\lambda)}\right) + \frac{1}{\mu}\right] \quad (28)$$

where $W_q = w_{sk} - \frac{1}{\mu}$
Note that preemption does not affect the service process in any way because of lack of memory property of exponential distribution. Hence, the expected total service time for any customer is still $\frac{1}{\mu}$. and the expected service cost S_c per unit time is; $S_c = C_t\mu$
The total cost

$$\frac{\partial C_{PP}}{\partial \mu} = C_w\left[\frac{\lambda}{\mu^2 - \mu\lambda} - \frac{(\lambda^2 + \lambda\mu - \lambda^2)(2\mu - \lambda)}{(\mu^2 - \mu\lambda)^2}\right] + C_t = 0$$

$$C_{pp} = w_c + S_c = C_w\left(\frac{\lambda^2}{\mu(\mu-\lambda)} + \frac{\lambda}{\mu}\right) + C_t\mu = C_w\left(\frac{\lambda^2 + \lambda\mu - \lambda^2}{\mu(\mu-\lambda)}\right) + C_t\mu \quad (29)$$

Similarly, from theorem 1:

$$C_{PP} = C_w\left[\lambda(\mu^2 - \mu\lambda) - (\lambda^2 + \lambda\mu - \lambda^2)\right] + C_tC_w(\mu^2 - \mu\lambda)^2 = 0 \quad (30)$$

$$C_t\mu^4 - (2\lambda C_t)\mu^3 + (\lambda^2C_t + C_w\lambda)\mu^2 - 2C_w\lambda\mu = 0 \quad (31)$$

Also, the resulting polynomial equation in (31) can be solved for real values of μ^* as the optimum service rate which minimizes the total cost of providing service and waiting for service in PP queuing model by substituting the constants C_t and λ.

4. Application

We consider queuing data on emergency consulting unit of University of Uyo Teaching Hospital in Nigeria. The following parameters were obtained thereof; λ = 2.76; s = 2, C_t = 17.36; r = 1.38 and C_w = 4.37. The associated cost component was obtained as the mean of the salaries of patients/patients's benefactors. By substituting these parameters in the results in (25) and (31) we obtained the respective service rate equations of the non-preemptive and preemptive priority queuing models as follows;

$$N_{pp}: \quad 3{,}402.56\mu^6 - 4{,}928\mu^5 - 1{,}656.47\mu^4 + 2{,}784.31\mu^3 - 815.21\mu^2 + 445.09\mu - 39.95 = 0 \quad (32)$$

$$P_p: \quad 17.36\mu^4 - 1.736\mu^3 + 0.2619\mu^2 - 0.437\mu = 0 \quad (33)$$

The solution of (32) yields the following rational values as roots of μ^* for the Nonpreemptive priority queuing model; 1. 4447, -0.8607, 0.7011 and 0.1029. The optimum cost service rate which yield the minimum service rate function that minimizes the waiting time cost of patients and the service time cost of doctors is $\mu^* = 0.7011$.
Similarly, the solution of (33) yields a rational value of $\mu^* = -0.2786 \approx 0$ as the optimum cost service rate that yields a minimum service rate function for the Preemptive priority queuing model that minimizes the waiting time cost of patients and the service time cost of doctors.

5. Conclusion

The total expected waiting time cost per unit time and the expected service time cost per unit time for priority queuing models: M/M/2: ∞/NPP and M/M/2: ∞/PP were formulated and optimized to obtain optimum service rate that both minimizes the expected waiting time cost and the expected service time cost per unit time for the models. The resulting polynomial equations were solved to obtain real values of μ as the optimum cost service rate (μ^*) that minimizes the total cost of time of providing service by doctors and waiting for service by patients at the emergency unit in hospital. The optimal value of $\mu^* = 0.7011$ provide the optimum service time that guarantees minimum waiting time cost and service time cost for the non-preemptive priority model while $\mu^* = -0.2786 \approx 0$ provides an optimum service time that guarantees a minimum waiting time cost and service time cost using the Preemptive priority model. This, however suggest an overstretched of available medical facilities.

References

[1] Bedford, A and Zeephongsekul, P. On a Dual Queuing System with Preemptive Priority Service Discipline. European Journal of Operational Research, 2003.

[2] Blake, J. T. and Carter, M. W. An Analysis of Emergency Room Waiting Time Issues via Computer simulation. Information System and Operational Research (INFORM) 34(4):263-273, 1996. https://doi.org/10.1080/03155986.1996.11732308.

[3] Hagen, M. S., Jopling, J. K., Buchmen, T. G and Lee, E. K. Priority Queuing Models for Hospital Intensive Care Units and Impact to Severe Case Patients. AMIA Annual Symposium Proceedings; 841-850, 2013.

[4] Hillier, S., Frederick and Lieberman, J. Gerald. Introduction to Operations Research. Sixth edition McGraw-Hill, Inc. Singapore, 1995.

[5] Ke, P., Li, Y. X. and Ni, F. Multiple Services Scheduling with Priority-Queuing Model. IJCIT.org or ijcit.uap-bd.edu (Online) 3(1), 2012.
[6] Li et al. A Priority-Queuing Model for Heterogeneous Traffic Scheduling in Inter-Vehicular Communication. Information Technology Journal 12: 419-423, 2013. https://doi.org/10.3923/itj.2013.419.423.
[7] Udoh, N. S. Priority Queuing Models for Emergency Cases in the Hospitals. Journal of National Association of Mathematical Physics, 16(2): 191-196, 2010.
[8] Siddarthan, K., Jones, W. J. and Johnson, J. A. A priority Queuing Model to Reduce Waiting Times in Emergency Care. International Journal of Healthcare Quality Assurance, 9(5): 10-16, 2006. https://doi.org/10.1108/09526869610124993.
[9] Taha, H. A. Operations Research: An Introduction, 7th edition. Pearson Education Inc., India, 2007.