

Transient Analysis of Batch Service Queue with Second Optional Service and Reneging

P. Vijaya Laxmi¹, Andwilile Abrahamu George²

¹Associate Professor, Dept. of Applied Mathematics, Andhra University, Visakhapatnam 530003, India

²PhD Scholar, Dept. of Applied Mathematics, Andhra University, Visakhapatnam 530003, India
gandwilile@gmail.com²

Abstract

This paper investigates a single server batch service queueing system with second optional service and reneging. The transient state probabilities of the queueing model are computed by using fourth order Runge Kutta method. Cost analysis is performed to determine the optimal values of the service rates of First Essential Service (FES) and Second Optional Service (SOS) simultaneously at the minimum total expected cost per unit time. Some important performance measures and numerical illustrations are provided in order to show the managerial intuitions of the model.

Keywords: Transient analysis, Queue, First essential service, Second optional service, Reneging

1.Introduction

In queueing theory, a service process is among the important characteristics of a queueing system. Recently, most of the service systems adopted the provision of more than one service. For instance, in transport industry passengers get transport service from buses/airline and other services like food, soft drinks, etc. However, majority of them may prefer only main service while few of them may demand an optional service. Complete details of these situations have been presented by [1-5] and considered an $M/G/1$ queue with second optional service which assumed to follow exponential distribution while first essential service is a general distribution. [6-8] considered both SOS and FES having a general service time distribution.

Balking and reneging are two impatient acts where a customer tends to leave without receiving the service due to longer waiting. Queueing problems with balking and reneging was studied by [2]. A single server queueing model with impatience has been presented by [9,10], where the customers lose

patience if the wait is more than the threshold they fixed. Later on, [11,12] studied a finite buffer multiple working vacation queues with balking, reneging and vacation interruption under N-Policy and obtained the solution for the steady state probabilities using recursive.

The study of the transient solution of the performance measures of the queue system has been analysed by [14-18] etc. A time dependent solution of a single server queueing system with a Poisson input process has been considered by [19,20]. The transient behaviour of the infinite capacity $M/G/1$ model with batch arrivals and server vacations has been discussed by [21]. The transient behaviour analysis of an $M/M/1/N$ queue with working breakdowns and server vacations has been studied by [22].

From the above literature study, most of the heterogeneous server queueing systems are analysed in steady state. However, the steady state measures do not explain the real picture of the system behaviour, since they ignore the transient and the initial behaviour of a system [18]. In most of the practical applications, the state of the system experiences a change and such changes can be comprehended by the transient behaviour. Therefore, this article aims to study a $M/M^{[b]}/1/$

N queue with SOS and reneging in transient state. The transient probabilities of the queueing model are obtained using fourth order Runge-Kutta method.

2. Model Description

We consider an $M/M^{[b]}/1/N$ queueing model with SOS and reneging customers in which the customers arrive according to a Poisson process with mean rate λ . The service times distribution of both FES and SOS are exponential with mean rates μ_1 and μ_2 , respectively and the services are given in batches of size not more than b such that if the server finds the customers less or equal to b in the waiting queue, the server takes all of them in the batch for service, but if the server finds the customers more than b waiting in the queue, then the he takes a batch of size b while others remain waiting in the queue. The FES is required by all arriving customers and after completing FES, they may opt SOS with probability r or may leave the system with probability $(1 - r)$. The capacity of the system is finite (N). A customer upon joining the queueing system and waiting in queue for some time, may get impatient and renege from the queue if his waiting time is beyond the expected time limit. Reneging times follow an exponential distribution with rate α .

2.1 Formulation of Mathematical Model

The $M/M^{[b]}/1/N$ queue with SOS and reneging can be modeled by a two-dimensional continuous time Markov process $\{(N(t), J(t)); t \geq 0\}$, where, $N(t)$ is the number of customers in the queue, and $J(t)$ is the server state with

$$J(t) = \begin{cases} 1, & \text{if the server is providing FES} \\ 2, & \text{if the server is providing SOS} \end{cases}$$

The state space of the Markov process is given by

$$\Omega = \{(n, i); n \geq 0; i = 1, 2\}$$

The transient probabilities are defined as

$$P_{n,i}(t) = Pr\{N(t) = n, J(t) = i; 0 \leq n \leq N, i = 1, 2\}$$

Here, $P_{n,i}(t)$ is the transient probability that there are n customers in the queue at time t and the server is providing FES (SOS) service, and $Q(t)$ is the probability that the queue is empty at time t and the server is idle. Using Markov theory, the differential-difference equations of the model are as follows:

$$Q'(t) = -\lambda Q(t) + (1 - r)\mu_1 P_{0,1}(t) + \mu_2 P_{0,2}(t) \dots \dots \dots 1$$

$$P'_{0,1}(t) = -(\lambda + \mu_1)P_{0,1}(t) + (1 - r)\mu_1 \sum_{i=1}^b P_{i,1}(t) + \mu_2 \sum_{i=1}^b P_{i,2}(t) + \alpha P_{1,1}(t) + \lambda Q(t) \dots \dots \dots 2$$

$$P'_{n,1}(t) = -(\lambda + \mu_1 + \alpha)P_{n,1}(t) + (1 - r)\mu_1 P_{n+b,1}(t) + \mu_2 P_{n+b,2}(t) + \alpha P_{n+1,1}(t) + \lambda P_{n-1,1}(t), \quad 1 \leq n \leq N - b \dots \dots \dots 3$$

$$P'_{n,1}(t) = -(\lambda + \mu_1 + \alpha)P_{n,1}(t) + \alpha P_{n+1,1}(t) + \lambda P_{n-1,1}(t), \quad N - b + 1 \leq n \leq N - 1 \dots \dots \dots 4$$

$$P'_{N,1}(t) = -(\mu_1 + \alpha)P_{N,1}(t) + \lambda P_{N-1,1}(t) \dots \dots \dots 5$$

$$P'_{0,2}(t) = -(\lambda + \mu_2)P_{0,2}(t) + r\mu_1 P_{0,1}(t) + \alpha P_{1,2}(t) \dots \dots \dots 6$$

$$P'_{n,2}(t) = -(\lambda + \mu_2 + \alpha)P_{n,2}(t) + \lambda P_{n-1,1}(t) + r\mu_1 P_{n,1}(t) + \alpha P_{n+1,2}(t), \quad 1 \leq n \leq N - 1 \dots \dots \dots 7$$

$$P'_{N,2}(t) = -(\mu_2 + \alpha)P_{N,2}(t) + \lambda P_{N-1,1}(t) + r\mu_1 P_{N,1}(t), \dots \dots \dots 8$$

The above is an implicit set of differential equations whose analytical solution is a difficult task. However, we shall use the numerical technique to derive the approximate solutions assuming some initial conditions.

3. Transient Solution of the Model

In this section, we apply the fourth order Runge-Kutta method (R-K-4) to obtain the numerical solutions to the above equations. We assume that the system is empty at the starting point when the server is idle and the probability of the state of the initial system at time $t = 0$ are given as follows:

$$Q(0) = 1, P_{n,i}(0) = 0, i = 1, 2, \quad 0 \leq n \leq N$$

The set of differential difference equations (1) to (8) can be expressed as:

$$\frac{dx}{dt} = f(t, x); \quad x(t_0) = x_0$$

The formula used for Runge-Kutta method of fourth order is given by:

$$x_j + 1 = x_j + \left[\frac{k_1 + 2k_2 + 2k_3 + k_4}{6} \right] \Delta t$$

where; Δt is the step size, and $0 \leq j \leq N - 1$

$$k_1 = f(t_j, x_j) \Delta t$$

$$k_2 = f\left(t_j + \frac{\Delta t}{2}, x_j + \frac{k_1}{2}\right) \Delta t$$

$$k_3 = f\left(t_j + \frac{\Delta t}{2}, x_j + \frac{k_2}{2}\right) \Delta t$$

$$k_4 = f(t_j + \Delta t, x_j + k_3) \Delta t$$

In our model, we consider x as $P_{n,i}(t)$ and $f(t, x)$ as $\frac{dP_{n,i}(t)}{dt}$

4. Performance Measures

In this section, we present the performance measures for anticipating the system behaviour. The various performance measures are as given below.

- (a) Expected number of customers in the queue during FES, SOS and the overall queue length are given, respectively by

$$L_q FES(t) = \sum_{n=0}^N n P_{n,1}(t),$$

$$L_q SOS(t) = \sum_{n=0}^N n P_{n,2}(t)$$

$$L_q(t) = L_q FES(t) + L_q SOS(t)$$

- (b) The average renegeing rate is given by

$$R.R(t) = \sum_{n=0}^N \alpha P_{n,1}(t) + \sum_{n=0}^N \alpha P_{n,2}(t)$$

- (c) The blocking probability and the effective arrival rate are, respectively given by:

$$P_{block}(t) = P_{N,1}(t) + P_{N,2}(t),$$

$$\lambda'(t) = \lambda(1 - P_{block}(t))$$

$$= \lambda \left[Q(t) + \sum_{n=0}^{N-1} (P_{n,1}(t) + P_{n,2}(t)) \right]$$

- (d) As a result of renegeing, the Average Load (AL) on the server is given by

$$AL(t) = \lambda'(t) - R.R(t)$$

Using Little's Law, we can determine $W_q(t)$ as follows:

$$W_q(t) = \frac{L_q(t)}{AL(t)}$$

4.1 Cost Analysis

In this section, we develop the total expected cost function per unit time based on the steady state system performance measure when $t \rightarrow \infty$. Our objective is to determine the optimal value of service rates of FES and SOS i.e., $(\mu_1$ and $\mu_2)$, respectively, so that the total cost function is minimized. Let us define the following cost elements:

$C_q \equiv$ Cost per unit time per customer present in the queue,

$C_1 \equiv$ Cost per unit time when the server is serving in FES, $C_2 \equiv$ Cost per unit time when the server is serving in SOS, $C_r \equiv$ Cost per unit time per lost customer due to renegeing.

From the definitions of each cost element listed above, the steady state expected cost function $f(\mu_1, \mu_2)$ per customer per unit time can be written mathematically as:

$$f(\mu_1, \mu_2) = C_q L_q + C_1 \mu_1 + C_2 \mu_2 + C_r E[R].$$

The above expected cost function is highly complex and non-linear in terms of $f(\mu_1, \mu_2)$ and the derivatives of the cost function are not easily available. Therefore, we solve the above optimization problem using Quadratic Fit Search Method (QFSM), since this method is highly suitable for the non-differentiable objective functions.

Given the point $\xi = (\xi_1, \xi_2)$, through QFSM, we fit a quadratic function that corresponds to functional values having a unique minimum, $\xi^q = (\xi_1^q, \xi_2^q)$, for the given objective function $f(\xi) = f(\xi_1, \xi_2)$. Quadratic fit makes use of this approximation to improvise the current 3-point pattern by replacing one of its points with approximate optimum ξ^q . The unique optimum ξ^q of the quadratic function agreeing with $f(\xi)$ at 3-point operation (ξ^l, ξ^m, ξ^h) occurs at

$$\xi^q \cong \frac{1}{2} \frac{[f(\xi^l)[s^m - s^h] + f(\xi^m)[s^h - s^l] + f(\xi^h)[s^l - s^m]]}{[f(\xi^l)[\xi^m - \xi^h] + f(\xi^m)[\xi^h - \xi^l] + f(\xi^h)[\xi^l - \xi^m]}$$

Where, $s^l = (\xi^l)^2, s^m = (\xi^m)^2, s^h = (\xi^h)^2$. For a detailed algorithm, one may refer Rardin (1997).

5. Numerical Investigation

In this section we demonstrate the applicability of the Runge-Kutta method in transient state models using MATLAB software. The numerical computations are presented in the form of tables and graphs. The values of the system parameters

are assumed to be $\lambda = 2, \mu_1 = 3, \mu_2 = 2.5, \alpha = 0.1, r = 0.36, N = 10, b = 5$.

As observed in Table 1, the transient probabilities obtained by Runge-Kutta method tend to the corresponding steady state probabilities for large values of t . This confirms correctness, accuracy and efficiency of the model.

Table 3: Optimum cost for different b and λ ($N = 10, r = 0.3$) at steady state for $t \rightarrow \infty$

Table 1: Probabilities using Runge Kutta method

$P_{n,i}(t)$	Transient solution		
	$t = 3$	$t = 10$	$t \rightarrow \infty$
$Q(t)$	0.3471	0.3457	0.3457
$P_{0,1}(t)$	0.2733	0.2731	0.2731
$P_{0,2}(t)$	0.0668	0.0668	0.0668
$P_{1,1}(t)$	0.1096	0.1097	0.1097
$P_{1,2}(t)$	0.0555	0.0555	0.0556
$P_{10,1}(t)$	0.0000	0.0000	0.0000
$P_{10,2}(t)$	0.0001	0.0000	0.0000
Total	1.0000	1.0000	1.0000

Table 2: Variation in different measures of effectiveness with the change in arrival rate (λ)

(b, λ)	$(\mu_1^*, \mu_2^*, f^*(\mu_1, \mu_2))$	L_q	$R.R$	P_{block}	AL	W_q
<u>$b = 1; \alpha = 0.3; r = 0.3$</u>						
(1, 2.0)	(5.34219, 3.13311, 38.9977)	0.607336	0.0857459	0.000493297	1.91327	0.317434
(1, 2.5)	(6.35925, 3.05648, 45.0453)	0.940152	0.110784	0.00212125	2.38391	0.394374
(1, 3.0)	(7.68885, 3.51123, 51.4935)	1.02275	0.115751	0.00279075	2.87588	0.355631
(1, 3.5)	(8.6376, 4.14159, 57.9119)	1.09593	0.120875	0.00327201	3.36767	0.325426
<u>$b = 3; \alpha = 0.5; r = 0.3$</u>						
(3, 2.0)	(2.83679, 1.80106, 27.4659)	0.940152	0.110784	0.00212125	2.38391	0.394374
(3, 2.5)	(3.63091, 2.0872, 31.3237)	0.782394	0.181567	0.00111562	2.31564	0.337873
(3, 3.0)	(4.08024, 2.1111, 34.8848)	0.988905	0.206987	0.0026694	2.785	0.355082
(3, 3.5)	(4.50026, 2.61151, 38.5909)	1.00624	0.211201	0.00254425	3.27989	0.306789

Table 3 provides the sensitivity analysis for optimum service μ_1^*, μ_2^* , optimum cost f^* and some performance measures. We observe that for fixed batch size the optimum service rates, cost, $L_q, R.R, Pblock, AL$ and W_q are increasing as the

arrival rate is increasing, which is in accordance with the practical situations. The trend is reversed when we fix λ and increase batch size value b .

Table 3 provides the sensitivity analysis for optimum service μ_1^*, μ_2^*

λ	L_q	$Pblock$	λ'	Rr	AL	W_q
1	0.1822	0.0000	0.9999	0.0413	0.9587	0.1900
2	0.5937	0.0003	1.9994	0.0654	1.9340	0.3070
3	1.1182	0.0029	2.9914	0.0783	2.9130	0.3839
4	1.7160	0.0114	3.9543	0.0850	3.8693	0.4435
5	2.3557	0.0284	4.8578	0.0879	4.7699	0.4939
6	3.0034	0.0541	5.6754	0.0882	5.5871	0.5376
7	3.6295	0.0868	6.3925	0.0869	6.3056	0.5756
8	4.2135	0.1241	7.0070	0.0844	6.9226	0.6087
9	4.7447	0.1638	7.5255	0.0813	7.4442	0.6374
10	5.2199	0.2041	7.9590	0.0778	7.8812	0.6623

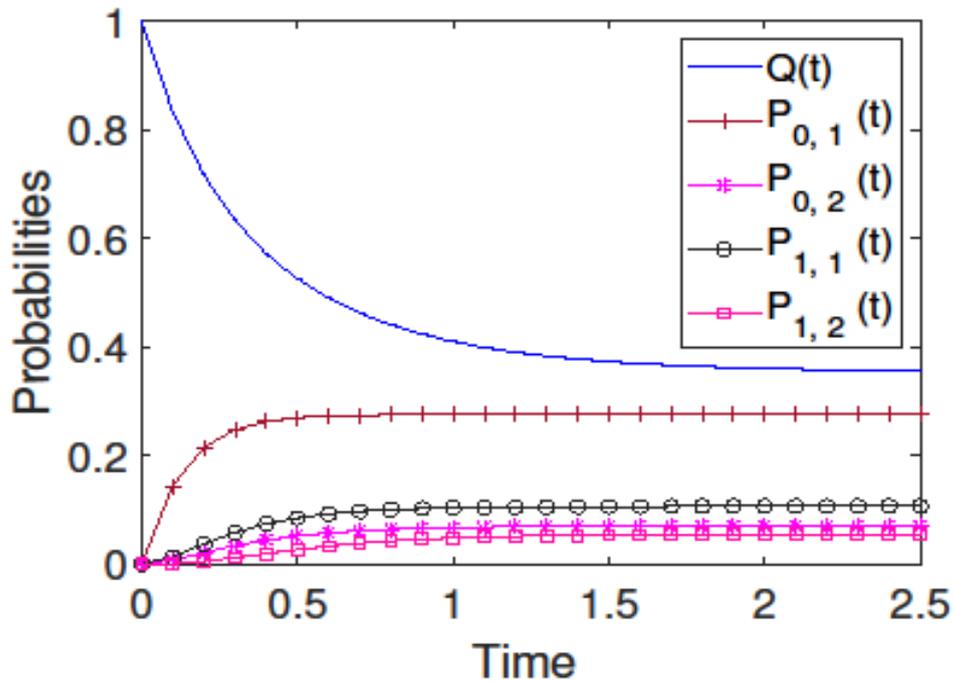


Fig.1: Transient state probability for $\lambda = 2$

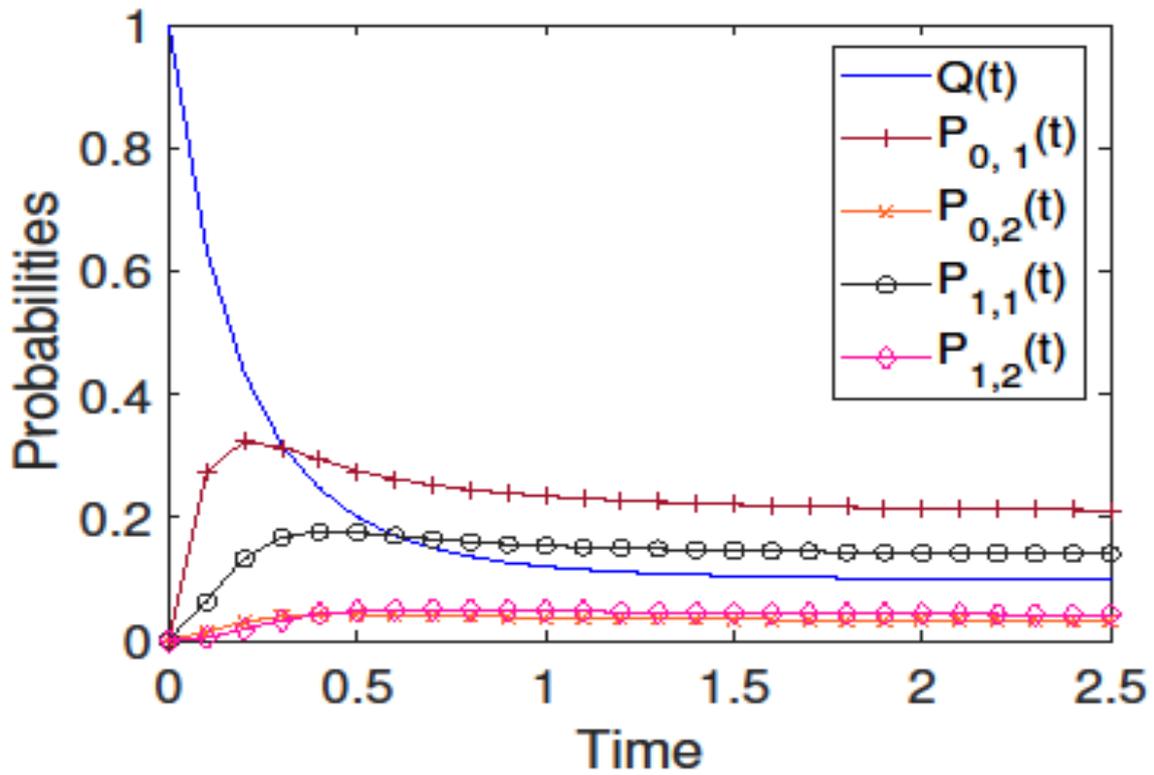


Fig.2: Transient state probability for $\lambda = 5$

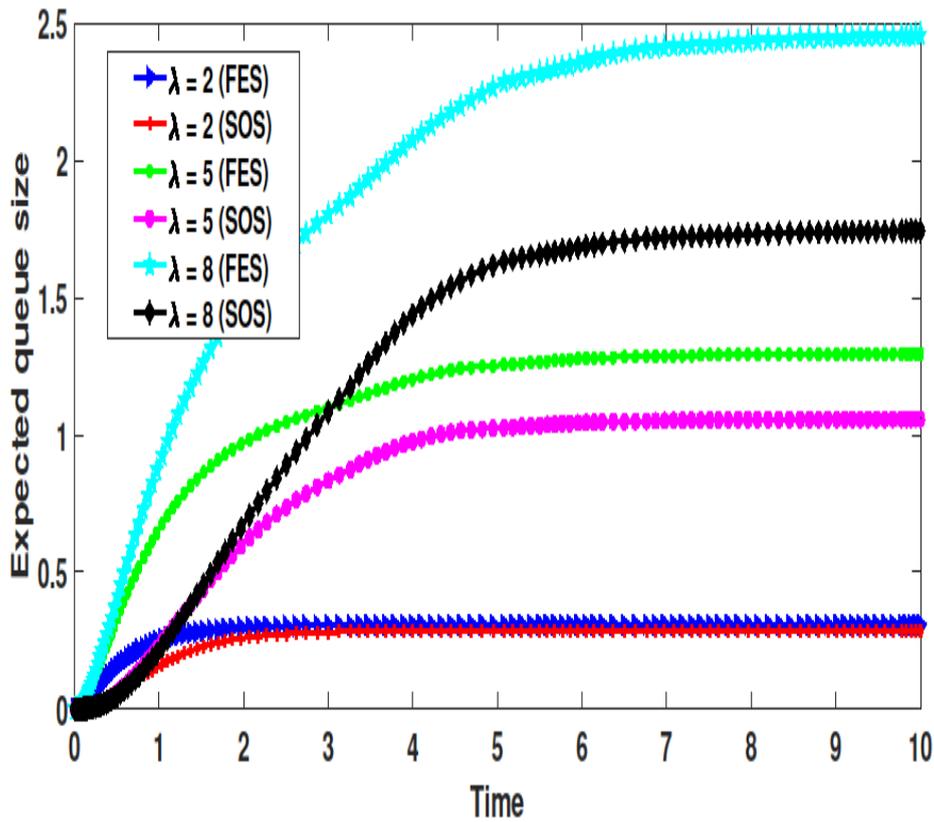


Fig.3: Expected queue size versus time

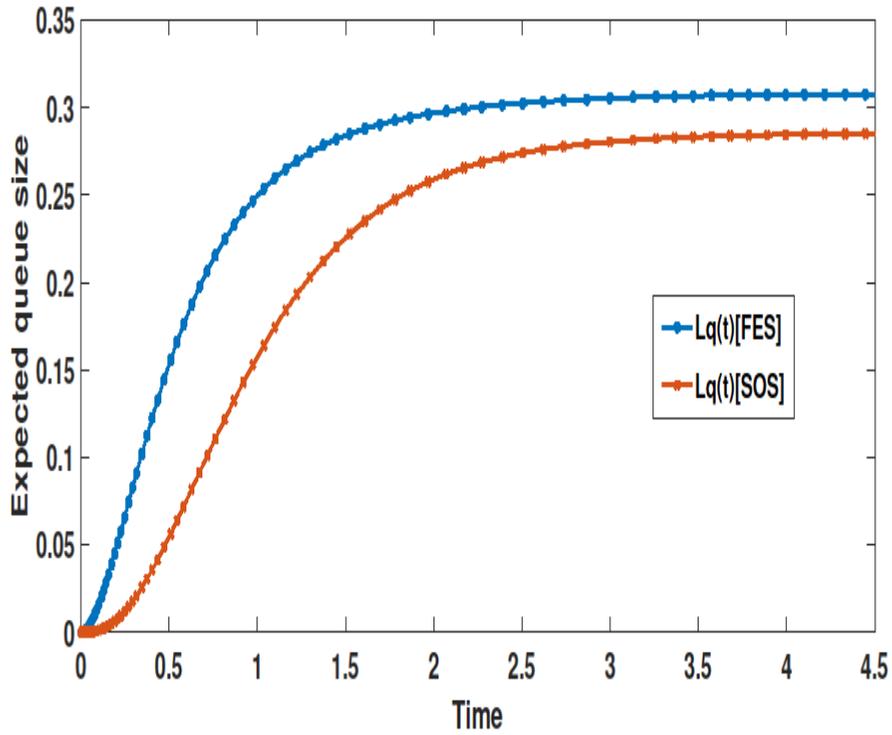


Fig.4: Expected queue size versus time

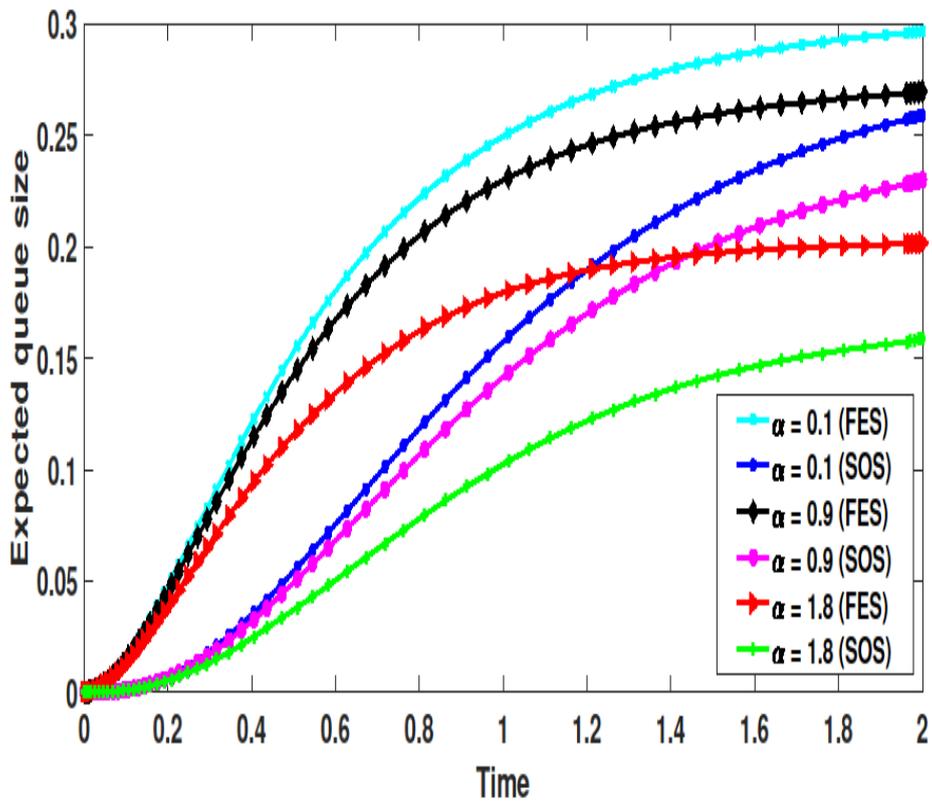


Fig.5: Effect of α on the expected queue size with respect to time

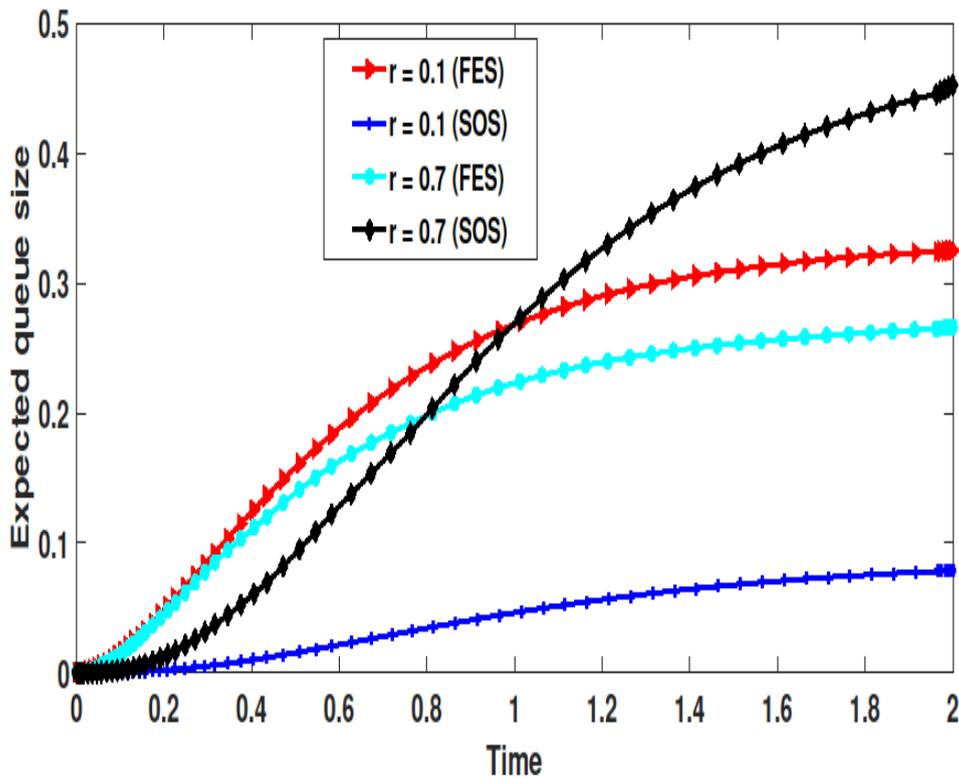


Fig.6: Effect of r on the expected queue size with respect to time

Figures 1 and 2, plots the transient state probabilities versus time for different $\lambda = 2.0$ and 5.0 , respectively with $\alpha = 0.35$. This graph is to show the sharp decrease in the idleness probability and the increase in FES probabilities as is the case when the rate of arrivals increases. Figure 3 shows the effect of arrival rate λ on the expected queue size with respect to time. We observe that the expected number of customers in the queue gradually increases as λ increases. In figure 4 one can observe that the expected queue size increases progressively from zero initially up to a certain value where it attains the steady state. In addition, we observed that the expected queue size in SOS is lower than in FES. Figures 5 and 6 demonstrates the impact of α and r , respectively on the expected number of customers in the queue versus time. We can see that as α increase the expected queue size decrease. Furthermore, we observed that initially the customers were patience with a variation of reneging rate, but as time goes on, some customers' loss patience and decide to leave the queue. In figure 6 we observe that as r increases with time the queue size in both case SOS and FES increase. In addition, when r increase the expected queue size in SOS increase,

this is due to the fact that more customers demand SOS.

6. Practical Application of the model

The model has numerous applications such as in hospital management systems, bank services, computer and communication networks, production system, etc. For example, in the hospital management systems the model can be applied to the situation where the outpatients request the appointment in the clinic system. The clinic officer monitors the length of booking windows for appointment of the outpatients, but since in most cases there is limited number of physicians/doctors, it leads to the unbalanced ratio between the number of outpatients and that of the physicians/doctors. This situation leads to an increase in the length of the booking window and brings the long waiting for appointments. Therefore, in order to shorten outpatients waiting time, we limit the length of the booking window and we assume the system has a limited slots capacity N during a limited length of booking window. The system capacity N can be divided into x equal amount of slots *b* i.e. ($b = \frac{N}{x}$). The slots b are termed as a maximum size of a batch of

the outpatients appointment and x is equal to the number of sessions/periods. In practice, the new requests for appointments come individually and the service can be provided from single patients up to the maximum batch size b of the patients. And after the first service the patients may opt for another service provided by a clinic. On the other hand, some patients cancel the appointment and those patients are termed as renegeing patients.

Conclusion

In this paper, we present $M/M^{[b]}/1/N$ queue with second optional service and renegeing in transient state. The fourth order Runge Kutta method is used to determine the transient solution of the queueing model and obtained the transient probabilities of the model. We developed the cost analysis to determine the optimal service rates of FES and SOS. Furthermore, we presented some performance measures along with the numerical results to show the effect of various system parameters on the transient behaviour of the model.

References

- [1] Al-Jararah, J and Madan, K.C. (2003). An M/G/1 queue with second optional service with general service time distribution, International Journal of Information and Management Sciences, 14, 47–56.
- [2] Ancker, C.J and Gafarian A.V. (1963). Some queueing problems with balking and renegeing, Operations Research, 11, 88–100.
- [3] Ke. J.C., Wu, C.H and Wen Lea Pearn, W.L. (2013). Capacity multi server queue with second optional service channel, International Conference on Mathematical Modelling in Physical Sciences, Conference Series.
- [4] Madhan, K.C. (2000). An M/G/1 queue with second optional service, Queueing Systems, 34, 37–46.
- [5] Madhan, K.C. (1994). An M/G/1 queueing system with additional optional service and no waiting capacity, Microelectronics and Reliability, 34, 521–527.
- [6] Manoharan, P and Sankarasasi, K. (2015). An M/G/1 renegeing queueing system with second optional service and second optional vacation, Applied Mathematical Sciences, 9, 3313–3325
- [7] Manoharan, Pand Sankarasasi, K. (2015). An M/G/1 feedback queueing system with second optional service and second optional vacation, International Journal of Fuzzy Mathematical Archieve, 8, 101–113.
- [8] Medhi, J. (2002). A single server Poisson input queue with second optional channel, Queueing Systems, 42, 239–242.
- [9] Wang, J. (2004). An M/G/1 queue with second optional service and server breakdowns, Computers and Mathematics with Applications, 47, 1713–1723.
- [10] Choudhury, A. (2008). Impatience in single server queueing model, American Journal of Mathematical and Management Sciences, 28, 177–211.
- [11] Neuts, M.F. (1967). A general class of bulk queues with poisson input, The Annals of Mathematical Statistics, 38, 759–770.
- [12] Laxmi, P. V, Goswami, V and Jyothsna, K. (2013). Optimization of balking and renegeing queue with vacation interruption under N-Policy, Journal of Optimization, Vol. 2013.
- [13] Nagarajan, R and Kurose, J. F. (1992). On defining, computing and guaranteeing quality of service in high speed networks, Proceeding of INFOCOM92, Florence, Italy, 2013, 2016–2025.
- [14] Holt, B. V and Blondia, C. (2005). Approximated transient queue length and waiting time distribution via steady state analysis, Stochastic models, 21, 725–744.
- [15] Hofkens, T, Spacy, K and Blondia, C. (2004). Transient Analysis of the DBMAP/G/1 queue with an application to the dimensioning of video playout buffer for VBR Traffic, Proceeding of Networking, Athens Greece.
- [16] Lucantoni, D. M, Choudhury, G. L, and Witt, W. (1994). The transient BMAP/P H/1 queue, Stochastic Models 10 461–478.
- [17] Bohm, W and Mohanty, S. G. (1994). Transient analysis of queues with heterogeneous arrivals, Queueing Systems, 18, 27–45.
- [18] Whitt, W. (1983). Untold horrors of the waiting room, what the equilibrium

distribution will never tell about the queue length process, *Management Science*, 29, 395–408.

- [19] Parathasavathy, P. P. (1987). A transient solution of an M/M/1 queue: A simple approach, *Advanced Applied Probability*, 19, 997–998.
- [20] Bai, Y, Song, J and Zhang, G. (2015). Design appointment in outpatient department with bulk service, *IEEE International Conference on Automation Science and Engineering (CASE)*, 746–751.
- [21] Jacob, M. J. and Mathusoodanan, T. P. (1988). Transient solution of an M/G/1 queueing system with group arrivals and vacations to the server, *OPSEARCH*, 25, 279–284.
- [22] Yang, D. Y and Ying, Y. W. (2014). Transient behaviour analysis of a finite capacity queue with working breakdowns and server vacations, *Proceedings of the International Multi Conference of Engineers and Computers Scientists, Hong Kong*, 2, 12–14.