J. Assam Sc. Soc. Vol. 57. No. 1 ISSN 0587-1921
December 2016; Pp. 90-107

MODELING CUSTOMERS' IMPATIENCE IN M/M/1 QUEUE UNDER GENERAL RENEGING DISTRIBUTION

Amit Choudhury, Pallabi Medhi
Dept. of Statistics, Gauhati University, Guwahati-781014, Assam, India
****Corresponding author***: achoudhury@rediffmail.com, pmedhi16@gmail.com
Mobile: 91-9864101261

ABSTRACT

In the analysis of queuing systems with reneging customers, it is traditionally assumed that the reneging rate remains constant irrespective of the position of the customer. In real life this is seldom true. This paper is an attempt to model such a reneging phenomenon along with balking in the M/M/1 queue. The reneging rates are a function of the position of the customer in the system. Steady state distributions along with some newly designed performance measures are analyzed. Traditional performance measures have been derived. Some new ones have been proposed. Explicit closed form expressions have been presented. A design oriented numerical example has been discussed to focus on usefulness of results derived.

Key words: Balking, Impatience, Queuing, Reneging.
2000 Mathematics Subject Classification: 60K25, 68M20, 90B22

INTRODUCTION AND REVIEW OF LITERATURE

Customers by nature are impatient entities. In our fast-paced life, one can observe that customers desiring some service from a service facility are often required to queue up. This act of queuing up induces impatience into the customer as a result of which customer may either refuse to join the queue or even if it joins, it may leave without completely receiving service. In queuing parlance, the phenomenon of customers arriving at a non-empty queuing system and leaving without joining the queue is known as balking. Haight (1957) has provided a rationale, which might influence a person to balk. It relates to the perception of the importance of being served which induces an opinion somewhere in between urgency, so that a queue of certain length will not be joined, to indifference where a non-zero queue is also joined. On the other hand, if the customer joins the queue but leaves without completely receiving service, the customer is said to have reneged. Even though one can observe reneging and balking in our day-to-day life, it is not that one can locate papers on these features of customer behavior in queuing literature very frequently.

In this paper, we shall consider balking in a manner such that any customer arriving at a non-empty queuing system will have a fixed probability of balking from it. As for modeling reneging phenomena, one approach in literature is to assume that each customers joining the system has a Markovian patience time. Second, conventionally, it has also been assumed that customers joining the system are not aware of their position on the same so that the reneging rate is state independent. However, it is our common day observation that there are systems where the customer is aware of its state in the system. For example customers queuing at the O.P.D. (out patient department) clinic of a hospital would know of their position in the queue. This invariably causes waiting customers to have higher rates of reneging in case their position in the queue is towards the end. It is not unreasonable then to expect that such customers who are positioned at a distance from the service facility have reneging rates which are higher than reneging rates of customers who are near the service facility. In other words, we assume that customers are "state aware" and in this paper we model the reneging phenomenon in such a manner that the Markovian reneging rate is a function of the state of the customer in the system. Customers at higher states will be assumed to have higher reneging rates.

Reneging are of two types-viz. reneging till beginning of service (henceforth referred to as R_BOS) and reneging till end of service (henceforth referred to as R_EOS) (Choudhury and Medhi, 2011). In R_BOS, a customer can renege only as long as it is in the queue. It cannot renege once it begins receiving service. A common example is customers queuing at OPD clinic. A customer can renege while he is waiting in queue. However once service starts i.e. examination by medical staff begins, the customer cannot leave till it is over. On the other hand in R_EOS, a customer can renege not only while waiting in queue but also while receiving service. Such a situation may occur in the processing or merchandising of perishable goods, hospital emergency room/O.T. handling critical patients etc. A critical patient may expire not only in while waiting in queue (waiting for emergency medical staff to attend to him/her) but may also expire at the O.T. table of the emergency wing of the hospital, this would be an example of reneging while service is being delivered .

The subsequent sections of this paper are structured as follows. In section 2, we formalize the model by outlining the specifics. Section 3 contains a brief review of the literature. Section 4 and section 5 contains the derivation of steady state probabilities and performance measures. We perform sensitivity analysis in section 6. A numerical example is discussed in section 7. Concluding statements are present in section 8. The appendix contains some derivation.

One of the earliest work on reneging was by Barrer (1957) where he considered deterministic reneging with single server Markovian arrival and service rates. Customers were selected randomly for service. In his subsequent work, Barrer (1957) also considered deterministic reneging (of both R_BOS and R_EOS type) in a multi server scenario with FCFS discipline. The general method of solution was extended to two related queuing problems. Another early work was by Haight (1959). Ancher and Gafarian (1963) carried out an early work on Markovian reneging with Markovian arrival and service pattern. Ghosal (1963) considered a D/G/1 model with deterministic reneging. Gavish and Schweitzer (1977) also considered a deterministic reneging model with the additional assumption that arrivals can be labeled by their service requirement before joining the queue and arriving customers are admitted only if their waiting plus service time do not exceed some fixed amount. This assumption is met in communication systems. Kok and Tijms (1985) considered a single server queuing system where a customer becomes a lost customer when its service has not begun within a fixed time.

Haghighi et al. (1986) considered a Markovian multi server queuing model with balking as well as reneging. Each customer had a balking probability, which was independent of the state of the system. Reneging discipline considered by them was R_BOS. Liu et al. (1987) considered an infinite server Markovian queuing system with reneging of type R_BOS. Brandt et al. (1999) considered a S-server system with two FCFS queues, where the arrival rates at the queues and the service may depend on number of customers 'n' being in service or in the first queue, but the service rate was assumed to be constant for n>s. Customers in the first queue were assumed impatient customers with deterministic reneging. Boots and Tijms (1999) considered an M/M/C queue in which a customer leaves the system when its service has not begun within a fixed interval after its arrival. In this paper, they have given the probabilistic proof of 'loss probability', which was expressed in a simple formula involving the waiting time probabilities in the standard M/M/C queue.

Ke and Wang (1999) considered the machine repair problem in which failed machines balk with probability (1-b) and renege according to a negative exponential distribution. Bae et al. (2001) considered an M/G/1 queue with deterministic reneging. They derived the complete formula of the limiting distribution of the virtual waiting time explicitly. Choi et al. (2001) introduced a simple approach for the analysis of the M/M/C queue with a single class of customers and constant patience time by finding simple Markov process. Applying this approach, they analyzed the M/M/1 queue with two classes of customer in which class 1 customer have impatience of constant duration and class 2 customers have no impatience and lower priority than class 1 customers. Performance measures of both M/M/C and M/M/1 queues were discussed. Zhang et al. (2005) considered an M/M/1/N framework with Markovian reneging where they derived the steady state probabilities and formulated a cost model. Some performance measures were also discussed. A numerical example was discussed to demonstrate how the various parameters of the cost model influence the optimal service rates of the system. Choudhury (2008) analyzed a single server Markovian queuing system with the added complexity of customers who are prone to giving up whenever its waiting time is larger than a random threshold-his patience time. He assumed that these individual patience times were independent and identically distributed exponential random variables. Reneging till beginning of service was considered. A detailed and lucid derivation of the distribution of virtual waiting time in the system was presented. Some performance measures were also presented. El-Paoumy (2008) also derived the analytical solution of $M^x/M/2/N$ queue for batch arrival system with Markovian reneging. In this paper, the steady state probabilities and some performance measures of effectiveness were derived in explicit forms. Another paper on Markovian reneging was by Altman and Yechiali (2008).

Xiong et al. (2008) considered a single server queue with a deterministic reneging time motivated by the timeout mechanism used in application servers in distributed computing environments. They derived the probability generating function of number of customers present in the system and some performance measures were discussed. Jouini et al. (2009) considered two multi-class call center models with and without reneging. Choudhury (2009) considered a single server finite buffer queuing system (M/M/1/K) assuming reneging customers. Both rules of reneging (R_BOS and R_EOS) were considered and various performance measures presented under both rules of reneging. Xiong and Altiok (2009) studied a multi server queue with Poisson arrivals general service time distribution and deterministic reneging times. Via approximations, they provided the expression for mean waiting time. Other attempts at modeling reneging phenomenon include those by Baccelli et al (1984), Martin and Artalejo (1995), Shawky (1997), Choi, Kim and Zhu (2004), and Singh et al (2007), El-Sherbiny (2008) and El-Paoumy and Ismail (2009) etc.

An early work on balking was by Haight (1957). Another work using the concepts of balking and reneging in machine interference queue has been carried out by Al-Seedy and Al-Ibraheem (2001). Liu and Kulkarni (2008) considered an M/PH/1 queue with work load-dependent balking. They assumed that an arriving customer joined the queue and stayed until served if and only if the system workload was less than a fixed level at the time of his arrival. They illustrate the results derived with the help of numerical examples. Liu and Kulkarni (2008) also considered the virtual queuing time process in an M/G/s queue with impatient customers. Jouini et al. (2008) modeled a call center as an M/M/s+M queue with endogenized customer reactions to announcements. They assumed that customers react by balking upon hearing the delay announcement and may subsequently reneged if they realized waiting time exceeds the delay that has originally announced to them. They calculated the waiting time distribution i.e. announcement coverage and subsequent performance in terms of balking and reneging. Al-Seedly et al. (2009) presented an analysis for the M/M/c queue with balking and reneging. They assumed that arriving customers balked with a fixed probability and reneged according to a negative exponential distribution. The generating function technique was used to obtain the transient solution of system those results in a simple differential equation. Yue et al. (2009) considered an M/M/2 queuing system with balking and two heterogeneous servers, server 1 and server 2. In this paper, they obtained the stationary condition where the system reaches a steady state and derived the steady state probabilities in a matrix form by using matrix-geometric solution method. They produced explicit expressions of some performance measures such as the mean system size, the average balking rate and the probabilities that server 2 is in various states. Numerical illustrations were also provided.

Bae and Kim (2010) considered a G/M/1 queue in which the patience time of the customers is constant. The stationary distribution of the workload of the server, or the virtual waiting time was derived by the level crossing argument. Boxma et al. (2010) considered an M/G/1 queue in which an arriving customer does not enter the system whenever its virtual waiting time i.e. the amount of work seen upon arrival, was larger than a certain random patience time. They determined the busy period distribution for various choices of the patience time distribution. Cochran et al. (2010) proposed to enable strategic decision making on future Emergency department (ED) on the basis of patient safety (rather than congestion measure). They hypothesized that the Leave without treatment (LWOT) reneging percentage is captured by the balking probability (p_k) relationship of an M/M/1/K queue. Yue and Yue (2010) studied a two server Markovian network system with balking and a Bernoulli schedule under a single vacation policy servers had different rates. They obtained the steady state condition, the stationary distribution of the number of customers in the system and the mean system size by using a matrix-geometric method.

Some other papers which have considered both balking and reneging are the work by Shawky and El-Paoumy (2000), El- Paoumy (2008), El-Sherbiny (2008), Shawky and El-Paoumy (2008), Pazgal et al. (2008).

MODEL DESCRIPTION

The model we deal with in this paper is the traditional M/M/1 model with the restriction that customers may balk from a non-empty queue as well as may renege after they join the queue. We shall assume that the inter-arrival and service rates are ë and ì respectively. As for balking we shall assume that each customer arriving at the system has a probability 'p' of balking from a non-empty queue.

Customers joining the system are assumed to be of Markovian reneging type. We shall assume that on joining the system the customer is aware of its state in the system. Consequently, the reneging rate will be taken as a function of the customer's state in the system. In particular, a customer who is at state 'n' will be assumed to have random patience time following exp (í$_n$). Under R_BOS, we shall assume that

$$\nu_n = \begin{cases} 0 & \text{for } n = 0,1. \\ \nu + c^{n-1} & \text{for } n = 2,3,... \end{cases}$$

and under R_EOS,

$$\nu_n = \begin{cases} 0 & \text{for n = 0.} \\ \nu + \mathrm{c}^{\mathrm{n-1}} & \text{for n = 1,2,3,...} \end{cases}$$

where c>1 is a constant.

Our aim behind this formulation is to ensure that higher the current state of a customer, higher is the reneging rate. Then it is clear that the constant c has to satisfy c>1. This reneging formulation also requires that as a customer progresses in the queue from state n (ne"3) to (n-1), the reneging distribution would shift from exp (í+c^{n-1}) to (í+c^{n-2}) under R_BOS. Similarly for R_EOS. In view of the memory less property, this shifting of reneging distribution is mathematically tractable, as we shall demonstrate in the subsequent sections. To the best of our knowledge, this formulation of reneging distribution has not been attempted in literature. Advantages of the same are however obvious.

The System State Probabilities

In this section, the steady state probabilities are derived by the Poisson process method (Cooper, 1981; page 17). Under R_BOS, let p_n denote the probability that there are 'n' customers in the system. Applying the Poisson process method (Cooper, 1981; page 17), we obtain the following set of steady-state equations.

$$\lambda p_0 = \mu p_1 \tag{3.1}$$

$$\lambda p_0 + (\mu + \nu + c) p_2 = \lambda(1-p) p_1 + \mu p_1 \tag{3.2}$$

$$\lambda(1-p) p_{n-1} + \{\mu + n\nu + c(c^n - 1)/(c-1)\} p_{n+1} = \lambda(1-p) p_n + \{\mu + (n-1)\nu + c(c^{n-1} - 1)/(c-1)\} p_n$$

n=2,3,… (3.3)

Solving recursively, we get (under R_BOS)

$$p_n = \left[\lambda^n (1-p)^{n-1} \Big/ \prod_{r=1}^{n} \{\mu + (r-1)\nu + c(c^{r-1} - 1)/(c-1)\} \right] p_0 \text{; n=1,2,3,…} \tag{3.4}$$

where p_0 is obtained from the normalizing condition $\sum_{n=0}^{\infty} p_n = 1$ and is given as

$$p_0 = \left[1 + \sum_{n=1}^{\infty} \lambda^n (1-p)^{n-1} \Big/ \prod_{r=1}^{n} \{\mu + (r-1)\nu + c(c^{r-1} - 1)/(c-1)\} \right]^{-1} \tag{3.5}$$

Under R_EOS, let q_n denote the probability that there are n customers in the system. Applying the Markov theory, we obtain the following set of steady state equations.

$$\lambda q_0 = (\mu + \nu) q_1 \tag{3.6}$$

$$\lambda q_0 + (\mu + 2\nu + c) q_2 = \lambda(1-p) q_1 + (\mu + \nu) q_1 \tag{3.7}$$

$$\lambda(1-p) q_{n-1} + \{\mu + (n+1)\nu + c(c^n - 1)/(c-1)\} q_{n+1} = \lambda(1-p) q_n + \{\mu + n\nu + c(c^{n-1} - 1)/(c-1)\} q_n$$

n=2,3,… (3.8)

Solving recursively, we get (under R_EOS)

$$q_n = \left[[\lambda^n (1-p)^{n-1} \Big/ \prod_{r=1}^{n} \{\mu + r\nu + c(c^{r-1}-1)/(c-1)\} \right] q_0 ; \text{ n=1,2,3,...}$$

where q_0 is obtained from the normalizing condition $\sum_{n=0}^{\infty} q_n = 1$ and is given as

$$q_0 = \left[1 + \sum_{n=1}^{\infty} \lambda^n (1-p)^{n-1} \Big/ \prod_{r=1}^{n} \{\mu + r\nu + c(c^{r-1}-1)/(c-1)\} \right]^{-1} \quad (3.9)$$

Performance Measures

In general, performance measures are the specific representation of a capacity, process or outcome deemed relevant to the assessment of performance, which are quantifiable and can be documented. The main objective of any queuing study is to assess some well-defined parameters, which are designed at striking a good balance between customer satisfaction and economic considerations. In queuing theory, measures through which the nature of the quality of service can be studied are known as performance measures. Performance measures are important as issues or problems caused by queuing situations are often related to customer's dissatisfaction with service or may be the root cause of economic losses in a business. Analysis of the relevant performance measures of queuing models allows the cause of queuing issues to be identified and the impact of proposed changes to be assessed. Some of the performance measures of any queuing system that are of general interest for the evaluation of the performance of an existing queuing system and to design a new system in terms of the level of service a customer receives as well as the proper utilization of the service facilities include mean size, server utilization, customer loss and the like.

An important measure is 'L', which denotes the mean number of customers in the system. To obtain an expression for the same, we note that L=P´(1) where

$$P'(1) = \frac{d}{ds} P(s)\,|_{s=1}$$

Here P(S) is the p.g.f. of the steady state probabilities. The derivation of P´(1) is given in the appendix. Then from (A.5) we have,

$$L_{R\_BOS} = [\lambda - (\mu - \nu + \lambda p)(1-p_0) - p_0 + \{c - (p_0 / p_0(c\lambda, \mu, \nu))/(c-1)\}]/\nu \text{ for c>0, c} \neq 1$$

$$L_{q(R\_BOS)} = [\lambda - (\mu + \lambda p)(1-p_0) - p_0 + \{c - (p_0 / p_0(c\lambda, \mu, \nu))/(c-1)\}]/\nu \text{ for c>0, c1}$$

where p_0 and p_0(cë,ì,í) are defined in (3.5) and (A.3) respectively.

and from (B.1) we have,

for c>0, c1

for c>0, c1

where q_0 and q_0(cë,ì,í) are defined in (3.9) and (B.2) respectively.

The Little's formula for calculating average waiting time in the system and average waiting time in queue are given by (Sztrik, 2016, pp: 22)

$W = L / \lambda$

$W_q = L_q / \lambda$ respectively.

where L and L_q are mean system size and mean queue size respectively.

Using the above mentioned Little's formula, we can calculate the average waiting time in the system and average waiting time in queue under R_BOS and R_EOS.

Customers arrive into the system at the rate ë. However all the customers who arrive do not join the system because of balking. The effective arrival rate into the system is thus different from the overall arrival rate and is given by

$$\begin{aligned} \lambda^{e}{}_{(R\_BOS)} &= \lambda p_0 + \lambda(1-p)p_1 + \lambda(1-p)p_2 + \ldots. \\ &= \lambda p_0 + \lambda(1-p)\sum_{n=1}^{\infty} p_n \\ &= \lambda\{1 - p(1-p_0)\} \end{aligned} \tag{4.1}$$

Similarly,

$$\lambda^{e}{}_{(R\_EOS)} = \lambda\{1 - p(1-q_0)\} \tag{4.2}$$

We have assumed that each customer has a reneging distribution (of type R_BOS or R_EOS) following exp(í$_n$). Clearly then, the reneging rate of the system would depend on the state of the system as well as reneging rule. The average reneging rate (avgrr) is given by

$$\begin{aligned} Avgrr_{(R\_BOS)} &= \sum_{n=2}^{\infty}\{(n-1)\nu + c(c^{n-1}-1)/(c-1)\}p_n \\ &= \nu\{P'(1) - p_1\} - \nu(1-p_0-p_1) + [\{P(c) - cp_1 - p_0\} - c\{1-p_0-p_1\}]/(c-1) \\ &= \lambda - (\mu + \lambda p)(1-p_0) \end{aligned} \tag{4.3}$$

$$\begin{aligned} Avgrr_{(R\_EOS)} &= \sum_{n=2}^{\infty}\{n\nu + c(c^{n-1}-1)/(c-1)\}q_n \\ &= \nu Q'(1) + [\{Q(c) - q_0 - c(1-q_0)\}]/(c-1) \\ &= \lambda - (\mu + \lambda p)(1-q_0) \end{aligned} \tag{4.4}$$

In a real life situation, customers who balk or renege represent business lost. Management of any queuing system would like to know the proportion of total customers lost in order to have an idea of total business lost. Customers are lost to the system in two ways due to balking and due to reneging.

Hence the mean rate at which customers are lost (under R_BOS) is

$$\lambda - \lambda^{e}_{(R\_BOS)} + avgrr_{(R\_BOS)} = \lambda - \mu(1-p_0) \qquad (4.5)$$

and the mean rate at which customers are lost (under R_EOS) is

$$\lambda - \lambda^{e}_{(R\_EOS)} + avgrr_{(R\_EOS)} = \lambda - \mu(1-q_0) \qquad (4.6)$$

These rates helps in the determination of proportion of customers lost. This proportion (under R_BOS) is given by

$$\lambda - \lambda^{e}_{(R\_BOS)} + avgrr_{(R\_BOS)} \big/ \lambda = [\lambda - \mu(1-p_0)] / \lambda \quad \text{using (5.5)} \qquad (4.7)$$

and the proportion (under R_EOS) is given by

$$\lambda - \lambda^{e}_{(R\_EOS)} + avgrr_{(R\_EOS)} \big/ \lambda = [\lambda - \mu(1-q_0)] / \lambda \quad \text{using (5.6)} \qquad (4.8)$$

The proportion of customer completing receipt of service can now be easily determined from the above proportion.

Customers who leave the system do not receive service. Consequently, only those customers who reach the service station constitute the actual load of the server. From the server's point of view, this provides a measure of the amount of work he has to do. Let us call the rate at which customers reach the service station as ë^{s}. Then under R_BOS ë$^{s}_{(R_BOS)}$= ë$^{e}_{(R_BOS)}$(1-proportion of customers lost due to reneging out of those joining the system)

$$= \lambda^{e}_{(R\_BOS)} \left\{ 1 - \sum_{n=2}^{k} (n-1)\nu p_n / \lambda^{e}_{(R\_BOS)} \right\}$$

$$= \lambda^{e}_{(R\_BOS)} - avgrr_{(R\_BOS)}$$

$$= \mu(1-p_0) \quad \text{using (4.1) and (4.3)}$$

In case of R_EOS, one needs to recall that customers may renege even while being served and only those customers who renege from the queue will not constitute any work for the server. Thus ë$^{s}_{(R_EOS)}$= ë$^{e}_{(R_EOS)}$(1-proportion of customers lost due to reneging from the queue out of those joining the system)

$$= \lambda^{e}_{(R\_EOS)} \left\{ 1 - \sum_{n=2}^{k} (n-1)\nu q_n / \lambda^{e}_{(R\_EOS)} \right\} \quad \text{using (4.2) and (4.4)}$$

$$= \lambda^{e}{}_{(R\_EOS)} - avgrr_{(R\_EOS)}$$

$$= \mu(1 - q_0)$$

In order to ensure that the system is in steady state, it is necessary for the rate of customers reaching the service station to be less than the system capacity. This translates to

$$(\lambda^{s}/\mu) < 1$$

Sensitivity Analysis

We have assumed that there are essentially three parameters viz: relating to the stochastic nature of arrival, service and reneging patterns. Various reasons may influence these parameters so that on different occasions these may undergo change. From managerial point of view, an idle server is a waste. So also for low server utilization. It is therefore interesting to examine and understand how server utilization varies in response to change in system parameters. We place below the effect of change in these system parameters on server utilization. For this purpose, we shall follow the following notational convention in the rest of this section.

$p_n(\lambda, \mu, \nu)$ and $q_n(\lambda, \mu, \nu)$ will denote the probability that there are 'n' customers in a system with parameters λ, μ, ν in steady state under R_BOS and R_EOS respectively.

i) Let ë$_1$> ë$_0$. Then

$$\frac{p_0(\lambda_1, \mu, \nu)}{p_0(\lambda_0, \mu, \nu)} < 1$$

$$\Rightarrow \frac{(\lambda_0 - \lambda_1)}{\mu} + \frac{(1-p)}{\mu(\mu + \nu + c)}\left(\lambda_0^{\,2} - \lambda_1^{\,2}\right) + \ldots\ldots < 0$$

which is true and hence $p_0 \downarrow \ as\ \lambda \uparrow$.

ii) Let ì$_1$> ì$_0$. Then

$$\frac{p_0(\lambda, \mu_1, \nu)}{\mathrm{p}_0(\lambda, \mu_0, \nu)} > 1$$

$$\Rightarrow \lambda\left(\frac{1}{\mu_0} - \frac{1}{\mu_1}\right) + \lambda^2(1-p)\left\{\frac{1}{\mu_0(\mu_0 + \nu + c)} - \frac{1}{\mu_1(\mu_1 + \nu + c)}\right\} + \ldots\ldots > 0$$

which is true and hence $p_0 \uparrow \quad as\ \mu \uparrow$

iii) Let í$_1$> í$_0$. Then

$$\frac{p_0(\lambda,\mu\ ,v_1,k)}{\mathrm{p}_0(\lambda,\mu,v_0,k)}>1$$

$$=\lambda\left(\frac{1}{\mu}-\frac{1}{\mu}\right)+\lambda^2(1-p)\left\{\frac{1}{\mu(\mu+v_0+c)}-\frac{1}{\mu(\mu+v_1+c)}\right\}+........>0$$

which is true and hence $p_0 \uparrow$ *as* $v \uparrow$

The following can similarly be shown.

iv) $q_0 \downarrow$ *as* $\lambda \uparrow$

v) $q_0 \uparrow$ *as* $\mu \uparrow$

vi) $q_0 \uparrow$ *as* $v \uparrow$

These results state that an increase in arrival rate would result in lowering of the fraction of time the server is idle. An increase in service rate would mean the server is able to work efficiently so that it can process same amount of work quickly. This translates to higher server idle time. An increase in reneging rate would mean the server has fewer work to do and hence higher fraction of idle time.

Numerical example

To illustrate the use of our results, we consider the following example from Ravindran, Phillips and Solberg (1987, page 338). It concerns a grain elevator at a small town in one of the plains states where wheat is the principal crop. During the harvest season, trucks loaded with wheat from the fields arrive at the elevator, where they must quickly deposit their load and return to fields for another. The check in process involves weighing the truck, drawing samples for moisture and contamination tests, and a few other details before the load can be dumped through a grate.

The farmers are very concerned about getting their crop off the field and into the elevator quickly. Once the wheat is ripe, it is highly vulnerable to rain or wind. Any delay could threaten a significant portion of a farmer's income for the whole year. Of course, all of the fields in a region ripen about the same time, so it is not surprising that a traffic problem could develop at the elevator. The average interarrival time is estimated to be 9/hr, further 10 trucks are serviced on the average.

Given the nature of the problem it is natural that there will be balking as well as reneging. The primary concern of the farmers is to ensure that the crop does not get destroyed. Consequently, the bigger the queue, the more impatient the farmer would be. Our model assumptions can model such a scenario. We shall assume that reneging distribution is state dependent following exp($í_n$) where $í_n$ is as described in section 2. Specifically, we shall assume í=2/hr and considered two scenarios with c viz: c=1.1 and c=1.2.

At the same time because of farmers concern, some of them will balk. We are assuming a balking probability (p) of 0.02. With these parameters, various performance measures of interest have been computed using FORTRAN 77 program coded by the authors. The results are presented in the following table:

Table 1: Table of Performance Measures
(assuming ë=9/hr, ì=10/hr, í=2/hr and p=0.02.)

Performance Measure	c=1.1	c=1.2
Average length of system	1.3487815	1.316907
Average length of queue	0.674331	0.647139
Mean reneging rate	2.143934	2.181762
Proportion of customer lost due to balking and reneging	0.251685	0.255813
Proportion of customer acquiring service	0.748315	0.744187
Effective arrival rate in to the system	8.878773	8.879442
Arrival rate of customer reaching service station	6.734839	6.697682
Fraction of time server is idle	0.326516	0.330232

From the above table it is clear that as the reneging rate increases, average length of system, average length of queue, proportion of customers acquiring service and arrival rate of customer reaching service station decreases. It is also seen that the effective arrival rate, proportion of customer lost due to reneging and due to balking and fraction of time server is idle increases with the increase in reneging rate. These results are on expected lines. The elevator management may look into the fact that at least one in four customers is lost due to balking and reneging. This is an area of concern for them.

CONCLUSION

We have provided closed form expressions of a number of performance measures in the single server Markovian queuing system with balking and position dependent reneging. To the best of our knowledge, this has not been attempted. This paper is an attempt to address this gap in the literature. The emphasis has been on understanding the extent of customer's loss. We believe that these performance measures can assist the system manager to reduce business loss. These are the merits of this work. The only demerit is that we have assumed a particular formulation (in section 2) to model position dependent reneging rates. We agree that other reneging situations may require different formulations. This is a pointer to extension of our work. Extension of our results to queues with non markovian arrival and service distributions can also be another extension of our work,

Appendix A: Derivation of $P'(1)$ under R_BOS

Let P(s) denote the probability generating function, which is given by

$$P(s) = \sum_{n=0}^{\infty} p_n s^n$$

From equation (3.3) we have

$$\lambda(1-p)p_{n-1} + \{\mu + n\nu + c(c^n - 1)/(c-1)\}p_{n+1} = \lambda(1-p)p_n + \{\mu + (n-1)\nu + c(c^{n-1} - 1)/(c-1)\}p_n$$

n=2,3,4……

Now multiplying both sides of the equation by s^n and summing over n

$$\lambda s\sum_{n=2}^{\infty}(1-p)p_{n-1}s^{n-1} + \frac{1}{s}\sum_{n=2}^{\infty}(\mu + n\nu + c(c^n-1)/(c-1))p_{n+1}s^{n+1} = \lambda\sum_{n=2}^{\infty}(1-p)p_n s^n + \sum_{n=2}^{\infty}\{\mu + (n-1)\nu + c(c^{n-1}-1)/(c-1)\}p_n s^n$$

$$\Rightarrow \lambda s(1-p)\sum_{n=2}^{\infty}p_{n-1}s^{n-1} - \lambda(1-p)\sum_{n=2}^{\infty}p_n s^n = \sum_{n=2}^{\infty}\{\mu + (n-1)\nu + c(c^{n-1}-1)/(c-1)\}p_n s^n - \frac{1}{s}\sum_{n=2}^{\infty}\{\mu + n\nu + c(c^{n-1}-1)/(c-1)\}p_{n+1}s^{n+1}$$

$$\Rightarrow \lambda s\{P(s) - p_0\} - \lambda(1-p)\{P(s) - p_0 - p_1 s\} = \mu\{P(s) - p_0 - p_1 s\} + \nu s\{P'(s) - p_1\} - \nu\{P(s) - p_0 - p_1 s\} + \sum_{n=2}^{\infty}\{c(c^{n-1}-1)/(c-1)\}p_n s^n$$
$$-1/s\left[\mu\{P(s) - p_0 - p_1 s - p_2 s^2\} + \nu s\{P'(s) - 2p_2 s - p_1\} - \nu\{P(s) - p_0 - p_1 s - p_2 s^2\} + \sum_{n=2}^{\infty}\{c(c^n - 1)/(c-1)\}p_{n+1}s^{n+1}\right]$$

$$\Rightarrow \lambda s\{P(s) - p_0\} - \lambda(1-p)\{P(s) - p_0 - p_1 s\} = \mu\{P(s) - p_0 - p_1 s\} + \nu s\{P'(s) - p_1\} - \nu\{P(s) - p_0 - p_1 s\} +$$
$$c/(c-1)\{(cp_2 s^2 + c^2 p_3 s^3 + ...) - (p_2 s^2 + p_3 s^3 + ...)\}$$
$$-1/s\left[\begin{array}{l}\mu\{P(s) - p_0 - p_1 s - p_2 s^2\} + \nu s\{P'(s) - 2p_2 s - p_1\} - \nu\{P(s) - p_0 - p_1 s - p_2 s^2\} + \\ c/(c-1)\{(c^2 p_3 s^3 + c^3 p_4 s^4 + ...) - (p_3 s^3 + p_4 s^4 + ...)\}\end{array}\right]$$

$$\Rightarrow \lambda s(1-p)P(s) - \lambda s(1-p)p_0 - \lambda(1-p)P(s) + \lambda(1-p)p_0 + \lambda(1-p)p_1 s = \mu P(s) - \mu p_0 - \mu p_1 s + \nu sP'(s) - \nu sp_1$$
$$-\nu P(s) + \nu p_0 + \nu sp_1 + 1/(c-1)\{P(cs) - p_0 - p_1 cs - cP(s) + cp_0 + cp_1 s\} - \mu/s\{P(s) - p_0 - p_1 s - p_2 s^2\}$$
$$-\nu\{P'(s) - 2p_2 s - p_1\} + \nu/s\{P(s) - p_0 - p_1 s - p_2 s^2\} -$$
$$1/(c-1)s\{P(cs) - p_0 - p_1 cs - p_2 c^2 s^2 - cP(s) + cp_0 + cp_1 s + cp_2 s^2\}$$

$$\Rightarrow \lambda s(1-p)P(s) - \lambda s(1-p)p_0 - \lambda(1-p)P(s) + \lambda(1-p)p_0 + \lambda(1-p)p_1 s = \mu P(s) - \mu p_0 - \mu p_1 s + \nu sP'(s)$$
$$-\nu P(s) + \nu p_0 - p_0/(c-1) + P(cs)/(c-1) - csp_1/(c-1) - cP(s)/(c-1) + csp_1/(c-1) + cp_0/(c-1) - \mu P(s)/s$$
$$+\mu p_0/s + \mu p_1 + \mu p_2 s - \nu P'(s) + 2\nu p_2 s + \nu p_1 + \nu/s\,P(s) - \nu p_0/s - \nu p_1 - \nu p_2 s - P(cs)/(c-1)s + p_0/(c-1)s +$$
$$cp_1/(c-1) + c^2 sp_2/(c-1) + cP(s)/(c-1)s - cp_0/(c-1)s - cp_1/(c-1) - csp_2/(c-1)$$

$$\Rightarrow \nu P'(s) - \nu sP'(s) = \mu P(s) - \mu p_0 - \mu p_1 s - \lambda s(1-p)P(s) + \lambda s(1-p)p_0 + \lambda(1-p)P(s) - \lambda(1-p)p_0 - \lambda(1-p)p_1 s$$
$$-\nu P(s) + \nu p_0 - p_0/(c-1) + P(cs)/(c-1) - cP(s)/(c-1) + cp_0/(c-1) - \mu P(s)/s$$
$$+\mu p_0/s + \mu p_1 + \mu p_2 s + \nu p_2 s + \nu/s\,P(s) - \nu p_0/s - P(cs)/(c-1)s + p_0/(c-1)s +$$
$$cp_1/(c-1) + c^2 sp_2/(c-1) + cP(s)/(c-1)s - cp_0/(c-1)s - cp_1/(c-1) - csp_2/(c-1)$$

$$\Rightarrow \nu(1-s)P'(s) = \nu P(s)(1-s)/s - \mu P(s)(1-s)/s + \lambda(1-p)P(s)(1-s) + \mu p_0(1-s)/s + \mu p_1(1-s)$$
$$-\nu p_0(1-s)/s + p_2 s(\mu + \nu + c) - \lambda(1-p)p_0(1-s) - \lambda^2 p_0 s(1-p)/\mu - P(cs)(1-s)/s(c-1)$$
$$+cP(s)(1-s)/s(c-1) - p_0(1-s)/s$$

$$\Rightarrow \nu(1-s)P'(s) = [\nu/s - \mu/s + \lambda(1-p) + c/(c-1)](1-s)P(s) + \mu p_0(1-s)/s + \lambda p_0(1-s)$$
$$- \nu p_0(1-s)/s - \lambda(1-p)p_0(1-s) - P(cs)(1-s)/s(c-1) - p_0(1-s)/s$$

$$\Rightarrow \nu P'(s) = \{\nu/s - \mu/s + \lambda(1-p) + c/(c-1)\}P(s) + \mu p_0/s + \lambda p_0 - \nu p_0 - \lambda(1-p)p_0 - P(cs)/s(c-1) - p_0/s$$

$$\Rightarrow P'(s) = (1/\nu)[\{\nu/s - \mu/s + \lambda(1-p) + c/(c-1)\}P(s) + \mu p_0/s + \lambda p_0 - \nu p_0 - \lambda(1-p)p_0 - P(cs)/s(c-1) - p_0/s]$$

Now (A.1)

$$\lim_{s \to 1-} p'(s) = \lim_{s \to 1-}(1/\nu)[\{\nu/s - \mu/s + \lambda(1-p) + c/(c-1)\}P(s) + \mu p_0/s + \lambda p_0 - \nu p_0 - \lambda(1-p)p_0 - P(cs)/s(c-1) - p_0/s]$$

$$\Rightarrow p'(1) = (1/\nu)[\{\nu - \mu + \lambda(1-p) + c/(c-1)\} + \mu p_0 + \lambda p_0 - \nu p_0 - \lambda(1-p)p_0 - P(c)/(c-1) - p_0]$$
$$= (1/\nu)[\lambda - (\mu - \nu + \lambda p)(1-p_0) - p_0 + \{c - P(c)/(c-1)\}]$$

Here $P(c) = \sum_{n=0}^{\infty} p_n(\lambda,\mu,\nu)c^n$ where the symbol $p_n(\lambda,\mu,\nu)$ is as described in section 5. We use p_n and interchangeably. However should any of the parameters change, it is explicitly stated. To obtain a closed form expression for P(c), let us for the time being, consider another queuing system with parameter and assumptions similar to the queuing system we are presently considering except that the arrival rate is 'cë'. For this new system, the steady state equations are same as (3.1), (3.2) and (3.3) with 'ë' replaced by 'cë'. Denoting the steady state probabilities of this new system by, we can obtain

$$p_n(c\lambda,\mu,\nu) = \left[(c\lambda)^n(1-p)^n \Big/ \prod_{r=1}^{n}\{\mu + (r-1)\nu + c(c^{r-1}-1)/(c-1)\}\right]p_0(c\lambda,\mu,\nu)$$

n=1,2,…. (A.2)

where

$$p_0(c\lambda,\mu,\nu) = \left[1 + \sum_{n=1}^{\infty} (c\lambda)^n(1-p)^{n-1} \Big/ \prod_{r=1}^{n}\{\mu + (r-1)\nu + c(c^{r-1}-1)/(c-1)\}\right]^{-1}$$

(A.3)

Let $P(S;c\lambda,\mu,\nu)$ denotes the probability generating function of the steady state probabilities of this new queuing system so that

$$P(S;c\lambda,\mu,\nu) = \sum_{n=0}^{\infty} p_n(c\lambda,\mu,\nu)s^n$$

Now,

$$P(c) = \sum_{n=0}^{\infty} p_n(\lambda, \mu, \nu) c^n$$

$$= p_0 + \sum_{n=1}^{\infty} \left[(c\lambda)^n (1-p)^n \Big/ \prod_{r=1}^{n} \left\{ \mu + (r-1)\nu + c(c^{r-1} - 1)/(c-1) \right\} \right] p_0$$

$$\Rightarrow \left[(P(c) - p_0)/p_0 \right] = \sum_{n=1}^{\infty} \left[(c\lambda)^n (1-p)^n \Big/ \prod_{r=1}^{n} \left\{ \mu + (r-1)\nu + c(c^{r-1} - 1)/(c-1) \right\} \right] \quad \text{(A.4)}$$

Using this in (A.1) we obtain (for c>1)

$$P'(1) = (1/\nu)\left[\lambda - (\mu - \nu + \lambda p)(1 - p_0) - p_0 + \left\{ c - \left(p_0 / p_0(c\lambda, \mu, \nu) \right) \right\} / (c-1) \right] \quad \text{(A.5)}$$

where p_0(cë,ì,í) is given in (A.3).

Appendix B: Derivation of Q´(1) under R_EOS

From equation (3.8) we have,

$$\lambda(1-p)q_{n-1} + \left\{ \mu + (n+1)\nu + c(c^n - 1)/(c-1) \right\} q_{n+1} = \lambda(1-p)q_n + \left\{ \mu + n\nu + c(c^{n-1} - 1)/(c-1) \right\} q_n$$

n=2,3…..

Multiplying both sides of this equation by s^n and summing over n from we get

$$\lambda s(1-p)\sum_{n=2}^{\infty} q_{n-1}s^{n-1} - \lambda(1-p)\sum_{n=2}^{\infty} q_n s^n = \sum_{n=2}^{\infty} \{\mu + n\nu + c(c^{n-1} - 1)/(c-1)\} p_n s^n - \frac{1}{s}\sum_{n=2}^{\infty} \left\{ \mu + (n+1)\nu + \left(c(c^n - 1)/(c-1) \right) \right\} p_{n+1} s^{n+1}$$

Proceeding in a manner similar to Appendix A, we obtain,

$$Q'(1) = (1/\nu)\left[\lambda - (\mu + \lambda p)(1 - q_0) + \left\{ c - \left(q_0 / q_0(c\lambda, \mu, \nu) \right) \right\} / (c-1) \right] \quad \text{(B.1)}$$

where

$$q_0(c\lambda, \mu, \nu) = \left[1 + \sum_{n=1}^{\infty} (c\lambda)^n (1-p)^{n-1} \Big/ \prod_{r=1}^{n} \left\{ \mu + r\nu + c(c^{r-1} - 1)/(c-1) \right\} \right]^{-1} \quad \text{(B.2)}$$

REFERENCES

1. Altman E. and Yechiali U. (2008) Infinite-Server Queues with System's Additional Tasks and Impatient Customers. *Probability in the Engineering and Informational Sciences*, 22(4): 477-493
2. Al-Seedy R.O. and Al-Ibraheem F.M. (2001) An Interarrival Hyperexponential Machine Interference with Balking, Reneging, State-dependent, Spares and an Additional Server for Longer Queues. *International Journal of Mathematics and Mathematical Sciences*, 27(12): 737-747
3. Al-Seedy R.O., El-Sherbiny A. A., El-Shehawy S. A. and Ammar S. I. (2009) Transient solution of the M/M/c queue with balking and reneging. Computers and Mathematics with Applications, 57(8):1280-1285

4. Ancker C.J. Jr. and Gafarian A.V. (1963) Queuing Problems with Balking and Reneging I, *Operations Research*, 11(1): 88-100
5. Bae J. and Kim S. (2010) The stationary workload of the G/M/1 queue with impatient customers. *Queuing Systems*, 64(3): 253-265
6. Bae J., Kim S. and Lee E.Y. (2001) The virtual waiting time of the M/G/1 Queue with Impatient Customers. Queuing Systems: Theory and Applications, 38(4):485-494
7. Baccelli F., Boyer P. and Hebuterne G. (1984) Single Server Queues with Impatients Customers. *Advances in Applied Probability*, 16(4): 887-905
8. Barrer, D.Y. (1957). Queuing with impatient customers and ordered Service. *Operations Research*, 5(5):650-656
9. Barrer D.Y. (1957) Queuing with impatient customers and indifferent clerks. *Operations Research*, 5(5): 644-649
10. Boots N.K. and Tijms H. (1999) An M/M/c Queues with Impatient Customers. *TOP*, 7(2): 213-220
11. Boxma O., Perry D., Stadje W. and Zacks S. (2010) The busy period of an M/G/1 queue with customer impatience. *Journal of Applied Probability*, 47(1): 130-145
12. Brandt A. and Brandt M. (1999) On a Two-Queue Priority System with Impatience and its Application to a Call Centre. *Methodology and Computing in Applied Probability*, 1(2):91-210
13. Choi B.D., Kim B. and Chung J. (2001) M/M/1 Queue with Impatient Customers of Higher Priority. *Queuing Systems*, 38(1): 49-66
14. Choi. B.D., Kim B. and Zhu D. (2004) MAP/M/C Queue with Constant Impatients Times. *Mathematics of Operations Research*, 29(2): 309-325
15. Choudhury A. (2008) Impatience in Single Server Queuing Model. *American Journal of Mathematical and Management Sciences*, 28: 177-211
16. Choudhury A. (2009) A few words on Reneging in M/M/1/K queues. *Contributions to Applied and Mathematical Statistics*, 4: 58-64
17. Choudhury A. and Medhi P. (2011) Balking and Reneging in multi server Markovian Queuing System. *International Journal of Mathematics in Operational Research*, 3(4): 377-394
18. Cochran J.K. and Broyles J.R. (2010) Developing non-linear queuing regressions to increase emergency department patient safety: Approximating reneging with balking. *Computers and Industrial Engineering*, 59(3): 378-386
19. Cooper B.R. (1981) Introduction to Queuing Theory, Second Edition, *Elsevier, North Holland, Inc.*
20. El-Paoumy M.S. (2008) On Poisson Bulk Arrival Queue: $M^X/M/2/N$ with Balking, Reneging and Heterogeneous Servers. *Applied Mathematical Sciences*, 2(24): 1169-1175
21. El-Paoumy M.S. (2008) On a Truncated Erlangian Queuing system with State-Dependent Service rate, Balking and Reneging, *Applied Mathematical Sciences*, 2(24): 1161-1167
22. El-Paoumy M.S. and Ismail M.M. (2009) On a truncated Erlang Queuing System with Bulk Arrivals, Balking and Reneging. *Applied Mathematical Sciences*, 3(23): 1103-1113
23. El-Sherbiny A.A. (2008) The Non-Truncated Bulk Arrival Queue $M^X/M/1$ with Reneging, Balking, State-Dependent and an Additional Server for Longer Queues. *Applied Mathematical Sciences*, 2(15):747-752
24. Gavish B. and Schweitzer P.J. (1977) The Markovian Queue with Bounded Waiting Time. *Management Science*, 23(12): 1349-1357

25. Ghoshal, A. (1963). Queues with Finite Waiting Time. Operations Research, Vol. 11, No. 6, pp: 919-921.
26. Hagighi, A. Montazer, Medhi, J. and Mohanty, S. G. (1986). On a multi server Markovian queuing system with Balking and Reneging. *Computer and Operational Research*, 13(4): 421-425
27. Haight F.A. (1957) Queuing with Balking. *Biometrika*, 44(3/4):360-369
28. Haight F.A. (1959) Queuing with Reneging. *Metrika*, 2(1): 186-197
29. Jouini O., Aksin Z. and Dallery Y. (2008) Call Centers with Delay Information: Models and Insights. Downloaded from the site www.uclouvain.be/cps/ucl /doc/core/documents /Jouini.pdf. visited on 03.08.2010.
30. Jouini O., Dallery Y. and Aksin Z. (2009) Queuing models for full-flexible multi-class call centers with real-time anticipated delays. *International Journal of Production Economics*, 120(2): 389-399
31. Ke J.C. and Wang K.H. (1999) Cost analysis of the M/M/R machine repair problem with balking, reneging and server breakdowns. *The Journal of the Operational Research Society*, 50(3): 275-282
32. Kok A.G.D and Tijms H. (1985) A Queuing System with Impatient customers. *Journal of Applied Probability*, 22(3): 688-696
33. Liu L. and Kulkarni V.G. (2008) Busy Period analysis for M/PH/1 queues with workload dependent Balking. *Queuing Systems,* 59(1): 37-51
34. Liu L. and Kulkarni V.G. (2008) Balking and reneging in m/g/s system exact analysis and approximations. *Probability in the Engineering and Informational Sciences*, 22(3): 355-371
35. Liu L, Kashyap B.R. K. and Templeton J.G.C. (1987) The service system $M/M^R/$" with impatient customers. *Queuing Systems*, 2(4): 363-372
36. Martin M. and Artalejo J. R. (1995) Analysis of an M/G/1 Queue with two types of Impatient units. *Advances in Applied Probability*, 27(3): 840-861
37. Medhi J. (2003). Stochastic Model in Queuing Theory, Second Edition, *Academic Press*
38. Pazgal A. and Radas S. (2008) Comparison of customer balking and reneging behavior to queuing theory predictions: An experimental study. *Computers and Operations Research*, 35(8): 2537-2548
39. Ravindran A, Phillips, Don T. and Solberg J.J. (1987) Operations Research, Principles and practice, Second Edition, John Wiley and Sons. (New York)
40. Shawky A.I. (1997) The single server machine interference model with balking, reneging and an additional server for longer queues. *Microelectronics Reliability*, 37(2): 355-357
41. Shawky A.I. and El- Paoumy M.S. (2000) The Interarrival Hyperexponential Queues: $H_k/M/c/N$ with Balking and Reneging. Stochastics *An International Journal of Probability and Stochastic Processes*, 69(1&2): 67-76
42. Shawky A.I. and El-Paoumy M.S. (2008) The Truncated Hyper-Poisson Queues: $H_k/M^{a,b}/c/N$ with Balking, Reneging and General Bulk-Service Rule. Yugoslav *Journal of Operations Research*, 8(1): 23-36
43. Singh C.J., Kumar B. and Jain M. (2007) Single server interdependent queuing model with controllable arrival rates and reneging. *Pakistan J. Statistics*, 23(2): 171-178
44. Sztrik J. (2016) Basic Queuing Theory, Foundation of System Performance Modeling, *Globe Edit*. (Omni Scriptum GmbH & Co. KG)

45. Xiong W., Jagerman D. and Altiok T. (2008) M/G/1 queue with deterministic reneging times. *Performance Evaluation*, 65(3&4): 308-316
46. Xiong W. and Altiok T. (2009) An approximation for multi-server queues with deterministic reneging times. *Annals of Operations Research,* 172(1): 143-151
47. Yue D. and Yue W. (2010) A heterogeneous two-server network system with balking and a bernoulli vacation schedule. Journal of Industrial and Management Optimization, 6(3): 501-516
48. Yue D., Yue W., Yu J. and Tian R. (2009) A heterogeneous two-server queuing system with balking and server breakdowns. Paper presented in the Eight International Symposium on Operations Research and its Applications (ISORA'09), Zhangjiajie, Chaina, Sept 20-22. Downloaded from the site www.aporc.org / LNOR / 10 / ISORA 2009 F31.pdf. on 04.08.2010.
49. Zhang Y., Yue D. and Yue W. (2005) Analysis of an M/M/1/N queue with balking, reneging and server vacations. Paper presented in the Vth. International Symposium on OR and its Application, 2005 and downloaded from www.aporc.org/LNOR/6/ISORA 2006F10.pdf. visited on 04.08.2010.