Numbers of Served and Lost Customers in Busy-Periods of $M/M/1/n$ Systems with Balking

Fátima Ferreira[1,4], António Pacheco[2,4], and Helena Ribeiro[3,4(✉)]

[1] University of Trás-os-Montes and Alto Douro, Vila Real, Portugal
mmferrei@utad.pt
[2] Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
apacheco@math.tecnico.ulisboa.pt
[3] Escola Superior de Tecnologia e Gestão do Politécnico de Leiria, Leiria, Portugal
helena.ribeiro@ipleiria.pt
[4] CEMAT, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

Abstract. In this work we analyze single server Markovian queueing systems with finite capacity and balking, that is $M/M/1/n$ systems with balking. In these systems, the admission of customers is modulated by the state of the system at the instants of customer arrivals. Depending on the size of the queue upon arrival, customers that find place to join the system decide to enter the system with a certain probability. The number of customers in the system amounts to a Markov chain whose transition probabilities incorporate the balking probabilities. Using the Markovian regenerative property of the chain embedded at the instants of arrival or departure of customers, we characterize the joint probability distribution of the number of customers served and the number of customers lost in busy-periods, that is, during continuous occupation periods of the server. This is accomplished implementing a priori a recursive algorithmic procedure for computing the respective probability-generating function. Finally, a numerical illustration of the derived results is presented for different balking policies.

Keywords: Queueing system · Balking · Busy-period · Markov chain

1 Introduction

The perception of the queue length by an arriving customer plays a relevant role in their decision to whether join (or not) a queueing system. In fact, in current rush days there is a growing tendency of customers to balk (i.e., to not join the system) when facing at arrival a long queue size (that exceeds their patience level) or a short queue if a loaded system is perceived as a good condition.

This research was partially supported by Fundação para a Ciência e a Tecnologia (FCT) through project UID/Multi/04621/2019.

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12251, pp. 147–156, 2020.
https://doi.org/10.1007/978-3-030-58808-3_12

As the balking phenomenon may result in significant loss of customers, the analysis of queueing systems with balking is important in practice.

In this work we address the study of finite capacity Markovian queues with a single server under different customer blocking policies when the customers arrive at the system. Using the notation introduced by David G. Kendall, this corresponds to a $M/M/1/n$ system with balking or reverse balking. These systems have many applications because in many real-life situations customers often have the possibility to postpone or give up a given type of service when, on arrival at the system, the level of server congestion is considered unsatisfactory by the same customers.

The notion of customer balking was introduced in the work of Haight [5] that considers an $M/M/1$ queue in which customers do not join the system whenever they face at arrival a queue length larger than a fixed threshold. After this pioneering work, the study of queues under different balking and other types of customer's impatience policies has aroused the interest of several authors (*cf* [2–4,7–10,12,14,15] and references therein). A review on queueing systems with impatient customers can be found in [13].

The analysis of queues with balking in busy-periods, i.e., in continuous periods of effective use of the server, is relevant from the operator's point of view and provides crucial information for its management. We consider multi busy-periods, *i.e.*, busy-periods initiated with multiple customers in the system. Specifically, by a busy-period initiated by i customers, hereinafter referred to as i-busy-period, we mean the period of time that starts at an arrival instant that makes the system stay with i customers, and ends at the first subsequent time at which the system becomes empty, with a customer initiating service after the arrival instant. This definition is in line with that of remaining busy-period from state i given in Harris [6], of residual busy-period provided in Hanbali [1] and of busy-period initiated with i customers considered in Peköz *et al.* [11].

The aim of this paper is to characterize the joint probability distribution function of the number of customers served and the number of customers lost in busy-periods of $M/M/1/n$ system with balking. In Sect. 2, after describing the system we present the transition probability matrix of the discrete time Markov chain (DTMC) embedded at the moments of arrival or departure of customers. Taking advantage of the Markovian structure of these queues and by using the probability-generating function technique, we obtain in Sect. 3 the joint probability function of the number of customers served and the number of customers lost in busy-periods.

In Sect. 4 we illustrate the results obtained and, finally, we present some concluding remarks in Sect. 5.

2 Model Description

We consider $M/M/1/n$ systems with balking of customers. These are queues with finite capacity, n, at which the customers arrive (one by one) according to a Poisson process with rate λ and are served (one by one) by a single-server, in

a first-come first-served discipline. Service times are independent and identically exponential distributed variables with rate μ, and are independent of the arrival process.

These systems differ from the traditional $M/M/1/n$ given that the customer admission is modulated by the state of the system at the time of arrival. As in the traditional $M/M/1/n$ systems, if on arrival a customer finds the system full (with n customers) he is blocked with probability 1. However, if on arrival a customer finds $i < n$ customers in the system, even though there is place in the system to accommodate the customer, he either decides to join the queue with probability e_i or not to enter (balk) with probability $1 - e_i$. For completeness we let $e_n = 0$.

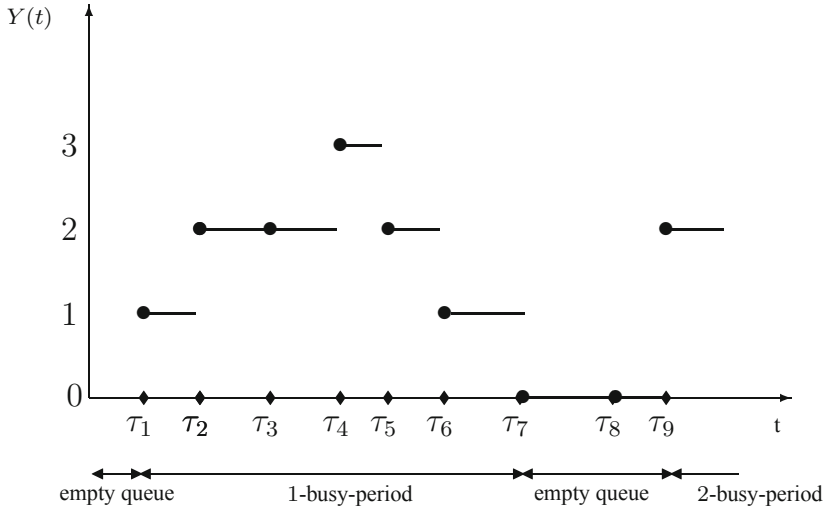


Fig. 1. Typical sample path of $M/M/1/n$ system with balking.

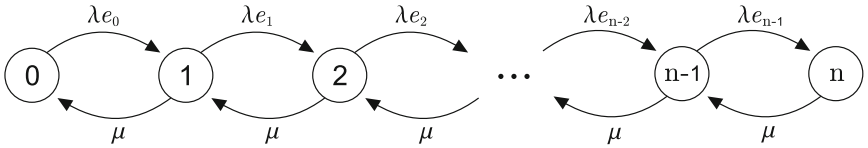


Fig. 2. The transition-rate diagram for an $M/M/1/n$ system with balking.

We let $Y(t)$ denote the number of customers in the system at time t. A typical sample path and the transition-rate diagram of $Y(t)$ are given in Figures 1 and 2, respectively. In Fig. 1, the sequence $(\tau_n)_{n\in\mathbb{N}}$ is the sequence of arrival or

departure instants. As the interarrival and the service times have exponential distributions, the process $Y = \{Y(t), t \geq 0\}$ is a continuous time Markov chain with state space $E = \{0, 1, \ldots, n\}$, rate transition matrix $R = (r_{ij})_{i,j \in E}$ with

$$r_{ij} = \begin{cases} \lambda e_i, & j = i+1 \\ \lambda(1-e_i), & j = i \\ \mu, & j = i-1 \\ 0, & j \in E \backslash \{i-1, i, i+1\} \end{cases},$$

and vector of transition rates out of states $r = (r_i)_{i \in E}$ with $r_i = \lambda + \mu \mathbf{1}_{\{i>0\}}$, where $\mathbf{1}_A$ denotes the indicator function of statement A. As a consequence, the process $\bar{Y} = (Y(\tau_k))_{k \in \mathbb{N}}$ embedded at the sequence $(\tau_k)_{k \in \mathbb{N}}$ of instants of arrival or departure of customers is a DTMC with state space E and transition probability matrix $P = (p_{ij})_{i,j \in E}$, such that

$$p_{ij} = \begin{cases} e_0 & j = 1 \wedge i = 0 \\ 1 - e_0 & j = i = 0 \\ \frac{\lambda e_i}{\lambda+\mu}, & j = i+1 \wedge i \neq 0 \\ \frac{\lambda(1-e_i)}{\lambda+\mu}, & j = i \neq 0 \\ \frac{\mu}{\lambda+\mu}, & j = i-1 \wedge i \neq 0 \\ 0, & j \in E \backslash \{i-1, i, i+1\} \end{cases}. \tag{1}$$

3 Probability Function of (S_i, L_i)

In this section we derive the joint probability function of the random vector (S_i, L_i) where, for $i \in E \backslash \{0\}$, S_i denotes the number of customers served during an i-busy-period and L_i denotes the number of customers lost during an i-busy-period.

To this purpose, we introduce the probability-generating function of (S_i, L_i),

$$g_i(u, v) = E(u^{S_i} v^{L_i}) = \sum_{s \in \mathbb{N}} \sum_{l \in \mathbb{N}_0} u^s v^l \, P(S_i = s, L_i = l) \tag{2}$$

with $|u| \leq 1$ and $|v| \leq 1$, from which, by derivation, we obtain the probability function of (S_i, L_i),

$$P(S_i = s, L_i = l) = \frac{1}{s!\, l!} \left. \frac{\partial^{s+l} g_i(u, v)}{\partial u^s \partial v^l} \right|_{u=0, v=0}, \qquad (s, l) \in \mathbb{N} \times \mathbb{N}_0.$$

The following theorem presents properties of the probability-generating function of (S_i, L_i).

Theorem 1. *The probability-generating function of the number of customers served and the number of customers lost during an i-busy-period of an $M/M/1/n$ system with balking satisfies, for $i \in E\backslash\{0\}$, with the convention that $g_0(u,v) = \theta_{n+1}(u,v) = 1$,*

$$g_i(u,v) = \theta_i(u,v)\, g_{i-1}(u,v) \tag{3}$$

with

$$\theta_i(u,v) = \frac{u\mu}{\lambda + \mu - v\lambda(1-e_i) - \lambda e_i\, \theta_{i+1}(u,v)}. \tag{4}$$

Proof. Denoting by X the type of event that occurs at the instant τ_k of the first transition after starting a i-busy-period, we let

$$X = \begin{cases} -1, & \text{if a customer exits the system at } \tau_k \\ 0, & \text{if a customer arrives without entering the system at } \tau_k \\ 1, & \text{if a customer arrives and enters the system at } \tau_k \end{cases}.$$

Conditioning on the event $\{X = x\}$, and letting $=_{st}$ denote equality in distribution, we obtain

$$(S_n, L_n)|_{X=x} =_{st} \begin{cases} (1 + S_{n-1}, L_{n-1}), & x = -1 \\ (S_n, 1 + L_n), & x = 0 \end{cases}$$

and, for $i \in E\backslash\{0, n\}$,

$$(S_i, L_i)|_{X=x} =_{st} \begin{cases} (1 + S_{i-1}, L_{i-1}), & x = -1 \\ (S_i, 1 + L_i), & x = 0 \\ (S_{i+1}, L_{i+1}), & x = 1 \end{cases}.$$

From the total probability law, it follows that, for $(s,l) \in \mathbb{N} \times \mathbb{N}_0$,

$$\begin{aligned} P(S_n = s, L_n = l) =& P(S_{n-1} = s-1, L_{n-1} = l)P(X = -1) \\ &+ P(S_n = s, L_n = l-1)P(X = 0) \end{aligned} \tag{5}$$

and, for $i \in E\backslash\{0, n\}$,

$$\begin{aligned} P(S_i = s, L_i = l) =& \mathbf{1}_{\{(i,s,l)=(1,1,0)\}} P(X = -1) \\ &+ \mathbf{1}_{\{i>1\}} P(S_{i-1} = s-1, L_{i-1} = l)P(X = -1) \\ &+ P(S_i = s, L_i = l-1)P(X = 0) \\ &+ P(S_{i+1} = s, L_{i+1} = l)P(X = 1). \end{aligned} \tag{6}$$

We will now prove the statement of the theorem using induction, starting with the case $i = n$. Taking into account (1), (2) and (5), we obtain:

$$g_n(u,v) = \frac{u\mu}{\lambda + \mu} g_{n-1}(u,v) + \frac{v\lambda}{\lambda + \mu} g_n(u,v).$$

This implies that

$$g_n(u,v) = \theta_n(u,v) g_{n-1}(u,v)$$

with

$$\theta_n(u,v) = \frac{u\mu}{\lambda + \mu - v\lambda}.$$

Therefore, the Eqs. (3)–(4) are valid for $i = n$.

Let us now assume that the Eqs. (3)–(4) are valid for $i = j + 1$ with j being a fixed number such that $1 < j < n$. Taking into account (1), (2) and (6), we obtain:

$$g_j(u,v) = \frac{u\mu}{\lambda+\mu} g_{j-1}(u,v) + \frac{v\lambda(1-e_j)}{\lambda+\mu} g_j(u,v) + \frac{\lambda e_j}{\lambda+\mu} g_{j+1}(u,v).$$

Thus,

$$g_j(u,v) = \frac{u\mu}{\lambda + \mu - v\lambda(1-e_j)}\, g_{j-1}(u,v) + \frac{\lambda e_j}{\lambda + \mu - v\lambda(1-e_j)}\, g_{j+1}(u,v).$$

Now, as the induction hypothesis gives $g_{j+1}(u,v) = \theta_{j+1}(u,v) g_j(u,v)$, we obtain:

$$g_j(u,v) = \frac{u\mu}{\lambda + \mu - v\lambda(1-e_j)}\, g_{j-1}(u,v) + \frac{\lambda e_j}{\lambda + \mu - v\lambda(1-e_j)}\, \theta_{j+1}(u,v) g_j(u,v).$$

Thus, isolating $g_j(u,v)$, we obtain Eqs. (3)–(4) for $i = j$. Therefore, by induction, the statement of the theorem is valid.

We conclude this section by presenting, in Fig. 3, a recursive procedure to obtain the probability generating function of (S_i, L_i) in $M/M/1/n$ systems with balking, departing from $g_0(u,v) = 1$.

4 Numerical Illustration

In this section, using the above-derived results, we compute the joint probability function of the number of customers served and the number of customers lost in busy-periods of $M/M/1/n$ systems with unit service rate ($\mu = 1$) and the following three customer balking policies.

i. Partial blocking:

$$e^{(1)} = (e_0, e_1, \ldots, e_n) \text{ with } e_i = \begin{cases} 1 & i \in E \backslash \{n\} \\ 0 & i = n \end{cases}$$

that represents the standard partial blocking policy of $M/M/1/n$ systems in which a customer is blocked if and only if the system is full at is arrival to the system.

Input: n, λ, μ, $e = (e_0, e_1, \ldots, e_{n-1}, 0)$

$\theta_n(u,v) = \dfrac{u\mu}{\lambda + \mu - v\lambda}$;

For $i = n-1 : -1 : 1$

$$\theta_i(u,v) = \frac{u\mu}{\lambda + \mu - v\lambda(1-e_i) - \lambda e_i \theta_{i+1}(u,v)};$$

End for

$g_1(u,v) = \theta_1(u,v)$;

For $i = 2 : n$

$$g_i(u,v) = \theta_i(u,v)\, g_{i-1}(u,v);$$

End for

Output: $(g_i(u,v))_{i=1,2,\ldots,n}$

Fig. 3. Algorithm for computing the probability-generating function of (S_i, L_i) in $M/M/1/n$ systems with balking.

ii. Increasing balking probabilities:

$$e^{(2)} = (e_0, e_1, \ldots, e_n) \text{ with } e_i = \begin{cases} \frac{1}{i+2} & i \in E\backslash\{n\} \\ 0 & i = n \end{cases}$$

in which the balking of customers increases with the size of the queue, that is frequent whenever the customers are in a hurry to be served and tend not to enter the system if they have to wait long periods of time.

iii. Decreasing balking probabilities:

$$e^{(3)} = (e_0, e_1, \ldots, e_n) \text{ with } e_i = \begin{cases} \frac{1}{n+1-i} & i \in E\backslash\{n\} \\ 0 & i = n \end{cases}$$

a reverse balking that favors the entry of customers whenever the system has greater demand and may occur, for instance, in investments in the stock market.

Tables 1 and 2 show the joint probability function of the number of customers served and the number of customers lost in a 1-busy-period of $M/M/1/7$ systems with balking policy $e^{(2)} = (\frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{8}, 0)$ and arrival rates $\lambda = 0.5$ and $\lambda = 1.1$, respectively. In the system with low traffic intensity (Table 1), as the arrival rate is much lower than the service rate, the queue has a small trend to fill up and to have a high number of losses. Consequently, the higher joint probabilities are associated to the lowest values of the number of customers served and of the number of customers lost. In fact, we observe that $P(S_1 \leq 3, L_1 \leq 3) = 0.9831$.

Table 1. Joint probability function of the number of customers served and the number of customers lost in 1-busy-period of a $M/M/1/7$ system with service rate $\mu = 1$ and arrival rate $\lambda = 0.50$, $P(S_1 = s, L_1 = l)$ for $s = 1, \ldots, 9$ and $l = 0, 1, \ldots, 9$.

$$\begin{bmatrix}
0.6667 & 0.1481 & 0.0329 & 0.0073 & 0.0016 & 0.0004 & 0.0001 & 0.0000 & 0.0000 & 0.0000 \\
0.0494 & 0.0343 & 0.0159 & 0.0061 & 0.0021 & 0.0007 & 0.0002 & 0.0001 & 0.0000 & 0.0000 \\
0.0064 & 0.0076 & 0.0054 & 0.0030 & 0.0014 & 0.0006 & 0.0002 & 0.0001 & 0.0000 & 0.0000 \\
0.0010 & 0.0016 & 0.0015 & 0.0011 & 0.0007 & 0.0004 & 0.0002 & 0.0001 & 0.0000 & 0.0000 \\
0.0001 & 0.0003 & 0.0004 & 0.0004 & 0.0003 & 0.0002 & 0.0001 & 0.0001 & 0.0000 & 0.0000 \\
0.0000 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000
\end{bmatrix}$$

In contrast, the system with higher utilization rate (Table 2) has a greater tendency to fill up and to experience more balking and customers blocked. In this case, we observe that $P(S_1 \leq 3, L_1 \leq 3) = 0.8933$ as higher values of the number of customers served and of the number of customers lost still have positive joint probabilities.

Table 2. Joint probability function of the number of customers served and the number of customers lost in 1-busy-period, of a $M/M/1/7$ system with service rate $\mu = 1$ and arrival rate $\lambda = 1.10$, $P(S_1 = s, L_1 = l)$ for $s = 1, \ldots, 9$ and $l = 0, 1, \ldots, 9$.

$$\begin{bmatrix}
0.4762 & 0.1663 & 0.0581 & 0.0203 & 0.0071 & 0.0025 & 0.0009 & 0.0003 & 0.0001 & 0.0000 \\
0.0396 & 0.0432 & 0.0315 & 0.0191 & 0.0104 & 0.0053 & 0.0026 & 0.0012 & 0.0006 & 0.0003 \\
0.0058 & 0.0107 & 0.0120 & 0.0105 & 0.0078 & 0.0053 & 0.0033 & 0.0019 & 0.0011 & 0.0006 \\
0.0010 & 0.0025 & 0.0039 & 0.0044 & 0.0042 & 0.0035 & 0.0027 & 0.0019 & 0.0013 & 0.0008 \\
0.0002 & 0.0006 & 0.0011 & 0.0016 & 0.0018 & 0.0018 & 0.0016 & 0.0014 & 0.0010 & 0.0008 \\
0.0000 & 0.0001 & 0.0003 & 0.0005 & 0.0007 & 0.0008 & 0.0008 & 0.0008 & 0.0007 & 0.0006 \\
0.0000 & 0.0000 & 0.0001 & 0.0001 & 0.0002 & 0.0003 & 0.0004 & 0.0004 & 0.0004 & 0.0003 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0001 & 0.0001 & 0.0001 & 0.0002 & 0.0002 & 0.0002 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0001 & 0.0001 & 0.0001 & 0.0001
\end{bmatrix}$$

Figure 4 illustrates the sensitivity of the joint distribution function of (S_1, L_1) at point $(3, 3)$ of $M/M/1/10$ systems, as function of the utilization rate ($\rho = \lambda/1$), considering three customer balking policies: $e^{(1)} = (1, 1, \ldots, 1, 0)$, $e^{(2)} = (\frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{11}, 0)$ and $e^{(3)} = (\frac{1}{11}, \frac{1}{10}, \ldots, \frac{1}{2}, 0)$. As expected, for the three balking policies, the joint distribution at the considered point decreases as the utilization rate increases. This decrease is more highlighted in the traditional $M/M/1/10$ systems, that is the system with $e^{(1)}$ policy, where customers always

enter in the system while it is not full up. Among the $M/M/1/10$ systems considered, the systems with balking, and in particular the system with $e^{(3)}$ policy (reverse balking), show higher joint distribution of (S_1, L_1) at $(3, 3)$ in contrast to the traditional $M/M/1/10$ system.

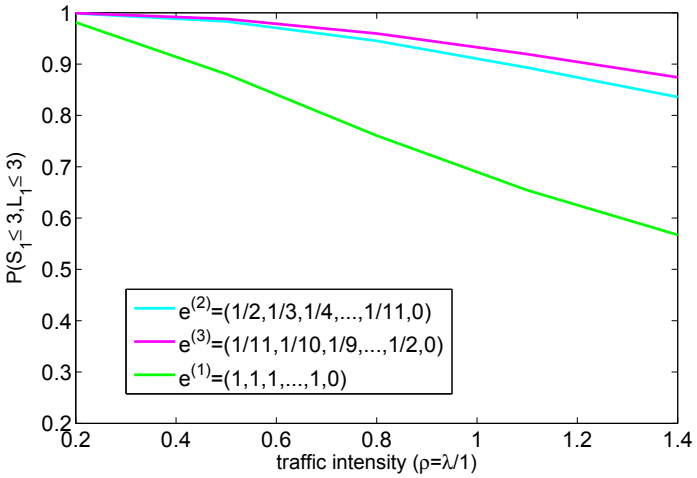


Fig. 4. Joint probability function of the number of customers served and of the number of customers lost in a 1-busy-period of a $M/M/1/10$ system with $\mu = 1$, $P(S_1 \leq 3, L_1 \leq 3)$, as a function of the utilization rate.

5 Conclusion

The number of customers served and the number of customers lost during busy-periods are important queueing performance measures. Although the analysis of their joint behavior constitutes an added value for the characterization of the queueing system, to our knowledge, the joint probability distribution of such measures has not been derived in the literature.

In this work the joint probability function of the number of customers served and the number of customers lost in busy-periods was obtained for single server Markovian queueing systems with finite capacity and balking. In particular, we derived a recursive procedure to compute the probability-generating function of the number of customers served and the number of customers lost in busy-periods that can be initiated by multiple customers in the system.

The derived recursion was applied to numerically compute the joint probability function of the number of customers served and the number of customers lost in busy-periods of $M/M/1/n$ systems under partial blocking and monotonic (increasing/decreasing) customer balking policies.

The approach followed in the paper can be generalized to obtain the joint probability distribution of the number of customers served and the number of

customers lost during busy-periods of Markovian queues with batch arrivals and balking and reneging.

References

1. Al Hanbali, A.: Busy period analysis of the level dependent $PH/PH/1/K$ queue. Queueing Syst. **67**(3), 221–249 (2011). https://doi.org/10.1007/s11134-011-9213-6
2. Ancker, C.J., Gafarian, A.V.: Some queuing problems with balking and reneging II. Oper. Res. **11**(6), 928–937 (1963)
3. Ferreira, F., Pacheco, A.: Analysis of $GI^X/M(n)//N$ systems with stochastic customer acceptance policy. Queueing Syst. **58**, 29–55 (2008). https://doi.org/10.1007/s11134-007-9057-2
4. Guha, D., Goswami, V., Banik, A.D.: Algorithmic computation of steady-state probabilities in an almost observable $GI/M/c$ queue with or without vacations under state dependent balking and reneging. Appl. Math. Model. **40**, 4199–4219 (2016)
5. Haight, F.A.: Queuing with balking I. Biometriika **44**(3–4), 360–369 (1957)
6. Harris, T.J.: The remaining busy period of a finite queue. Oper. Res. **19**, 219–223 (1971)
7. Jain, N.K., Kumar, R., Som, B.K.: An $M/M/1/N$ queuing system with reverse balking. Am. J. Oper. Res. **4**(2), 17–20 (2014)
8. Kumar, R., Sharma, S.: Transient analysis of an $M/M/c$ queuing system with balking and retention of reneging customers. Commun. Stat. Theory Methods **47**(6), 1318–1327 (2018)
9. Kumar, R., Som, B.K.: An $M/M/1/N$ queuing system with reverse balking and reverse reneging. AMO-Adv. Model. Optim. **16**(2), 339–353 (2014)
10. Laxmi, P.V., Gupta, U.C.: Analysis of finite-buffer multi-server queues with group arrivals: $GI^X/M/c/N$. Queueing Syst. **36**(1–3), 125–140 (2000). https://doi.org/10.1023/A:1019179119172
11. Peköz, E.A., Righter, R., Xia, C.H.: Characterizing losses during busy periods in finite buffer systems. J. Appl. Probab. **40**(1), 242–249 (2003)
12. Som, B.K., Kumar, R.: A heterogeneous queuing system with reverse balking and reneging. J. Ind. Prod. Eng. **35**(1), 1–5 (2018)
13. Wang, K., Li, N., Jiang, Z.: Queuing system with impatient customers: a review. In: IEEE International Conference on Service Operations and Logistics and Informatics, 15–17 July 2010, Shandong, pp. 82–87 (2010)
14. Wang, Q., Zhang, B.: Analysis of a busy period queuing system with balking, reneging and motivating. Appl. Math. Model. **64**, 480–488 (2018)
15. Yue, D., Zhang, Y., Yue, W.: Optimal performance analysis of an $M/M/1/N$ queue system with balking, reneging and server vacation. Int. J. Pure Appl. Math. **28**, 101–115 (2006)