Customer impatience in the $GI/M(k)/1/N$ queue with working vacation: Economic and performance analysis

Abdelhak Guendouzi[1,*] and **Allal Saadaoui**[1]

[1] *Institute of Science, University Center of Aflou, B. P. 603, 03001, Aflou, Algeria*
E-mail: ⟨{a.guendouzi, a.saadaoui}@cu-aflou.edu.dz⟩

Abstract. This study investigates a $GI/M(k)/1/N$ queueing system with Bernoulli feedback and multiple working vacation policy. The model incorporates customer impatience and accounts for the retention of reneged customers. By employing the supplementary variable technique and recursive methods, we derive the steady-state distributions of the customer count at both pre-arrival and arbitrary epochs. Various system performance measures are analyzed. In addition, an economic evaluation is carried out. Numerical illustrations are provided using the R software to confirm the analytical findings and highlight the influence of queueing parameters on performance indicators and system costs.

Keywords: The $GI/M(k)/1/N$ queue, impatient customers, working vacation, Bernoulli feedback, cost model

Received: October 18, 2024; accepted: May 24, 2025; available online: June 3, 2025

DOI: 10.17535/crorr.2026.0002

Original scientific paper.

1. Introduction

Queueing theory offers effective tools for analyzing and addressing congestion issues across numerous real-world and industrial domains. Applications span a broad spectrum, including emergency response systems, computing environments, customer service centers, online platforms, communication infrastructures, and various service facilities such as banks and governmental institutions [8, 11, 20, 17].

Queueing systems involving working vacations have been widely studied due to their operational importance in various sectors, such as manufacturing, service networks, transportation, telecommunications and computing systems. Unlike traditional vacation models, where the server becomes completely inactive, working vacation systems permit service continuation at a reduced rate during vacation periods. This approach enhances resource utilization efficiency and minimizes total system downtime. Given their practical significance, these systems have attracted considerable academic interest, beginning with the foundational work of Servi and Finn [19], followed by further developments by Jain and Agrawal [12], Baba [1, 2], Banik et al. [3], Bouchentouf et al. [5], among others.

Despite the advantages offered by these systems, customer impatience remains one of the major challenges affecting their performance. Prolonged waiting times may prompt customers to exhibit behaviors such as reneging—leaving the system before receiving service—or balking—deciding not to enter the system at all. Such behaviors can significantly reduce service efficiency and result in substantial revenue losses, making customer impatience a critical factor in queueing system design. In recent years, working vacation queues that incorporate customer impatience

[*] Corresponding author.

have been the subject of numerous analytical studies. For instance, Selvaraju and Goswami [18] discussed a $M/M/1$ queue by evaluating the impact of single versus multiple working vacation policies in the presence of both balking and reneging. Then, Laxmi and Jyothsna [15] investigated a $GI/M/1/N$ queue incorporating balking behavior and multiple working vacations. Building on this work, Goswami [10] undertook a systematic analysis of a $GI/M/1/N$ queueing model with multiple working vacations, emphasizing the joint effects of balking and reneging on performance metrics.

As mentioned earlier, customer impatience can result in the loss of potential clients and a decline in a firm's revenue. To mitigate this issue, businesses adopt various customer retention strategies with varying degrees of effectiveness. Retaining customers is particularly vital for organizations that frequently encounter impatience-related challenges. Companies employ different approaches, such as increasing service rates or introducing additional service channels, to encourage customers to remain in the system. For a comprehensive discussion on customer retention mechanisms, interested readers are referred to the works of Kumar and Sharma [14], Madheswari et al. [16], Yang and Wu [21], Bouchentouf et al. [4], among others. In real-life scenarios, customers may re-enter the system due to dissatisfaction with the initial service, as observed in supermarkets, hospitals, post offices, banks, and computer networks. The Bernoulli feedback mechanism models this behavior by describing situations in which customers, after receiving unsatisfactory service, either permanently exit the system or rejoin the queue with a certain probability. Additional insights can be found in [13, 6, 7, 9]. In the present paper, we perform a stationary analysis of a $GI/M(k)/1/N$ Bernoulli feedback queue model that incorporates multiple working vacations, state-dependent service rates, customer reneging due to impatience, and the retention of reneged customers. Our study investigates the complex interplay among these factors, which are common in practical service systems, yet rarely examined together in the queueing literature. This work makes a contribution in this regard. Using supplementary variable and recursive techniques, we derive steady-state probabilities at pre-arrival and arbitrary epochs. Furthermore, we present various system performance metrics and establish the economic model. Moreover, we carry out a numerical analysis to illustrate how queueing parameters influence key performance measures and economic characteristics.

The rest of this paper is structured as follows. Section 2 outlines the main assumptions and introduces the notation used throughout the model. Section 3 presents a detailed analytical study and derives the key performance metrics. Section 4 develops the cost evaluation framework. Numerical results and discussions are given in Section 5. Finally, Section 6 concludes the paper.

2. Description of the model

We consider a $GI/M(k)/1/N$ queueing model that incorporates Bernoulli feedback and customer impatience behaviors (balking and reneging). The service rates are state-dependent during normal busy periods and working vacations, which include multiple working vacations. The assumptions and notations introduced below will be used consistently throughout the paper.

- The inter-arrival times of successive arrivals are assumed to be independent and identically distributed (i.i.d.) random variables with cumulative distribution function $A(u)$, probability density function $a(u)$ for $u \geq 0$, Laplace-Stieltjes transform (LST) $A^*(s)$, and mean inter-arrival time $1/\lambda = -A^{*(1)}(0)$, where $h^{(1)}(0)$ denotes the first derivative of $h(s)$ evaluated at $s = 0$.
- The system's capacity is limited to N customers. When a customer arrives and finds k customers in the system, they join the queue with probability θ_k (where $\overline{\theta}_k = 1 - \theta_k$ is the balking probability). Specifically, $\theta_0 = 1$, $0 \leq \theta_{k+1} \leq \theta_k \leq 1$ for $1 \leq k \leq N-1$, and $\theta_N = 0$ to prevent joining when the buffer is full.

- Service is provided by a single server following a FIFO queue discipline. Service durations are exponentially distributed with parameters μ_k. When the server has no customers to serve and becomes idle, it proceeds to a working vacation period. The working vacation times follow an exponential distribution with mean of $1/\phi$. During a working vacation period, the server operates at a reduced service rate compared to regular periods, reflecting lower resource availability. The service durations during the working vacation period are also exponentially distributed with parameter ν_k, where $\nu_k < \mu_k$ for $1 \leq k \leq N$. During both the busy period and the working vacation period, the server can change its service rate when serving the same customer, depending on the system state. The average service rates are $\nu = \sum_{k=1}^{N} \nu_k/N$ for working vacations and $\mu = \sum_{k=1}^{N} \mu_k/N$ for regular busy periods.
- Upon entering the system, if a customer encounters an unavailable server—whether during normal busy periods or working vacations—they activate an impatience timer T, assumed to follow an exponential distribution with rate parameter ξ. If their service is not completed before the timer expires, they may abandon the queue with probability α and will never return to the system. Alternatively, with probability $\overline{\alpha} = 1 - \alpha$, they may be retained in the system through a retention mechanism. It is assumed that the impatience timers of all customers are independent of the number of customers waiting in the queue.
- After receiving unsatisfactory service—whether during the working vacation or normal periods—a customer may return to the system with probability $\overline{\beta}$ as a feedback customer for another regular service, or leave the system permanently with probability β, where $\overline{\beta} + \beta = 1$. No distinction is made between new arrivals and feedback arrivals.

All random variables involved in the system–namely, impatience timers, service times, working vacation periods and inter-arrival times–are considered to be mutually independent.

3. Analysis of the model

In this section, we study the system in steady-state using the supplementary variable technique. The state of the system at time t can be described by a continuous-time Markov process $\{(L(t), S(t), U(t)), t \geq 0\}$, with state space expressed as

$$\Omega = \{(0,0,u) \cup (k,0,u) \cup (k,1,u) : 1 \leq k \leq N, u \geq 0\},$$

where:

- $L(t)$ denotes the number of customers in the system at time t
- $U(t)$ denotes the remaining inter-arrival time until the next customer arrival
- $S(t)$ denotes the state of the server at time t, defined as

$$S(t) = \begin{cases} 0, & \text{if the server is in a working vacation period,} \\ 1, & \text{if the server is in a normal busy period.} \end{cases}$$

Let us define the joint probabilities as

$$P_{k,0}(u,t)du = \mathbb{P}(L(t) = k, u \leq U(t) < u + du, S(t) = 0), u \geq 0, 0 \leq k \leq N,$$

$$P_{k,1}(u,t)du = \mathbb{P}(L(t) = k, u \leq U(t) < u + du, S(t) = 1), u \geq 0, 1 \leq k \leq N.$$

Then, in steady-state, we have

$$P_{k,0}(u) = \lim_{t\to\infty} P_{k,0}(u,t), \;\; 0 \leq k \leq N, \quad P_{k,1}(u) = \lim_{t\to\infty} P_{k,1}(u,t), \;\; 1 \leq k \leq N.$$

3.1. Governing equations

By employing the probabilistic arguments and using the remaining inter-arrival time as the supplementary variable, we observe the state of the system at two consecutive time epochs, t and $t+dt$. Taking $\lim_{t\to\infty}$ and after some simplification, we have the following set of differential difference equations.

$$-P^{(1)}_{0,0}(u) = \beta\mu_1 P_{1,1}(u) + \beta\nu_1 P_{1,0}(u), \tag{1}$$

$$\begin{aligned} -P^{(1)}_{k,0}(u) = &-[\beta\nu_k + \phi + (k-1)\alpha\xi]P_{k,0}(u) + [\beta\nu_{k+1} + k\alpha\xi]P_{k+1,0}(u) \\ &+a(u)[\theta_{k-1}P_{k-1,0}(0) + \overline{\theta}_k P_{k,0}(0)] \;\;, 1 \le k \le N-1, \end{aligned} \tag{2}$$

$$-P^{(1)}_{N,0}(u) = -[\beta\nu_N + \phi + (N-1)\alpha\xi]P_{N,0}(u) + a(u)[\theta_{N-1}P_{N-1,0}(0) + P_{N,0}(0)], \tag{3}$$

$$-P^{(1)}_{1,1}(u) = -\beta\mu_1 P_{1,1}(u) + \phi P_{1,0}(u) + (\beta\mu_2 + \alpha\xi)P_{2,1}(u) + a(u)\overline{\theta}_1 P_{1,1}(0), \tag{4}$$

$$\begin{aligned} -P^{(1)}_{k,1}(u) = &-[\beta\mu_k + (k-1)\alpha\xi]P_{k,1}(u) + [\beta\mu_{k+1} + k\alpha\xi]P_{k+1,1}(u) \\ &+\phi P_{k,0}(u) + a(u)[\theta_{k-1}P_{k-1,1}(0) + \overline{\theta}_k P_{k,1}(0)], \;\; 2 \le k \le N-1, \end{aligned} \tag{5}$$

$$\begin{aligned} -P^{(1)}_{N,1}(u) = &-[\beta\mu_N + (N-1)\alpha\xi]P_{N,1}(u) + \phi P_{N,0}(u) \\ &+a(u)[\theta_{N-1}P_{N-1,1}(0) + P_{N,1}(0)], \end{aligned} \tag{6}$$

where $P_{k,0}(0), 0 \le k \le N$ and $P_{k,1}(0), 1 \le k \le N$ are the respective rate probabilities with the remaining inter-arrival time equal to zero denoting that an arrival is about to occur. Then, we introduce the following Laplace-Stieltjes transforms of the steady-state probabilities as

$$P^*_{k,0}(s) = \int_0^\infty e^{-su}P_{k,0}(u)du, \;\; 0 \le k \le N, \;\; P^*_{k,1}(s) = \int_0^\infty e^{-su}P_{k,1}(u)du, \;\; 1 \le k \le N.$$

Let $P_{k,0} = P^*_{k,0}(0), 0 \le k \le N$ and $P_{k,1} = P^*_{k,1}(0), 1 \le k \le N$, be the steady-state probabilities of k customers in the system when the server is in state $j = 0, 1$ at an arbitrary epoch.
Let $P^-_{k,0}, 0 \le k \le N$, and $P^-_{k,1}, 1 \le k \le N$, denote the steady-state probabilities at a pre-arrival epoch, that is, an arrival sees k customers in the system and the server is in state $j = 0, 1$, at the arrival epoch.

Theorem 1. The steady-state probabilities at an arbitrary epoch in terms of the steady-state probabilities at a pre-arrival epoch are respectively expressed as follows:

$$P_{N,0} = \frac{\lambda}{\eta_N}\theta_{N-1}P^-_{N-1,0},$$

$$P_{k,0} = \frac{\eta_{k+1} - \phi}{\eta_k}P_{k+1,0} + \frac{\lambda}{\eta_k}\left(\theta_{k-1}P^-_{k-1,0} - \theta_k P^-_{k,0}\right), k = N-1, \ldots, 1,$$

$$P_{N,1} = \frac{\phi}{\gamma_N}P_{N,0} + \frac{\lambda}{\gamma_N}\theta_{N-1}P^-_{N-1,1},$$

$$P_{k,1} = \frac{\gamma_{k+1}}{\gamma_k}P_{k+1,1} + \frac{\phi}{\gamma_k}P_{k,0} + \frac{\lambda}{\gamma_k}\left(\theta_{k-1}P^-_{k-1,1} - \theta_k P^-_{k,1}\right), k = N-1, \ldots, 2,$$

$$P_{1,1} = \frac{\gamma_2}{\gamma_1}P_{2,1} + \frac{\phi}{\gamma_1}P_{1,0} - \frac{\lambda}{\gamma_1}\theta_1 P^-_{1,1},$$

where $P_{0,0}$ can be computed by using the normalization condition, that is,

$$P_{0,0} = 1 - \left(\sum_{k=1}^{N} P_{k,0} + \sum_{k=1}^{N} P_{k,1} \right),$$

with $\eta_k = \beta\nu_k + \phi + (k-1)\alpha\xi$ and,$\gamma_k = \beta\mu_k + (k-1)\alpha\xi$, for $1 \le k \le N$.

Proof. The steady-state probabilities at arbitrary epochs are obtained by following the same analytical approach adopted in [3, 15, 10], which is based on the supplementary variable technique combined with recursive methods. □

3.2. Measures of performance

In this subsection, we derive and outline several important performance metrics for the proposed model, based on the steady-state probabilities at arbitrary epochs as established in the preceding subsection. These performance measures are as follows:

- The probabilities that the server is idle during a working vacation period (Π_I), busy during a working vacation period (Π_W), and busy during a normal busy period (Π_B) are given by:

$$\Pi_I = P_{0,0}, \quad \Pi_W = \sum_{k=1}^{N} P_{k,0}, \quad \Pi_B = \sum_{k=1}^{N} P_{k,1}.$$

- The average number of customers in the system (L_s) and the average number of customers in the queue (L_q) are given by:

$$L_s = \sum_{k=1}^{N} k(P_{k,0} + P_{k,1}), \quad L_q = \sum_{k=2}^{N} (k-1)(P_{k,0} + P_{k,1}).$$

- The expressions for the average balking rate (B_r) and number of customers served (S_c) are as follows:

$$B_r = \sum_{k=1}^{N} \lambda\overline{\theta}_k(P_{k,0} + P_{k,1}), \quad S_c = \beta\sum_{k=1}^{N} \nu_k P_{k,0} + \beta\sum_{k=1}^{N} \mu_k P_{k,1}.$$

- The expressions for the average reneging rate (R_r) and retention rate (R_e) are as follows:

$$R_r = \sum_{k=1}^{N} (k-1)\alpha\xi(P_{k,0} + P_{k,1}), \quad R_e = \sum_{k=1}^{N} (k-1)\overline{\alpha}\xi(P_{k,0} + P_{k,1}).$$

Remark. Under the assumption of state-independent service rates ($\mu_k = \mu$ and $\nu_k = \nu$ for all $k = 1, \dots, N$), with $\beta = 1$ (implying no feedback customers) and $\alpha = 1$ (indicating no retention of reneged customers), our model coincides with the queueing system examined by Goswami [10]. Furthermore, if we additionally assume that customers are patient ($\xi = 0$), our model aligns with the framework introduced by Laxmi and Jyothsna [15]. Finally, by further imposing the absence of balking behavior ($\theta_k = 1$ for all $0 \le k \le N-1$), our model simplifies to the system studied by Banik et al. [3].

4. Economic model

To construct the economic model, we consider the following cost (in unit) elements associated with different events:

- Costs incurred based on the server's operational state and waiting customers:

- C_1: Cost when the server is idle during a working vacation period.

- C_2: Cost when the server is busy during a working vacation period.

- C_3: Cost when the server is busy during a normal busy period.

- C_4: Cost when customers join the queue and wait for service.

- Costs associated with impatient customers:

- C_5: Cost when a customer balks. - C_6: Cost when a customer reneges.

- C_7: Cost when a reneging customer is retained.

- Service-related costs:

- C_8: Cost per service. - C_9: Cost of serving a feedback customer.

Let R represent the revenue earned per serviced customer. The economic measures of the system are defined as follows:

- The total expected cost incurred by the system per unit time (T_C) is computed as follows:

$$T_C = C_1\Pi_I + C_2\Pi_W + C_3\Pi_B + C_4L_q + C_5B_r + C_6R_r + C_7R_e + (\mu + \nu)(C_8 + \overline{\beta}C_9).$$

- The total expected revenue generated by the system per unit time (T_R) and the corresponding total expected profit (T_P) are computed as follows:

$$T_R = R \times S_c, \qquad T_P = T_R - T_C.$$

5. Numerical illustrations and discussion

This section presents numerical examples for the state-dependent service rates and multiple working vacations in a $GI/M(k)/1/N$ queue. Various inter-arrival time distributions, including deterministic (D) and Erlang-3 (E_3), are considered. The outcomes, illustrated through graphs and tables, reveal the influence of changing system parameters on key performance metrics such as total expected cost, total expected revenue, and total expected profit. To achieve this, we developed an R program, authored by us, to demonstrate the practical applicability of the formulas derived in the previous sections. For this numerical study, we selected arbitrary values for different system parameters and costs. The cost elements used in the study are $C_1 = 5$, $C_2 = 7$, $C_3 = 10$, $C_4 = 10$, $C_5 = 4$, $C_6 = 4$, $C_7 = 8$, $C_8 = 10$, $C_9 = 5$, and $R = 100$. The following cases are considered:

- $\theta_k = 1 - \dfrac{k}{N}$, $\mu_k = 1.2k$, $\nu_k = 0.8k$, $\phi = 4.2$, $\overline{\beta} = 0.3$, $\overline{\alpha} = 0.4$, and $N = 8$. This case is presented in Table 1 and Figure 1.

- $\theta_k = e^{-0.1k}$, $\mu_k = 1.4k$, $\nu_k = 0.7k$, $\phi = 5.0$, $\lambda = 3.0$, $\xi = 0.8$, and $N = 9$. This case is presented in Figure 2 and Table 2.

		With balking			Without balking		
		$\lambda = 2.6$	$\lambda = 2.9$	$\lambda = 3.2$	$\lambda = 2.6$	$\lambda = 2.9$	$\lambda = 3.2$
$\xi = 0.6$	Π_I	0.186887	0.162467	0.142026	0.133298	0.107545	0.086755
	Π_W	0.105808	0.102906	0.099548	0.075771	0.068406	0.061075
	Π_B	0.707305	0.734628	0.758426	0.790931	0.824049	0.852169
	L_q	0.920353	1.049017	1.175376	1.418935	1.654920	1.892920
	B_r	0.563376	0.683875	0.813340	0.005055	0.010301	0.019089
	R_r	0.331327	0.377646	0.423135	0.510817	0.595771	0.681451
	R_e	0.220885	0.251764	0.282090	0.340545	0.397181	0.454301
	T_C	126.797565	129.129316	131.447604	131.583395	134.907988	138.248770
$\xi = 0.9$	Π_I	0.203947	0.179824	0.159471	0.152830	0.126448	0.104680
	Π_W	0.115384	0.113815	0.111691	0.086810	0.080370	0.073639
	Π_B	0.680669	0.706361	0.728838	0.760360	0.793182	0.821681
	L_q	0.820113	0.935488	1.049293	1.228458	1.432768	1.640300
	B_r	0.525254	0.636428	0.755929	0.002580	0.005380	0.010211
	R_r	0.442861	0.505164	0.566618	0.663367	0.773695	0.885762
	R_e	0.295241	0.336776	0.377745	0.442245	0.515797	0.590508
	T_C	126.569633	128.874890	131.172654	130.961746	134.197009	137.466633
$\xi = 1.2$	Π_I	0.218805	0.195175	0.175147	0.170192	0.143644	0.121390
	Π_W	0.123704	0.123443	0.122579	0.096603	0.091233	0.085331
	Π_B	0.657492	0.681382	0.702273	0.733204	0.765123	0.793279
	L_q	0.739351	0.843670	0.946880	1.081576	1.260746	1.443499
	B_r	0.494177	0.597580	0.708693	0.001392	0.002957	0.005717
	R_r	0.532333	0.607443	0.681754	0.778735	0.907737	1.039320
	R_e	0.354888	0.404962	0.454502	0.519157	0.605158	0.692880
	T_C	126.373521	128.650281	130.923132	130.448754	133.599580	136.795234

Table 1: *The impact of ξ and λ on system characteristics in the $D/M(k)/1/N$ queue model.*

Table 1 and Figure 1 illustrate the influence of the arrival rate λ and the reneging rate ξ on several key performance measures, as well as on the total expected cost, revenue, and profit, under the assumption of a deterministic inter-arrival time distribution.

- When the reneging rate ξ is fixed, an increase in the arrival rate λ results in higher values of the average number of customers in the system L_s and in the queue L_q, the probability that the server is in a normal busy state Π_B, the average number of customers served S_c, and the average rates of balking B_r, retention R_r, and reneging R_e. Conversely, the probabilities that the server is idle Π_I or working during a vacation period Π_W decrease. This is because a higher arrival rate increases the system congestion, which leads to greater impatience among customers and consequently a rise in reneging. Implementing customer retention strategies may help retain more customers in the system, thereby improving the average retention rate. Moreover, as L_s increases with λ, the average balking rate also rises. In addition, a higher λ increases the number of served customers, which in turn leads to higher values of total expected cost T_C, total expected revenue T_R, and total expected profit T_P.

- For a fixed arrival rate λ, increasing the reneging rate ξ causes decreases in B_r, L_q, L_s, S_c, and Π_B. Meanwhile, the average reneging and retention rates R_r and R_e, along with the probabilities Π_I and Π_W, increase. This behavior is due to the fact that a higher reneging rate reduces the average number of customers in the system and shortens waiting times, which leads to fewer customers choosing to wait (balking), and consequently lowers the average balking rate. However, the increase in customer loss caused by impatience reduces the total expected revenue T_R and total expected cost T_C, which ultimately results in a lower total expected profit T_P. The results conclusively demonstrate that customer

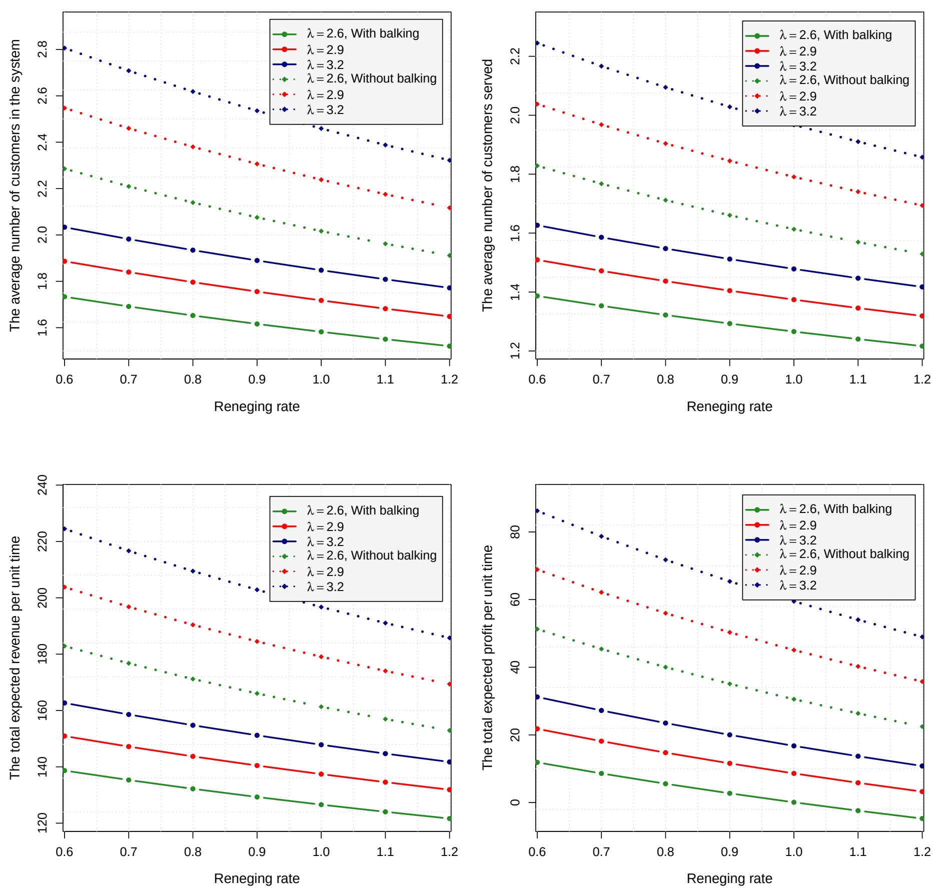

Figure 1: The impact of ξ and λ on L_s, S_c, T_R and T_P.

impatience significantly degrades the system's economic performance metrics, confirming its substantial adverse effect on operational profitability.

Figure 2 and Table 2 illustrate the influence of the feedback probability $\overline{\beta}$ and the retention probability of reneged customers $\overline{\alpha}$ on various system performance indicators, as well as on the total expected cost, revenue, and profit, assuming that the inter-arrival times follow an Erlang-3 distribution. The main observations are as follows:

- When $\overline{\beta}$ is held constant, an increase in $\overline{\alpha}$ leads to growth in the average number of customers present in the system L_s and in the queue L_q, the average balking rate B_r, and the average retention rate R_e. In contrast, the average reneging rate R_r decreases, which aligns with expectations. Consequently, the probability that the server is operating in a regular busy period Π_B increases, while the probabilities of the server being idle Π_I or working during a vacation period Π_W decrease. These changes also result in an increase in the average number of customers served S_c. As $\overline{\alpha}$ increases, the total expected cost T_C, total expected revenue T_R, and total expected profit T_P all rise. Thus, a higher retention probability positively contributes to both operational efficiency and profitability.

- When $\overline{\alpha}$ is fixed, raising the feedback probability $\overline{\beta}$ results in increases in L_s, L_q, B_r, R_r, R_e, and S_c, as would be intuitively anticipated. This causes a drop in the probabilities

		With balking			Without balking		
		$\overline{\alpha} = 0.3$	$\overline{\alpha} = 0.5$	$\overline{\alpha} = 0.7$	$\overline{\alpha} = 0.3$	$\overline{\alpha} = 0.5$	$\overline{\alpha} = 0.7$
$\overline{\beta} = 0.3$	Π_I	0.186502	0.172795	0.157503	0.149592	0.134941	0.118978
	Π_W	0.104780	0.097120	0.088562	0.084230	0.076013	0.067049
	Π_B	0.708718	0.730085	0.753935	0.766178	0.789046	0.813974
	L_q	0.948604	1.043480	1.158561	1.244632	1.380842	1.545851
	B_r	0.471158	0.500069	0.535518	0.046996	0.065215	0.092240
	R_r	0.531218	0.417392	0.278055	0.696994	0.552337	0.371004
	R_e	0.227665	0.417392	0.648794	0.298712	0.552337	0.865677
	T_C	108.595008	110.813442	113.502055	111.336317	114.319574	117.965873
$\overline{\beta} = 0.5$	Π_I	0.098580	0.085862	0.072150	0.072766	0.060514	0.047900
	Π_W	0.056439	0.049173	0.041333	0.041727	0.034712	0.027485
	Π_B	0.844980	0.864964	0.886517	0.885506	0.904774	0.924616
	L_q	1.306120	1.461307	1.654829	1.707834	1.920901	2.180379
	B_r	0.590257	0.637761	0.699348	0.106555	0.153162	0.224286
	R_r	0.731427	0.584523	0.397159	0.956387	0.768360	0.523291
	R_e	0.313469	0.584523	0.926704	0.409880	0.768360	1.221012
	T_C	122.068469	125.476554	129.738197	125.995139	130.510280	136.115247
$\overline{\beta} = 0.7$	Π_I	0.034778	0.026437	0.018263	0.022303	0.015592	0.009653
	Π_W	0.020289	0.015426	0.010659	0.013024	0.009107	0.005639
	Π_B	0.944933	0.958137	0.971078	0.964673	0.975300	0.984707
	L_q	1.829330	2.088379	2.418199	2.355374	2.674136	3.050476
	B_r	0.757664	0.845329	0.968873	0.254126	0.375068	0.559279
	R_r	1.024425	0.835352	0.580368	1.319009	1.069654	0.732114
	R_e	0.439039	0.835352	1.354191	0.565290	1.069654	1.708267
	T_C	137.924211	143.335867	150.314194	143.443008	150.197202	158.496285

Table 2: *The impact of $\overline{\beta}$ and $\overline{\alpha}$ on system characteristics in the $E_3/M(k)/1/N$ queue model.*

Π_I and Π_W, and a corresponding rise in Π_B. Furthermore, increases in $\overline{\beta}$ translate into higher values of T_C, T_R, and T_P. These observations are in full agreement with real-world behavior, indicating that higher feedback probabilities strengthen the economic performance of the system.

Analyzing the data presented in Tables 1–2 and Figures 1–2, it is evident that the performance indicators Π_B, L_s, L_q, R_r, R_e, S_c, as well as T_C, T_R, and T_P attain higher values in the absence of balking compared to the scenario with balking. In contrast, the metrics Π_I, Π_W, and B_r exhibit increased values when balking behavior is included. These findings are fully consistent with the expected theoretical behavior of the system.

6. Conclusions and future work

This article investigates a non-Markovian queueing system with a limited buffer, where service rates vary depending on the system state during both normal busy and working vacation periods. The model incorporates the behavior of impatient customers as well as a Bernoulli feedback, all under a multiple working vacation policy. The developed framework is applicable to a broad range of practical environments, including industries facing congestion challenges, such as customer service centers, production lines, and communication infrastructures. Steady-state equations and system size probabilities at pre-arrival and arbitrary epochs are derived using the supplementary variable technique and recursive methods. Key performance measures are obtained, and an economic model is established. Numerical experiments evaluate the sensitivity of the system's performance metrics and economic indicators to variations in key parameters. The model can be further extended to accommodate multiple servers, as well as

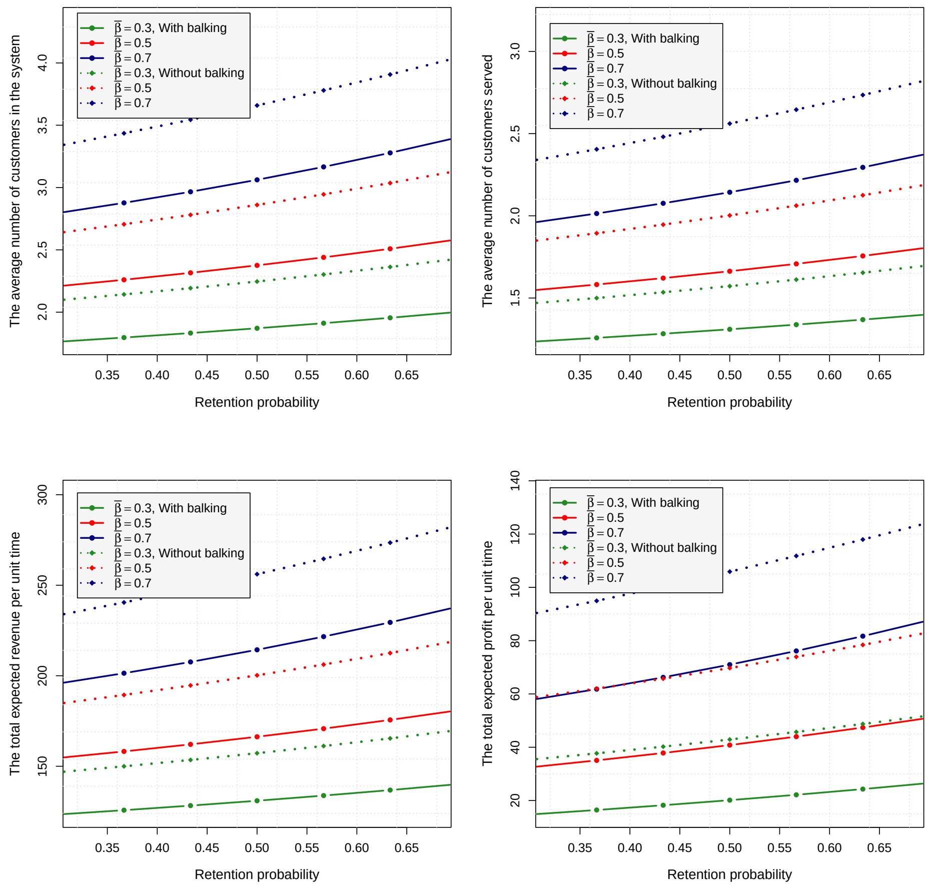

Figure 2: The impact of $\overline{\beta}$ and $\overline{\alpha}$ on L_s, S_c, T_R and T_P.

server breakdowns and repairs.

References

[1] Baba, Y. (2005). Analysis of a GI/M/1 queue with multiple working vacations. Operations Research Letters, 33(2), 201—209. doi: 10.1016/j.orl.2004.05.006

[2] Baba, Y. (2012). The M^X/M/1 queue with multiple working vacation. American Journal of Operations Research, 2(2), 217–224. doi: 10.4236/ajor.2012.22025

[3] Banik, A. D., Gupta, U. C., and Pathak, S. S. (2007). On the GI/M/1/N queue with multiple working vacations: analytic analysis and computation. Applied Mathematical Modelling, 31(9), 1701-–1710. doi: 10.1016/j.apm.2006.05.010

[4] Bouchentouf, A. A., Houalef, M., and Guendouzi, A. (2022). Analysis of a batch arrival multi-server queueing system with waiting servers, synchronous working vacations and impatient customers. Acta Universitatis Sapientiae, Mathematica, 14(1), 37-–60. doi: 10.2478/ausm-2022-0003

[5] Bouchentouf, A. A., Guendouzi, A., and Majid, S. (2020). On impatience in Markovian M/M/1/N/DWV queue with vacation interruption. Croatian Operational Research Review, 21-–37. doi: 10.17535/crorr.2020.0003

[6] Bouchentouf, A. A. and Guendouzi, A. (2019). Cost optimization analysis for an $M^X/M/c$ vacation queueing system with waiting servers and impatient customers. SeMA Journal, 76, 309—341. doi: 10.1007/s40324-018-0180-2

[7] Bouchentouf, A. A. and Guendouzi, A. (2021). Single server batch arrival Bernoulli feedback queueing system with waiting server, K-variant vacations and impatient customers. Operations Research Forum, 2, 1-–23. doi: 10.1007/s43069-021-00057-0

[8] Daigle, J. N. (2005). Queueing Theory with Applications to Packet Telecommunication. New York: Springer. doi: 10.1007/b99875

[9] Dehmi, A., Boualem, M., Kahla, S., and Berdjoudj, L. (2024). ANFIS computing and cost optimization of an M/M/c/M queue with feedback and balking customers under a hybrid hiatus policy. Croatian Operational Research Review, 15(2), 159-170. doi: 10.17535/crorr.2024.0013

[10] Goswami, V. (2015). Study of customers' impatience in a GI/M/1/N queue with working vacations. International Journal of Management Science and Engineering Management, 10(2), 144-–154. doi: 10.1080/17509653.2014.926229

[11] Haghighi, A. M., and Mishev, D. P. (2016). Delayed and Network Queues. New Jersey: John Wiley & Sons.

[12] Jain, M., and Agrawal, P. K. (2007). M/E_k/1 queueing system with working vacation. Quality Technology & Quantitative Management, 4(4), 455–470. doi: 10.1080/16843703.2007.11673165

[13] Kumar, R., Jain, N. K., and Som, B. K. (2014). Optimization of an M/M/1/N feedback queue with retention of reneged customers. Operations Research and Decisions, 24(3), 45-–58. doi: 10.5277/ord140303

[14] Kumar, R. and Sharma, S. K. (2012). M/M/1/N queueing system with retention of reneged customers. Pakistan Journal of Statistics and Operation Research, 8(4), 859—866. doi: 10.18187/pjsor.v8i4.408

[15] Laxmi, P. V. and Jyothsna, K. (2013). Analysis of finite buffer renewal input queue with balking and multiple working vacations. Opsearch, 50(4), 548—565. doi: 10.1007/s12597-013-0123-8

[16] Madheswari, S. P., Suganthi, P., and Josephine, S. A. (2016). Retrial queueing system with retention of reneging customers. International Journal of Pure and Applied Mathematics, 106(5), 11-–20. doi: 10.12732/ijpam.v106i5.2

[17] Ramesh, A., and Manoharan, M. (2025). Aspects of blocking on time-varying tandem queueing network. Croatian Operational Research Review, 16(2), 1–13. doi: 10.17535/crorr.2025.0016

[18] Selvaraju, N. and Goswami, C. (2013). Impatient customers in an M/M/1 queue with single and multiple working vacations. Computers & Industrial Engineering, 65(2), 207–215. doi: 10.1016/j.cie.2013.02.016

[19] Servi, L. D., and Finn, S. G. (2002). M/M/1 queues with working vacations (M/M/1/WV). Performance Evaluation, 50(1), 41–52. doi: 10.1016/S0166-5316(02)00057-3

[20] Sreelatha, V., Mamatha, E., and Reddy, C. S. (2024). Multi server queuing model with dynamic power shifting and performance efficiency factor. Croatian Operational Research Review, 15(2), 171–183. doi: 10.17535/crorr.2024.0014

[21] Yang, D.-Y. and Wu, Y.-Y. (2017). Analysis of a finite-capacity system with working breakdowns and retention of impatient customers. Journal of Manufacturing Systems, 44, 207—216. doi: 10.1016/j.jmsy.2017.05.010