Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2025, 33 (1) 10–26
http://journals.rudn.ru/miph

Research article

UDC 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2025-33-1-10-26
EDN: AMSSFO

Modeling and optimization of an $M/M/1/K$ queue with single working vacation, feedback, and impatience timers under N-policy

Abir Kadi[1], Mohamed Boualem[2], Nassim Touche[2], Aimen Dehimi[1]

[1] University of Bejaia, Laboratory of Applied Mathematics, 06000 Bejaia, Algeria

[2] University of Bejaia, Research Unit LaMOS (Modeling and Optimization of Systems), 06000 Bejaia, Algeria

(received: July 8, 2024; revised: August 10, 2024; accepted: August 25, 2024)

Abstract. This work presents an intensive study of a single server finite-capacity queueing model with impatience timers which depend on the server's states, feedback, and a single working vacation policy operating under an N-policy discipline. We examine the scenario where the server must wait for the number of customers to reach N to start a regular busy period; otherwise, the server will initiate a working vacation or switch to the dormant state if the number of customers increases. By applying the Markov recursive method, the steady-state probabilities were derived. Various performance metrics were visually depicted to assess diverse system parameter configurations. After constructing the expected cost function of the model, Grey Wolf Optimization (GWO) algorithm is utilized to determine the optimum values of the service rates μ^* and μ_v^*. Numerical examples are provided to validate the theoretical findings, offering insights into this intricate system.

Key words and phrases: Queueing system with impatience, N-policy, vacation policy, feedback, GWO algorithm, cost optimization

For citation: Kadi, A., Boualem, M., Touche, N., Dehimi, A. Modeling and optimization of an $M/M/1/K$ queue with single working vacation, feedback, and impatience timers under N-policy. *Discrete and Continuous Models and Applied Computational Science* **33** (1), 10–26. doi: 10.22363/2658-4670-2025-33-1-10-26. edn: AMSSFO (2025).

1. Introduction

Queueing theory is a branch of applied mathematics focused on studying and analyzing processes within a wide range of service, production, management, and communication systems. These systems involve repetitive occurrences of homogeneous events. Examples include consumer services, information reception, processing, and transmission systems, automated production lines, telecommunication networks, among others.

Queueing theory offers invaluable solutions to mitigate long queues in real-life settings, providing mathematical frameworks to address such challenges [1–4].

In contemporary times, research on queueing systems with impatient customers has garnered increased attention. Impatience emerges as a prominent characteristic, as customers often feel anxious and restless while awaiting services. Hence, it is imperative for queueing systems research to incorporate patron impatience to accurately reflect real-world conditions. Typically, customer dissatisfaction is modeled through concepts like balking and reneging. When faced with a long queue, customers may opt to leave (balk) or refrain from entering the system altogether if the queue is excessively lengthy. This behavior is observed in practical systems such as hospital emergency rooms, especially when dealing with critically ill patients [2, 5–7].

The concept of the N-policy was first introduced by Yadin and Naor [8]. Meena et al. [9] developed a non-Markovian system for machine repair problems with finite capacity and the N-policy, a variant of the vacation queueing system under Bernoulli feedback. Bouchentouf et al. [10] studied an $M/M/c$ queue with customers' impatience and Bernoulli feedback, incorporating a variant of multiple vacations. Kathirvel [11] addressed a finite single-server model with an optional second service and obtained the waiting time distribution of customers in the waiting hall using Laplace–Stieltjes transforms. Kadi et al. [12] analyzed an $M/M/2$ machine repairable model for both single and multiple vacations under the triadic policy using a matrix geometric method. Adou et al. [13] applied a comparative study between peremptive and non-preemptive scheduling algorithms for the operation of a wireless network slicing model. Additionally, Sharma et al. [14] examined the N-policy with the vacation interruption concept and impatience behavior. Rajadurai et al. [15] introduced an $M/G/1$ queueing system with multiple working vacations, vacation interruption with feedback, and server breakdowns, applying the results to Simple Mail Transfer Protocol (SMTP) applications. Furthermore, Boualem [16] utilized stochastic orders to analyze an $M/G/1$ queueing model where the server undergoes breakdown and repair processes. Hilquias et al. [17] compared *RED* and *TailDrop* algorithms with the renovation mechanism for different models of queueing systems. Goswami [18] investigated the interrelationship between F-policy and N-policy considering inter-arrival times of customers and geometrically distributed service times. Vemuri et al. [19] determined the optimum value of the control parameter N for the expected cost function of an $M^X/M/1$ system. Bouchentouf et al. [5] analyzed balking and server state-dependent reneging queues using generating functions and obtained the steady-state solution. Additionally, Bouchentouf et al. [2] analyzed a multi-server model with finite capacity, multiple synchronous working vacations, and balking. As a highly effective approximation function, ANFIS (Adaptive neuro-fuzzy inference system) is considered the best tool of neural and fuzzy systems which ensures smoothness to reduce the optimization search space from fuzzy systems. ANFIS has been Commonly implemented for prediction, control, and optimization Assignments. Divya and Indhira [20] applied ANFIS computing to analyze an unreliable model under hybrid vacation and feedback. Moreover, Indumathi et al. [21] applied the ANFIS to assess the accuracy of cost results for an $M/M/2$ system with two heterogeneous servers and Catastrophe and Restoration phenomena. Recently, Dehimi et al. [22] studied a finite multi-server model operating under a hybrid hiatus policy and applied ANFIS to validate the accuracy of diverse performance metrics achieved.

Scientists believe that the Grey Wolf Optimizer (GWO) has an exceptional hunting mechanism. From this perspective, optimization using the grey wolf method has gained prominence. GWO is employed as a meta-heuristic algorithm known for its strong optimal search capability in studying queueing system costs. The GWO algorithm was initially introduced by [23], who demonstrated its ability to address issues of instability and convergence accuracy, highlighting its superior accuracy and faster convergence speed. As a meta-heuristic algorithm for queueing systems, GWO was introduced in the seminal work of [24], where the authors applied GWO to study repairable systems in cloud computing using an N-policy. Also Dehimi et al. [25] applied GWO to derive the optimal service rates for a finite

$M/M/c$ system with synchronous differentiated working vacation policy, Bernoulli feedback, and impatience.

The originality of the work done is to study a finite-capacity single-server Markovian queue with feedback and impatience operating under the N-policy discipline along with the provision to go on a working vacation (single working vacation policy), simultaneously, which was not studied in past literature. Our study utilized GWO (Grey Wolf Optimizer) to minimize the cost function for this queueing system. To the best of our knowledge, GWO has not been used so far in optimizing this category's queuing systems. We use it to determine service rates couple (μ^*, μ_v^*) to minimize the expected cost function. Our model accurately represents an airport security checkpoint, where passengers arrive randomly following a Poisson process. The main contributions of this work can be summarized as follows:

- Obtaining the steady-state solution for the system using the Markov recursive method, which provides a powerful approach for analyzing steady-state probabilities and stochastic processes.
- Deriving important performance metrics, then performing a numerical analysis to validate the analytical results and investigate the impact of different system parameters on the performance metrics.
- Applying the GWO algorithm to acquire the optimum values of μ and μ_v of the optimum cost function. This offers decision-makers significant management information for designing management policy.

The paper is structured as follows: Section 2 introduces the mathematical description of the proposed model along with a practical application. Section 3 establishes the analysis of the system. Section 4 examines various performance measures based on the steady-state probability distribution of the system. Section 5 provides numerical simulations for the performance metrics. Section 6 applies the GWO algorithm to determine the optimal parameters for the cost function. Finally, a comprehensive conclusion for the study is presented.

2. Description of the queueing model

We consider a finite-capacity $M/M/1/K$ queue with feedback, single working vacation, impatient customers, and N-policy. The assumptions of the proposed model are as follows:

1. Customers arrive according to a Poisson process with rate λ. Upon arrival, a customer decides to join the queue with probability β_i or balk (refuse to join) with probability $1-\beta_i$, where $0 \leqslant i \leqslant K$. Specifically, $\beta_0 = 1$ and $\beta_K = 0$.
2. Customers, unsatisfied with the service provided, either leave the system with probability θ or return with probability $\theta' = 1-\theta$. Feedback customers are treated as new arrivals.
3. Service times follow an exponential distribution with rate μ during regular busy periods and μ_v during vacation periods ($\mu_v < \mu$). Service is provided on a First-In-First-Out (FIFO) discipline.
4. Upon entering the queue, a customer activates a timer T_0 (during dormant periods) or T_2 (during working vacation periods). These timers follow exponential distributions with rates ξ_0 and ξ_2 respectively. A customer leaves the queue with probability α and may return to the system with probability $1-\alpha$. Impatient clients are only active during vacation and dormant states.
5. When the number of customers in the system drops to zero while at least one server is active, the server enter a working vacation period (WV):
   - During a WV, the server serves arriving customers at a rate lower than the regular service rate. At the end of the vacation period, if the system size is N, the server switches to a regular busy period and starts operating under the N-policy.

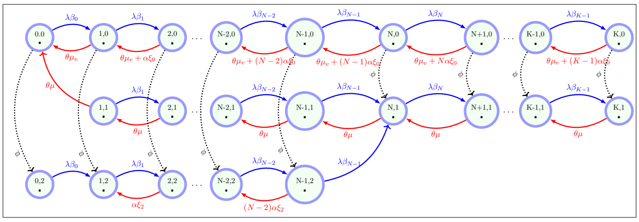


Figure 1. State transition rate diagram

- Under a Single Working Vacation (SWV) policy, server takes only one WV whenever the system becomes empty. If there are at least N customers at the end of the vacation period, the server resumes operation under the N-policy with the usual service rate. Otherwise, servers remain dormant in the system until N customers arrive instead of taking another WV.
- Vacation durations are assumed to follow an exponential distribution with rate ϕ.

We assume that inter-arrival times, service times, and vacation times are mutually independent. Figure 1 illustrates the state-transition rate diagram of the model.

2.1. Implementation of the model in practical scenarios

Consider an airport security checkpoint where passengers arrive randomly following a Poisson process with rate λ. Upon arrival, passengers decide whether to join the security queue based on its current length. Each passenger decides to join the queue based on a probability β_i, potentially opting to delay or avoid joining if the queue reaches its capacity K (balking). Passengers dissatisfied with wait times may leave (quit) with probability θ or return later (feedback) with probability $\theta' = 1 - \theta$. A single security agent performs screenings at rate μ during peak hours and a reduced rate μ_v during off-peak times. Impatient passengers may permanently leave the queue during quieter periods with probability α. If the queue becomes empty with one active agent, they enter a working vacation period, conducting screenings at a reduced rate for an exponentially distributed duration with rate ϕ. The agent resumes full-speed screening if enough passengers accumulate (N or more) by the end of the working vacation, ensuring effective passenger flow management under varying demand and operational conditions.

3. Queueing model analysis

Consider the state of the system at time t characterized by the random variables $N(t)$ representing the system size and $J(t)$ indicating the server's state, defined as:

$$j(t) = \begin{cases} 0, & \text{if the server is in WV period,} \\ 1, & \text{if the server is active during regular busy period,} \\ 2, & \text{if the server is dormant.} \end{cases}$$

We can define the process $\{N(t), j(t)\}$ with its state space as follows:

$$\Omega = \{(0,0) \cup (n,j) \,:\, 0 \leqslant n \leqslant K, j = 0,1,2\}.$$

Applying Markov process, we derive the following set of equations describing the steady state:

$$(\phi + \lambda\beta_0)P_{0,0} = \theta\mu_v P_{0,1} + \theta\mu P_{1,1}, \quad n = 0, \tag{1}$$

$$(\phi + \theta\mu_v + (n-1)a\xi_0 + \lambda\beta_n)P_{0,n} = (\theta\mu_v + na\xi_0)P_{0,n+1} + \lambda\beta_{n-1}P_{0,n-1}, \quad 1 \leqslant n \leqslant K-1,$$

$$(\phi + \theta\mu_v + (K-1)a\xi_0)P_{0,K} = \lambda\beta_{K-1}P_{0,K-1}, \quad n = K, \tag{2}$$

$$(\theta\mu + \lambda\beta_1)P_{1,1} = \theta\mu P_{1,2}, \quad n = 1,$$

$$(\lambda\beta_n + \theta\mu)P_{1,n} = \lambda\beta_{n-1}P_{1,n-1} + \theta\mu P_{1,n+1}, \quad 2 \leqslant n \leqslant N-1, \tag{3}$$

$$(\lambda\beta_N + \theta\mu)P_{1,N} = \lambda\beta_{N-1}P_{1,N-1} + \lambda\beta_{N-1}P_{2,N-1} + \theta\mu P_{1,N+1} + \phi P_{0,N}, \quad n = N, \tag{4}$$

$$(\lambda\beta_n + \theta\mu)P_{1,n} = \lambda\beta_{n-1}P_{0,n-1} + \theta\mu P_{1,n+1} + \phi P_{0,n}, \quad N+1 \leqslant n \leqslant K-1,$$

$$\theta\mu P_{1,K} = \lambda\beta_{K-1}P_{1,K-1} + \phi P_{0,K}, \quad n = K, \tag{5}$$

$$\lambda\beta_0 P_{2,0} = \phi P_{0,0}, \quad n = 0, \tag{6}$$

$$(\lambda\beta_n + (n-1)a\xi_2)P_{2,n} = \lambda\beta_{n-1}P_{2,n-1} + na\xi_2 P_{2,n+1} + \phi P_{0,n}, \quad 1 \leqslant n \leqslant N-2,$$

$$((N-2)a\xi_2 + \lambda\beta_{N-1})P_{2,N-1} = \lambda\beta_{N-2}P_{2,N-2} + \phi P_{0,N-1}, \; n = N-1. \tag{7}$$

The normalization condition:

$$\sum_{n=0}^{K} P_{0,n} + \sum_{n=1}^{K} P_{1,n} + \sum_{n=0}^{N-1} P_{2,n} = 1. \tag{8}$$

3.1. Steady-state solution

In this sub-section, we employ the Markov recursive method to obtain the steady-state distribution of the server states.

Theorem 1. *The steady-state probabilities of the system size during the working vacation period are provided as follows:*

$$P_{0,n} = \chi_n P_{0,K}, \tag{9}$$

where

$$\chi_n = \begin{cases} 1, & n = K, \\ \dfrac{\phi + \theta\mu_v + (K-1)a\xi_0}{\lambda\beta_{K-1}}, & n = K-1, \\ \left(\dfrac{\lambda\beta_{n+1} + \phi + \theta\mu_v + na\xi_0}{\lambda\beta_n}\right)\chi_{n+1} - \left(\dfrac{\theta\mu_v + (n+1)a\xi_0}{\lambda\beta_n}\right)\chi_{n+2}, & \\ 0 \leqslant n \leqslant K-2. & \end{cases}$$

The steady-state probabilities during the busy period are given by:

$$P_{1,n} = \Upsilon_n P_{0,K} + t_{n-2}P_{2,N-1},$$

where

$$\Upsilon_n = \begin{cases} \dfrac{\phi+\lambda\beta_0}{\theta\mu}\chi_0 - \dfrac{\theta\mu_v}{\theta\mu}\chi_1, & n=1, \\ \dfrac{\lambda\beta_1+\theta\mu}{\theta\mu}\Upsilon_1, & n=2, \\ \left(\dfrac{\lambda\beta_{n-1}+\theta\mu}{\theta\mu}\right)\Upsilon_{n-1} - \left(\dfrac{\lambda\beta_{n-2}}{\theta\mu}\right)\Upsilon_{n-2}, & 3 \leqslant n \leqslant N, \\ \left(\dfrac{\lambda\beta_{n-1}+\theta\mu}{\theta\mu}\right)\Upsilon_{n-1} - \left(\dfrac{\lambda\beta_{n-2}}{\theta\mu}\right)\Upsilon_{n-2} - \dfrac{\phi}{\theta\mu}\chi_{n-2}, & N+1 < n \leqslant K-1. \end{cases}$$

and

$$t_n = \begin{cases} 0, & 0 \leqslant n \leqslant N-2, \\ \dfrac{\lambda\beta_{N-1}}{\theta\mu}, & n = N-1, \\ \left(\dfrac{\lambda\beta_{N+1}+\theta\mu}{\theta\mu}\right)t_{N-1}, & n=N, \\ \left(\dfrac{\lambda\beta_{n+1}+\theta\mu}{\theta\mu}\right)t_{n-1} - \left(\dfrac{\lambda\beta_n}{\theta\mu}\right)t_{n-2}, & N+1 < n \leqslant K-2. \end{cases}$$

and the stationary probabilities for the dormant period verify the equation:

$$P_{2,n} = \Delta_n P_{2,N-1} + \Theta_n P_{0,K},$$

where

$$\Delta_n = \begin{cases} 0, & N \leqslant n \leqslant K, \\ 1 & n = N-1, \\ \left(\dfrac{(N-2)a\xi_2+\lambda\beta_{N-1}}{\lambda\beta_{N-1}}\right), & n=N-2, \\ \left(\dfrac{\lambda\beta_{n+1}+na\xi_2}{\lambda\beta_n}\right)\Delta_{n+1} - \left(\dfrac{(n+1)a\xi_2}{\lambda\beta_n}\right), & n=N-3, \\ \left(\dfrac{\lambda\beta_{n+1}+na\xi_2}{\lambda\beta_n}\right)\Delta_{n+1} - \left(\dfrac{(n+1)a\xi_2}{\lambda\beta_n}\right)\Delta_{n+2}, & 0 \leqslant n < N-3. \end{cases}$$

$$\Theta_n = \begin{cases} 0, & N-1 \leqslant n \leqslant K, \\ \left(\dfrac{\phi\psi_{N-1}}{\lambda\beta_{N-2}}\right), & n=N-2, \\ \left(\dfrac{\lambda\beta_{n+1}+na\xi_2}{\lambda\beta_n}\right)\psi_{n+2} - \left(\dfrac{\phi}{\lambda\beta_n}\right)\psi_{n+1}, & n=N-3, \\ \left(\dfrac{\lambda\beta_{n+1}+na\xi_2}{\lambda\beta_n}\right)\Theta_{n+1} - \left(\dfrac{(n+1)a\xi_2}{\lambda\beta_n}\right)\Theta_{n+2} - \dfrac{\phi}{\lambda\beta_n}\psi_{n+1}, & 0 \leqslant n < N-3. \end{cases}$$

$$\Theta_n = \Theta_{n+1} - \frac{\phi}{\lambda\beta_n}\chi_{n+1}, \quad 0 \leqslant n \leqslant N-2.$$

Then, using equation (7)*, we get:*

$$P_{2,N-1} = \left(\frac{\phi\chi_0 - \lambda\beta_0\Theta_0}{\lambda\beta_0\Delta_0}\right)P_{0,K}.$$

Finally, we have:

$$P_{0,K} = \left[\sum_{n=0}^{K} \chi_n + \sum_{n=1}^{K} (\Upsilon_n + \varsigma_0 t_n) + \sum_{n=0}^{N-1} (\varsigma_0 \Delta_n + \Theta_n) \right]^{-1}.$$

Where, $\varsigma_0 = \left(\dfrac{\phi \chi_0 - \lambda \beta_0 \Theta_0}{\lambda \beta_0 \Delta_0} \right)$.

4. Performance metrics

In this section, useful performance measures for different server states are presented.

1. The probability of the server being in a working vacation state:

$$P_{Vacation} = \sum_{n=0}^{K} P_{0,n} = \sum_{n=0}^{K} \chi_n P_{0,K}.$$

2. The probability of the server being in a busy state:

$$P_{Busy} = \left[\sum_{n=1}^{K} \Upsilon_n + \varsigma_0 \sum_{n=1}^{K-2} t_n \right] P_{0,K}.$$

3. The probability of the server being in a dormant state:

$$P_{Dormant} = \left[\varsigma_0 \sum_{n=0}^{N-1} \Delta_n + \sum_{n=0}^{N-2} \Theta_n \right] P_{0,K}.$$

4. The probability that the server is idle:

$$P_{idle} = \left[\chi_0 + \varsigma_0 \sum_{n=0}^{N-1} \Delta_n + \sum_{n=0}^{N-2} \Theta_n \right] P_{0,K}.$$

5. The expected number of customers in the queue:

$$L_s = \sum_{n=0}^{K} n P_{0,n} + \sum_{n=1}^{K} (n-1) P_{1,n} + \sum_{n=0}^{N-1} (n-1) P_{2,n} = \left[\sum_{n=0}^{K} \chi_n + \sum_{n=1}^{K} \Upsilon_n + \varsigma_0 \left(\sum_{n=0}^{K-2} t_n + \sum_{n=0}^{N-1} \Delta_n \right) + \sum_{n=0}^{N-2} \Theta_n \right] P_{0,K}.$$

6. The average balking rate:

$$BR = \sum_{n=0}^{K} n\lambda(1-\beta_n) P_{0,n} + \sum_{n=1}^{K} (n-1)\lambda(1-\beta_n) P_{1,n} + \sum_{n=0}^{N-1} n\lambda(1-\beta_n) P_{2,n} = \sum_{n=0}^{K} n\lambda(1-\beta_n)\chi_n P_{0,K} + $$
$$+ \sum_{n=1}^{K} (n-1)\lambda(1-\beta_n) \left(\sum_{n=1}^{K} \Upsilon_n + \varsigma_0 \sum_{n=1}^{K-2} t_n \right) P_{0,K} + n\lambda(1-\beta_n) \left(\varsigma_0 \sum_{n=0}^{N-1} \Delta_n + \Theta_n \right) P_{0,K}.$$

7. The average reneging rate:

$$RR = \xi_0 \sum_{n=0}^{K} (n-1) P_{0,n} + \xi_2 \sum_{n=0}^{N-1} (n-1) P_{2,n} = \left[\xi_0 \sum_{n=0}^{K} (n-1)\chi_n + \xi_2 \left(\varsigma_0 \sum_{n=0}^{N-1} (n-1)\Delta_n + \Theta_n \right) \right] P_{0,K}.$$

8. The average rate of lost customers:

$$AR = BR + RR.$$

9. The expression for the expected waiting time of customers in the system:

$$W_s = \frac{L_s}{\lambda'}, \text{ where } \lambda' = \lambda - AR.$$

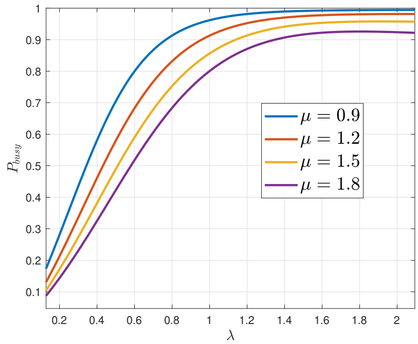


Figure 2. P_{busy} vs. λ for different values of μ

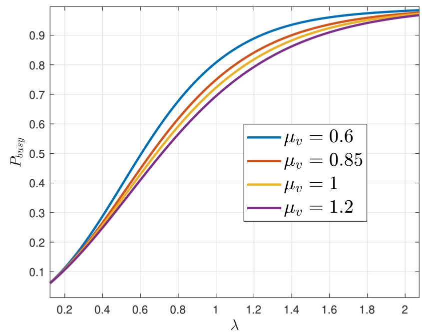


Figure 3. P_{busy} vs. λ for different values of μ_v

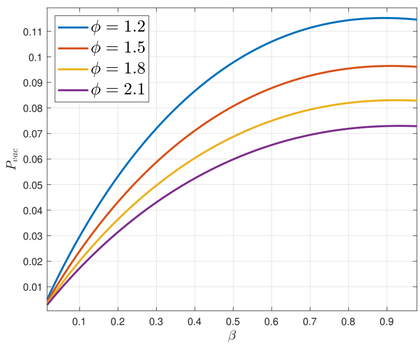


Figure 4. P_{vac} vs. β for multiple values of ϕ

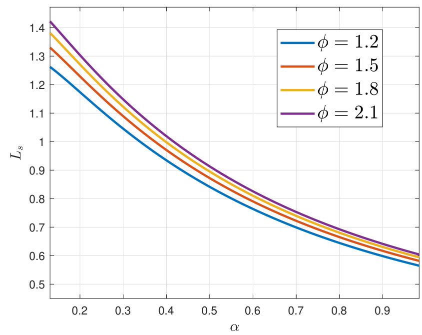


Figure 5. L_s vs. α for multiple values of ϕ

5. Numerical results

In this section, key numerical results are presented graphically to illustrate the impact of various system parameters on different performance metrics. These graphs were generated using the **MATLAB** program. For this purpose, the following set of system parameters was fixed: $\lambda = 0.4$, $\mu = 1.5$, $\mu_v = 0.7$, $\phi = 0.6$, $\xi_0 = 0.5$, $\xi_2 = 1.8$, $\theta = 0.7$, $\beta = 0.8$, $\alpha = 0.7$, $N = 3$, and $K = 5$.

5.1. Analysis of findings

The numerical experiments systematically analyzed the sensitivity of various system parameters on performance measures. Based on these analyses, the following key observations and potential managerial recommendations have been identified:

1. Effect of λ (arrival rate): With an increasing value of λ, several factors are significantly affected. The service rates μ and μ_v increase correspondingly, leading to an increase in the probability of the server being in the busy state P_{busy} (see Figures 2 and 3). Additionally, as the capacity K increases, the expected number of customers in the queue L_s also increases (see Figure 7).

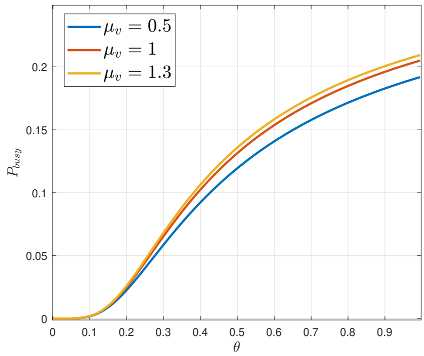


Figure 6. P_{busy} vs. θ for different values of μ_v

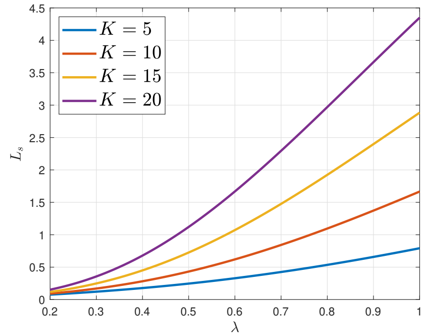


Figure 7. L_s vs. λ for different values of K

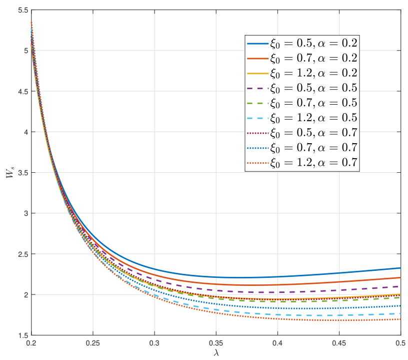


Figure 8. W_s vs. λ for different values of ξ_0 and α

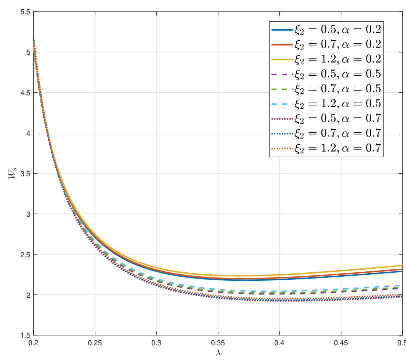


Figure 9. W_s vs. λ for different values of ξ_2 and α

2. Effect of parameter ϕ (vacation rate): As the balking rate β and vacation rate ϕ increase, P_{vac} gradually increases (see Figure 4). Conversely, an increase in the probability α while increasing the vacation rate ϕ leads to a decrease in the expected number of customers in the queue L_s (see Figure 5).
3. Effect of parameter θ (feedback rate): As the feedback rate θ increases, the probability P_{busy} of the server being in the busy state increases for different service rates μ_v (see Figure 6).
4. Effect of parameters ξ_0 and ξ_2 (impatience rate): A lower impatience rate ξ_0 (or ξ_2) and parameter α (probability of leaving the queue) lead to an increase in the expected waiting time of customers in the system W_s (see Figures 8 and 9). Consequently, due to this impatience, there is an increase in the number of customers L_s.

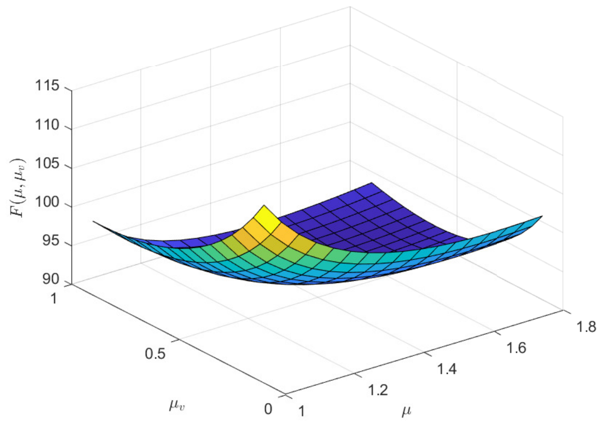


Figure 10. $F(\mu, \mu_v)$ vs. μ and μ_v

6. Numerical cost optimum

6.1. Cost model

The utilization of various cost factors in terms of cost function can play a crucial role in optimizing and enhancing the system availability by considering the benefits associated with different design variable options. Using the Grey Wolf Optimizer (GWO) algorithm . We will assess the expected total cost and determine the optimal values of (μ^*, μ_v^*) the decision variables based on the costs are fixed as: $C_{WV} = 30$, $C_B = 20$, $C_d = 15$, $C_s = 20$, $C_{AR} = 30$, $C_{\mu_v} = 10$, $C_\mu = 20$.

Then we define the total expected cost per unit of time of the system as follows:

$$F = C_{wv}P_{vac} + C_B P_{busy} + C_d(P_{dormant} + P_{idle}) + C_s L_s + C_{AR}AR + \mu C_{mu} + \mu_v Cv,$$

where,

- C_{wv}: denotes the cost per unit time when the server is on a working vacation period,
- C_B: denotes the cost per unit time when the server is on a busy period,
- C_d: denotes the cost per unit time when the server is on a dormant or idle period,
- C_s: denotes the cost per unit time when a customer joins the queue and waits for service,
- C_{AR}: denotes the cost per unit time when a customer leaves the queue,
- C_{μ_v} (resp. C_μ): denotes the cost per service per unit time during normal busy period (resp. working vacation period).

6.2. GWO—Grey Wolf optimizer

The Grey Wolf optimizer (GWO) algorithm is inspired by the leadership organization and hunting strategy of grey wolves in nature. GWO represents a recent innovation in cost optimization methods. This meta-heuristic algorithm effectively explores the search space and converges to the optimal solution by simulating the hunting behavior of grey wolves. This section presents a practical

Table 1

The optimal (μ^*, μ_v^*) and $F^*(\mu^*, \mu_v^*)$ vs. θ, when $\theta = 0.2 : 0.8$, $\xi_0 = 0.5$, $\xi_2 = 1.8$, $\lambda = 0.4$, $\phi = 0.6$, $\beta = 0.8$, $\alpha = 0.7$, $N = 3$ and $K = 5$

	μ^*	μ_v^*	$F(\mu^*, \mu_v^*)$
$\theta = 0.2$	3.8882	1.0039	143.9779
$\theta = 0.4$	2.4800	0.8102	104.9952
$\theta = 0.6$	2.3922	0.5771	80.0489
$\theta = 0.8$	2.3003	0.4035	83.5105

Table 2

The optimal (μ^*, μ_v^*) and $F^*(\mu^*, \mu_v^*)$ vs. K and N, when $\lambda = 0.4$, $\phi = 0.6$, $\xi_0 = 0.5$, $\xi_2 = 1.8$, $\theta = 0.7$, $\beta = 0.8$, $\alpha = 0.7$

		μ^*	μ_v^*	$F(\mu^*, \mu_v^*)$
$K = 10$	$N = 3$	2.2297	0.8871	126.2557
	$N = 5$	1.8132	0.9201	80.9963
	$N = 7$	1.6801	0.0225	69.3733
$K = 15$	$N = 3$	2.5312	1.3530	151.4877
	$N = 5$	1.8374	0.3141	90.3144
	$N = 7$	1.6846	0.0179	70.9343
$K = 18$	$N = 3$	2.7078	1.5911	164.5993
	$N = 5$	1.8545	0.4193	90.3144
	$N = 7$	1.6944	0.0105	71.4826
$K = 20$	$N = 3$	2.8737	1.7380	172.7564
	$N = 5$	1.8656	0.4832	94.6289
	$N = 7$	1.7018	0.0956	71.9032

application of GWO in the field of queueing systems. The objective of this study is to determine the optimal service rates (μ, μ_v) that minimize the expected cost function. Given the complexity of higher-order non-linear optimization problems, it is recommended to employ nonlinear optimization techniques to find these optimal solutions. For optimizing the cost model, the GWO algorithm is applied by initially setting parameters to obtain the optimal values of (μ^*, μ_v^*).

The optimization problem can be written as:

$$\min_{\mu, \mu_v} F(\mu, \mu_v) \text{s.t} \begin{cases} \mu - \mu_v > 0 \\ \mu_v > 0 \\ (\mu, \mu_v) \in \mathbb{R}_+^2. \end{cases}$$

Table 3

The optimal (μ^*, μ_v^*) and $F^*(\mu^*, \mu_v^*)$ vs. λ, when $\lambda = 0.6 : 1.8$, $\phi = 0.6$, $\xi_0 = 0.5$, $\xi_2 = 1.8$, $\theta = 0.7$, $\beta = 0.8$, $\alpha = 0.7$, $N = 3$ and $K = 5$

	μ^*	μ_v^*	$F(\mu^*, \mu_v^*)$
$\lambda = 0.6$	2.9196	0.9762	101.4854
$\lambda = 0.8$	2.7014	1.3150	117.0471
$\lambda = 1$	3.1103	1.6098	131.1372
$\lambda = 1.2$	3.4837	1.8806	144.1337
$\lambda = 1.4$	3.8283	2.1393	156.2874
$\lambda = 1.8$	4.1440	2.2980	167.7924

Table 4

The optimal (μ^*, μ_v^*) and $F^*(\mu^*, \mu_v^*)$ vs. ϕ, when $\phi = 0.5 : 5.5$, $\xi_0 = 0.5$, $\xi_2 = 1.8$, $\lambda = 0.4$, $\theta = 0.7$, $\beta = 0.8$, $\alpha = 0.7$, $N = 3$ and $K = 5$

	μ^*	μ_v^*	$F(\mu^*, \mu_v^*)$
$\phi = 0.5$	1.7138	0.5838	83.6245
$\phi = 1.5$	1.7043	0.7143	83.5471
$\phi = 2.5$	1.7059	0.7596	83.4869
$\phi = 3.5$	1.7070	0.7801	83.4537
$\phi = 4.5$	1.7082	0.7939	83.4330
$\phi = 5.5$	1.7090	0.8034	83.4188

- Table 1 clearly shows that higher feedback probabilities have a negative effect on the optimal cost and (μ^*, μ_v^*).
- By examining Table 2, it is evident that the optimal cost and (μ^*, μ_v^*) increase with the system capacity K and the number of customers N.
- The impact of arrival rates can be observed through Table 3. The optimal cost and (μ^*, μ_v^*) directly increase as the arrival rate takes larger values.
- Table 4 demonstrates that the vacation rate has a direct impact on the optimal cost and (μ^*, μ_v^*); as the vacation rate increases, so do the optimal cost and (μ^*, μ_v^*).
- Upon analyzing Table 5, it becomes clear that the optimal cost and (μ^*, μ_v^*) exhibit a significant increase as the values of the impatience rates ξ_0 and ξ_2 increase.
- Figure 11, shows that a higher arrival rate λ leads to an increase in the expected cost $F(K, N)$ and (K, N) values.

Table 5

The optimal (μ^*, μ_v^*) and $F^*(\mu^*, \mu_v^*)$ vs. ξ_0 and ξ_0, when $\xi_0 = [0.5; 1; 1.5]$, $\xi_2 = [1; 1.2; 1.4; 1.6; 1.8]$, $\lambda = 0.4$, $\phi = 0.6$, $\theta = 0.7$, $\beta = 0.8$, $\alpha = 0.7$, $N = 3$ and $K = 5$

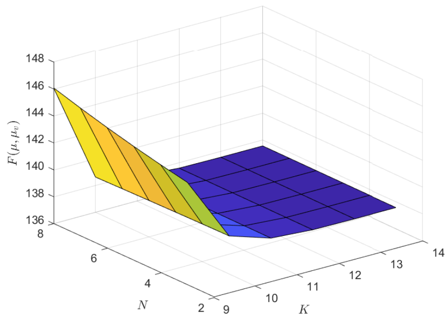
		μ^*	μ_v^*	$F(\mu^*, \mu_v^*)$
$\xi_0 = 0.5$	$\xi_2 = 1$	1.5167	0.4454	77.5487
	$\xi_2 = 1.2$	1.5718	0.4910	79.2290
	$\xi_2 = 1.4$	1.6209	0.5348	80.7922
	$\xi_2 = 1.6$	1.6673	0.5747	82.2544
	$\xi_2 = 1.8$	1.7117	0.6102	83.6292
$\xi_0 = 1$	$\xi_2 = 1$	1.4912	0.2287	77.4745
	$\xi_2 = 1.2$	1.5448	0.2778	79.1299
	$\xi_2 = 1.4$	1.5955	0.3210	80.6766
	$\xi_2 = 1.6$	1.6408	0.3601	82.1278
	$\xi_2 = 1.8$	1.6865	0.3968	83.4951
$\xi_0 = 1.5$	$\xi_2 = 1$	1.4666	0.0140	77.3672
	$\xi_2 = 1.2$	1.5203	0.0641	79.0078
	$\xi_2 = 1.4$	1.5713	0.1081	80.5449
	$\xi_2 = 1.6$	1.6182	0.1483	81.9898
	$\xi_2 = 1.8$	1.6608	0.1854	83.3530

7. Conclusion

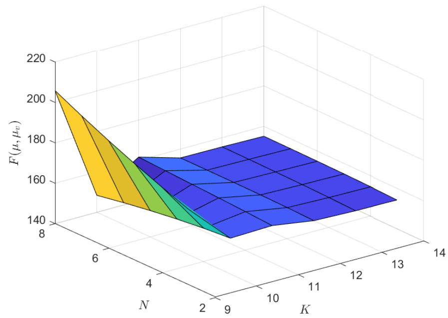
This study is based on a finite capacity queueing model tailored for application in the scenario of passengers at the airport. Our model represents single working vacations, N-policy, feedback during both dormant and working vacation periods, and impatience timers which depend on the server's states. By employing the Markov recursive method, we derived closed-form expressions for steady-state probabilities and performance metrics. Furthermore, various metrics were graphically shown and discussed. Additionally, we applied the GWO algorithm intending to obtain the optimum service rates for the expected cost function and also conducted a comprehensive numerical analysis to assess the influence of various parameters on the results obtained.

In the future, the model could be enhanced by incorporating characteristic customer behaviors such as priority mechanisms, as well as multiple optional services. This extension would broaden the model applicability but also increase the complexity of calculations.

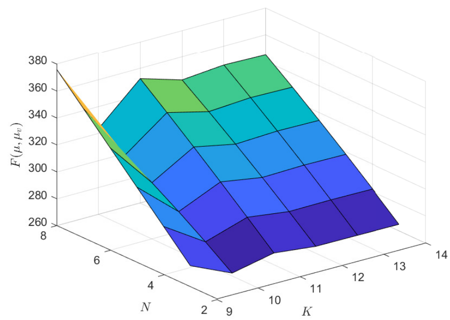
Author Contributions: Conceptualization, A. Kadi, M. Boualem and N. Touche; methodology, A. Kadi, M. Boualem and N. Touche; software, A. Kadi, N. Touche and A. Dehimi; formal analysis, A. Kadi, M. Boualem and A. Dehimi; writing-original draft preparation, A. Kadi, M. Boualem and A. Dehimi; writing-review and editing, A. Kadi, M. Boualem, N. Touche and A. Dehimi; visualization, A. Kadi, M. Boualem, N. Touche and A. Dehimi; supervision, M. Boualem and N. Touche. All authors have read and agreed to the published version of the manuscript.



(a) $\lambda = 0.4$



(b) $\lambda = 1.4$



(c) $\lambda = 2.4$

Figure 11. The expected cost $F^*(K, N)$ vs. λ, when $\phi = 0.6$, $\xi_0 = 0.5$, $\xi_2 = 1.8$, $\theta = 0.7$, $\beta = 0.8$, $\alpha = 0.7$

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: The authors would like to thank the editor for their support and guidance throughout the review process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bouchentouf, A. A., Cherfaoui, M. & Boualem, M. Analysis and Performance Evaluation of Markovian Feedback Multi-Server Queueing Model with Vacation and Impatience. *American Journal of Mathematical and Management Sciences* **40,** 261–282. doi:10.1080/01966324.2020.1842271 (Nov. 2020).
2. Bouchentouf, A. A., Boualem, M., Yahiaoui, L. & Ahmad, H. A multi-station unreliable machine model with working vacation policy and customers' impatience. *Quality Technology & Quantitative Management* **19,** 766–796. doi:10.1080/16843703.2022.2054088 (Apr. 2022).

3. Gorbunova, A., Zaryadov, I., Samouylov, K. & Sopin, E. A Survey on Queuing Systems with Parallel Serving of Customers. *Discrete and Continuous Models and Applied Computational Science* **25,** 350–362. doi:10.22363/2312-9735-2017-25-4-350-362 (Jan. 2017).
4. Bose, S. K. *An introduction to queueing systems* (Springer Science & Business Media, Dec. 2013).
5. Bouchentouf, A. A., Boualem, M., Cherfaoui, M. & Medjahri, L. Variant vacation queueing system with Bernoulli feedback, balking and server's states-dependent reneging. *Yugoslav Journal of Operations Research* **31,** 557–575. doi:10.2298/yjor200418003b (Jan. 2021).
6. Bouchentouf, A. A., Cherfaoui, M. & Boualem, M. Performance and economic analysis of a single server feedback queueing model with vacation and impatient customers. *OPSEARCH* **56,** 300–323. doi:10.1007/s12597-019-00357-4 (Feb. 2019).
7. Chettouf, A., Bouchentouf, A. A. & Boualem, M. *A Markovian Queueing Model for Telecommunications Support Center with Breakdowns and Vacation Periods* in *Operations Research Forum* **5** (2024), 22. doi:10.1007/s43069-024-00295-y.
8. Yadin, M. & Naor, P. Queueing Systems with a Removable Service Station. *Journal of the Operational Research Society* **14,** 393–405. doi:10.1057/jors.1963.63 (Dec. 1963).
9. Meena, R. K., Jain, M., Assad, A., Sethi, R. & Garg, D. Performance and cost comparative analysis for $M/G/1$ repairable machining system with N-policy vacation. *Mathematics and Computers in Simulation* **200,** 315–328. doi:10.1016/j.matcom.2022.04.012 (Oct. 2022).
10. Bouchentouf, A. A., Medjahri, L., Boualem, M. & Kumar, A. Mathematical analysis of a Markovian multi-server feedback queue with a variant of multiple vacations, balking and reneging. *Discrete and Continuous Models and Applied Computational Science* **30,** 21–38. doi:10.22363/2658-4670-2022-30-1-21-38 (Apr. 2022).
11. Kathirvel, J. A single server perishable inventory system with N additional options for service. *Journal of Mathematical Modeling* **2,** 187–216 (May 2015).
12. Kadi, A., Touche, N., Bouchentouf, A. A. & Boualem, M. Finite-capacity $M/M/2$ machine repair model with impatient customers, triadic discipline, and two working vacation policies. *Journal of Mathematical Modeling* **13,** 183–200. doi:10.22124/jmm.2024.27095.2384 (2024 Article in press).
13. Adou, K. Y. B., Markova, E. V. & Zhbankova, E. A. Performance analysis of queueing system model under priority scheduling algorithms within 5G networks slicing framework. *Discrete and Continuous Models and Applied Computational Science* **30,** 5–20. doi:10.22363/2658-4670-2022-30-1-5-20 (Apr. 2022).
14. Sharma, R. Optimal N-policy for unreliable server queue with impatient customer and vacation interruption. *Journal of Reliability and Statistical Studies* **10,** 83–96 (May 2017).
15. Rajadurai, P., Saravanarajan, M. & Chandrasekaran, V. A study on $M/G/1$ feedback retrial queue with subject to server breakdown and repair under multiple working vacation policy. *Alexandria Engineering Journal* **57,** 947–962. doi:10.1016/j.aej.2017.01.002 (June 2018).
16. Boualem, M. Stochastic analysis of a single server unreliable queue with balking and general retrial time. *Discrete and Continuous Models and Applied Computational Science* **28,** 319–326. doi:10.22363/2658-4670-2020-28-4-319-326 (Dec. 2020).
17. Hilquias, V. C. C., Zaryadov, I. S. & Milovanova, T. A. Queueing systems with different types of renovation mechanism and thresholds as the mathematical models of active queue management mechanism. *Discrete and Continuous Models and Applied Computational Science* **28,** 305–318. doi:10.22363/2658-4670-2020-28-4-305-318 (Dec. 2020).
18. Goswami, V. Relationship between randomized F-policy and randomized N-policy in discrete-time queues. *OPSEARCH* **53,** 131–150. doi:10.1007/s12597-015-0220-y (Sept. 2015).
19. Vemuri, V. K., Siva, V., Kotagiri, C. & Bethapudi, R. T. Optimal strategy analysis of an N-policy two-phase $M^X/M/1$ queueing system with server startup and breakdowns. *Opsearch* **48,** 109–122. doi:10.1007/s12597-011-0046-1 (June 2011).

20. Divya, K. & Indhira, K. Performance Analysis And ANFIS Computing Of An Unreliable Markovian Feedback Queueing Model Under A Hybrid Vacation Policy. *Mathematics and Computers in Simulation* **218,** 403–419. doi:10.1016/j.matcom.2023.12.004 (Apr. 2024).
21. Indumathi, P. & Karthikeyan, K. ANFIS-Enhanced $M/M/2$ Queueing Model Investigation in Heterogeneous Server Systems with Catastrophe and Restoration. *Contemporary Mathematics,* 2482–2502. doi:10.37256/cm.5220243977 (June 2024).
22. Dehimi, A., Boualem, M., Kahla, S. & Berdjoudj, L. 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,** 159–170. doi:10.17535/crorr.2024.0013 (2024).
23. Mirjalili, S., Mirjalili, S. M. & Lewis, A. Grey Wolf Optimizer. *Advances in Engineering Software* **69,** 46–61. doi:10.1016/j.advengsoft.2013.12.007 (Mar. 2014).
24. Chahal, P. K., Kumar, K. & Soodan, B. S. Grey wolf algorithm for cost optimization of cloud computing repairable system with N-policy, discouragement and two-level Bernoulli feedback. *Mathematics and Computers in Simulation* **225,** 545–569. doi:10.1016/j.matcom.2024.06.005 (Nov. 2024).
25. Dehimi, A., Boualem, M., Bouchentouf, A. A., Ziani, S. & Berdjoudj, L. Analytical and Computational Aspects of a Multi-Server Queue With Impatience Under Differentiated Working Vacations Policy. *Reliability: Theory & Applications* **19,** 393–407. doi:10.24412/1932-2321-2024-379-393-407 (2024).

Information about the authors

Abir Kadi—PhD student, at Department of Mathematics, Laboratory of Applied Mathematics, 06000 Bejaia, Algeria (e-mail: abir.kadi@univ-bejaia.dz, ORCID: 0009-0000-6668-0270)

Mohamed Boualem—Full Professor of Applied Mathematics at the Department of Automation, Telecommunications, and Electronics at the University of Bejaia, Algeria. Permanent Researcher at the Research Unit LaMOS (Modeling and Optimization of Systems). (e-mail: mohammed.boualem@univ-bejaia.dz, phone: +213770556978, ORCID: 0000-0001-9414-714X)

Nassim Touche—Full Professor of Mathematics at Department of Operations research, Faculty of Exact Sciences, Research Unit LaMOS (Modeling and Optimization of Systems) (e-mail: nassim.touche@univ-bejaia.dz, ORCID: 0000-0002-9185-3433)

Aimen Dehimi—PhD student, at Department of Mathematics, Laboratory of Applied Mathematics, 06000 Bejaia, Algeria (e-mail: aimen.dehimi@univ-bejaia.dz, ORCID: 0009-0009-1221-9898)

УДК 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2025-33-1-10-26 EDN: AMSSFO

Моделирование и оптимизация очереди $M/M/1/K$ с одиночным рабочим отпуском, обратной связью и таймерами нетерпимости в рамках N-политики

Абир Кади[1], Мохамед Буалем[2], Нассим Туш[2], Аймен Дехими[1]

[1] Университет Беджаи, Лаборатория прикладной математики, 06000 Беджаи, Алжир

[2] Университет Беджаи, Исследовательская группа LaMOS (Моделирование и оптимизация систем), 06000 Беджаи, Алжир

Аннотация. Этот труд представляет собой интенсивное исследование модели очереди с одним сервером и конечной ёмкостью, с таймерами нетерпимости, зависящими от состояний сервера, с обратной связью и политикой одиночного рабочего отпуска, функционирующей в рамках дисциплины N-политики. Мы рассматриваем сценарий, при котором сервер должен дождаться, пока количество клиентов не достигнет N, чтобы начать обычный рабочий период; в противном случае сервер начнёт рабочий отпуск или перейдёт в неактивное состояние, если количество клиентов увеличится. С помощью метода Марковской рекурсии были получены вероятности в установившемся состоянии. Различные показатели производительности были визуально изображены для оценки различных конфигураций параметров системы. После построения ожидаемой функции стоимости модели используется алгоритм Оптимизация серых волков (GWO) для определения оптимальных значений коэффициентов обслуживания μ и μ_υ. Приведены числовые примеры для проверки теоретических выводов, что позволяет глубже понять эту сложную систему.

Ключевые слова: система очередей с нетерпимостью, N-политика, политика отпуска, обратная связь, алгоритм GWO, оптимизация стоимости