Vol. 34, No. 3 (2024) | DOI: 10.37190/ord240311

OPEN ACCESS

Operations Research and Decisions

www.ord.pwr.edu.pl

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

Cost optimization of a $M/M/1/WV\&MAV$ queueing system using Newton–Raphson and particle swarm optimization techniques

Ramachandran Remya[1] Amina Angelika Bouchentouf[2] Kaliappan Kalidass[1, 3*]

[1] *Department of Mathematics, Karpagam Academy of Higher Education, Coimbatore-641 021, Tamil Nadu, India*
[2] *Mathematics Laboratory, Djillali Liabes University of Sidi Bel Abbes,*
Laboratory of Stochastic Models, Statistic and Applications, University of Saida – Dr. Moulay Tahar, Algeria
[3] *Department of Mathematics, Chikkanna Government Arts College, Tirupur-641 602, Tamil Nadu, India*
[*] *Corresponding author, email address: dassmaths@gmail.com*

Abstract

This paper is concerned with the optimal control of a Markovian queueing system subjected to multiple adaptive vacation and working vacation policies. This system is applicable in diverse modern technologies, in particular in call centers. We establish the steady-state solution as well as important system characteristics by means of probability generating functions technique. We also construct the expected total cost for this model and develop a procedure to determine the optimal service rate that yields the minimum cost. Further, we carried out a comparative analysis to obtain the minimum cost using the Newton–Raphson method and particle swarm optimization (PSO) algorithm.

Keywords: *Markovian queue, working vacation, adaptive vacations, cost model, optimization*

1. Introduction

In recent years, vacation queues have been largely analyzed by diverse researchers due to their wonderful applications in the problem of congestion in signal transmission, telecommunication, production, network transportation, computer science and signal system, quality control, manufacturing and telecommunication system design and control. The readers can refer to [6, 27, 29] for excellent surveys on the earlier works on the subject.

As to vacation queues, majority of the works were focused on models with multiple vacations policy and single vacation policy [2–4, 12–15, 17, 22, 28, 30, 31, 37]. Nevertheless, multiple adaptive vacation is more general than most classical vacation ones in the way that both multiple and single vacation

Received 2 January 2023, accepted 2 July 2024, published online 17 October 2024
ISSN 2391-6060 (Online)/© 2024 Authors

The costs of publishing this issue have been co-financed by the program *Development of Academic Journals* of the Polish Ministry of Education and Science under agreement RCN/SP/0241/2021/1

policies represent the extreme cases of this policy. The concept was first introduced by Zhang and Tian [36] where they dealt with discrete time geometric/$G/1$ queueing system with multiple adaptive vacations. Then, a $Geo/G/1$ queueing system with batch arrival and multiple adaptive vacations has been treated by Sun et al. [26]. After that, Ma and Xu [18] discussed a $M/G/1$ queue with multiple adaptive vacations. An $M/G/1$ reparable queue with multiple adaptive vacations and p-entering discipline has been treated by Cheng and Tang [5]. Later, Ma et al. [19] considered a general limited service $Geom/G/1$ queue with multiple adaptive vacations. Jeyakumar and Rameshkumar [10] dealt with a $M^X/G(a,b)/1$ queueing model with server breakdown without interruption and controllable arrivals during multiple adaptive vacations. Recent results on different queueing models with adaptive vacation have presented by Jeyakumar and Rameshkumar [7–9].

The basis of the research, in classical vacation queueing models, is the assumption that during the idle period of the server, the system completely shuts down a service. Whereas, in order to manage the service systems efficiently and economically, it may not be from system economics perspective to maintain idle servers in the system. It is from this point that working vacation policy has been borne in mind. This idea was first initiated by Servi and Finn [24], where an $M/M/1$ queue with multiple working vacation policy has been dealt with. Since then, vast research papers has been given on working vacation queueing systems. For recent devolvements on the subject, the reader may refer to [4, 11, 16, 20, 23, 25, 32–35] and the references therein. This paper treats a Markovian time queueing system with multiple adaptive hiatuses and working hiatus policy. The considerable advantages of the current research works in this article are along these lines:

- The proposed queueing system with adaptive hiatus policy appears in varied practical areas where through this model, we can analyse diverse hiatus policies between these two extremes in order to better allocate server time to perform primary works (serving queue) and do secondary works (vacations).
- The model suggested is subjected to working hiatus and multiple adaptive hiatus policy, this is more realistic in diverse modern technologies, in particular in call centers, at which when the busy period is ended, if no calls are received, the agents may do some other tasks as sending emails (working hiatus). When this period is over, he comes back to the system to see if there are new requests. If not, he returns to the hiatus at which he can take a rest. At the hiatus completion instant, if there exist new arrivals, he switches to the busy period, otherwise, he is permitted to take random number of hiatuses. When the hiatuses are complete, the agents return to the normal working period and stay there waiting for new calls.
- By taking into consideration multiple adaptive hiatus policy and working hiatus simultaneously, the queueing model becomes complex. This complexity, makes difficult to model the system and therefore to derive its steady-state probabilities. Accordingly, the current analysis has not been considered in the literature as yet.
- A cost function is developed, to optimize, via the Newton–Raphson and PSO techniques, the service rate during busy period, at the optimal cost. This grants decision-makers substantial management info for designing management policy.

The paper is organised as follows. Section 2 is devoted to the description of the model. A practical motivation of the model discussed in section 3. In Section 4, the balance equations are given, while the system analysis is provided in Section 5. Then, some particular cases of the suggested model are presented in Section 6. Useful performance measures are derived in Section 7. After that, a cost model is developed in Section 8. Numerical discussions are done in Section 9. Later, some detailed managerial insights and conclusions are given in Section 10.

2. The model formulation

We consider an $M/M/1$ queueing system with the following assumptions:

- Customers entered into the waiting line according to a Poisson process with rate λ.
- After entering the system, waiting customers, during normal mode, get service in a FCFS order following exponential distribution with rate μ_b.
- At the end of the busy period, the service provider immediately goes for a single hiatus following an exponential distribution with rate γ. During this mode, the server offers a service to the customers with a lower rate μ_{wv}; $\mu_{wv} < \mu_b$.
- The server takes a random maximum number, denoted by H, of hiatuses after emptying the system. The probability mass function (pmf) of H is $\mathbb{P}(H = j) = h_j, j = 1, 2, \ \ldots, \ m-1$. At each vacation completion instant, the server checks the system to see if there is any customer waiting and decides the action to take according to the state of the system. The following situations are considered:
  - If at the end of working hiatus completion instant, some customers are waiting in the line, the system returns to a regular busy period. Otherwise, the server takes a hiatus which is exponentially distributed with rate γ_1. During this period, new customers could not be served.
  - At the end of the first hiatus, if there are any beneficiaries waiting for service, the service provider suddenly starts their service. Otherwise, the server goes to the second type of hiatus with rate γ_2. Note that $\gamma_1 \leq \gamma_2$.
  - If the server finds a customer at the jth $(1 \leq j \leq m-1)$ hiatus completion instant, the server immediately switches to regular busy mode; otherwise, he takes another hiatus of type $(j+1)$th which is exponentially distributed with rate γ_{j+1} and continues to have m hiatuses sequentially.
  - As soon as the mth hiatus is completed, the server immediately switches to the regular busy mode and stay there waiting for new arrivals.

  This type of limited number of hiatuses policy reflects the compromise between the benefit of working on the queue and the benefit of performing other tasks represented by the hiatuses.

Remark 1. All random variables involving in the model are assumed to be i.i.d.

Using the matrix analysis method, we can prove that the condition for ergodicity for our vacation system is the same as the condition for the system without vacation, that is $\dfrac{\lambda}{\mu_b} < 1$.

3. Practical application

Consider a situation in which a server provides network services, such as web service, file transfer service, mail service, etc. Apart from the above main tasks, the server has many secondary jobs like virus scanning, assessing hard disk space, checking server log files, installing security software patches, updating antivirus software and so on, as parts of server maintenance. When there are no jobs, a virus scanning is performed to enhance the server performance. During the server performance time, if new job arrives, the server immediately offers them a service in a slow manner. If there are some jobs waiting after the scanning process is complete, the server starts processing them. Otherwise, the server proceeds with another secondary task, namely checking server log files. After the server log files check is complete, if any job is waiting, the server starts working on it, otherwise another secondary task, namely assessing hard disk space, begins. The server continues this processus up to completing all kinds of secondary jobs. Then, he starts a normal busy mode if there are some jobs waiting, or simply waits for new arrivals.

4. The balance equations

The bivariate process $\{(l(t), s(t)), t \geq 0\}$ confines a continuous-time Markov process (CTMP) with state space

$$\triangle = \{(l, i) : \ i = B, wv, h_1, h_2, \ldots, h_m, \ l \geq 0\}$$

where $l(t)$ is the number of beneficiaries in the system and $s(t)$ is the position of the server at time t with

$$s(t) = \begin{cases} h_j, & \text{the server is in } j^{th} \text{ hiatus } j = 1, \ldots, m \\ wv, & \text{the server is on working hiatus} \\ B & \text{the server is busy or available} \end{cases}$$

The state transition diagram of our model is given in Figure 1.

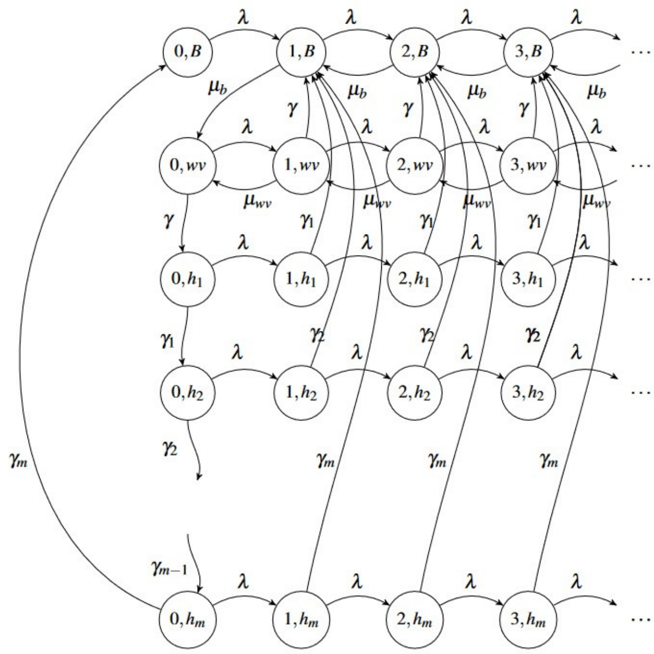

Figure 1. State transition diagram

Let

$$p_{l,\,i} = \lim_{t\to\infty} p_{l,\,i}(t)$$

denote the system state probabilities of the process $\{(l(t), s(t)), t \geq 0\}$. By the Markov theory, the steady-state balance equations of the proposed model are as follows:

$$\lambda p_{0,B} = \gamma_m p_{0,m} \tag{1}$$

$$(\lambda+\mu_b)p_{n,B} = \lambda p_{n-1,B} + \mu_b p_{n+1,B} + \gamma p_{n,wv} + \sum_{n=1}^{m} \gamma_j p_{n,h_j}, \ \ 1 \leq j \leq m, \ \ n \geq 1$$

$$(\lambda+\gamma)\, p_{0,\,wv} = \mu_b p_{1,B} + \mu_{wv} p_{1,\,wv} \tag{2}$$

$$(\lambda+\gamma+\mu_{wv})p_{n,wv} = \lambda p_{n-1,wv} + \mu_{wv} p_{n+1,wv}, \ \ n \geq 1 \tag{3}$$

$$(\lambda+\gamma_1)p_{0,\,h_1} = \gamma p_{0,\,wv} \tag{4}$$

$$(\lambda+\gamma_1)p_{n,h_1} = \lambda p_{n-1,h_1}, \ \ n \geq 1 \tag{5}$$

$$(\lambda+\gamma_j)p_{0,\,h_j} = \gamma_{j-1} p_{0,\,h_{j-1}} \tag{6}$$

$$(\lambda+\gamma_j)p_{n,h_j} = \lambda p_{n-1,h_j}, \ \ 2 \leq j \leq m, \ \ n \geq 1 \tag{7}$$

5. The steady state solution

Let

$$P_i(z) = \sum_{n=1}^{\infty} p_{n,i} z^n, \ \ i = B, wv, h_1, h_2, h_3, \ \dots, \ h_m$$

be the partial probability generating functions. Multiplying equations (1)–(4) by z^n and summing all possible values of n, then re-arranging all the terms, we have

$$P_B(z) = \frac{\mu_b z p_{1,B} - \lambda z^2 p_{0,B} - \gamma z P_{wv}(z) - \sum_{j=1}^{m} \gamma_j z P_{h_j}(z)}{\lambda z^2 - (\lambda+\mu_b)z + \mu_b} \tag{8}$$

In the same way, using (2)–(3) and (4)–(7), we respectively find

$$P_{wv}(z) = \frac{z(\mu_{wv} p_{1,\,wv} - \lambda z p_{0,\,wv})}{\lambda z^2 - (\lambda+\gamma+\mu_{wv})z + \mu_{wv}} \tag{9}$$

and

$$P_{h_j}(z) = \frac{z\lambda p_{0,\,h_j}}{\lambda(1-z)+\gamma_j}, \ \ \ 1 \leq j \leq m \tag{10}$$

Let

$$\alpha = \frac{(\lambda+\gamma+\mu_{wv}) - \sqrt{(\lambda+\gamma+\mu_{wv})^2 - 4\lambda\mu_{wv}}}{2\lambda}$$

be the of denominator root of $P_{wv}(z)$. This is also the root of numerator of (9). From this, we get

$$p_{1,\,wv} = \frac{\lambda\alpha}{\mu_{wv}} p_{0,\,wv} \tag{11}$$

Using equations (11) and (2), we obtain

$$p_{0,\,wv} = \frac{\mu_b p_{1,B}}{\lambda(1-\alpha)+\gamma} \tag{12}$$

Then, by making use of equations (1)–(2), (4), and (6) and after some transformations we obtain

$$p_{0,B} = \frac{\gamma\gamma_m}{\lambda(\lambda+\gamma_1)} \frac{\mu_b p_{1,B}}{\lambda(1-z1)+\gamma} \prod_{j=2}^{m} \left(\frac{\gamma_{j-1}}{\lambda+\gamma_j} \right) \tag{13}$$

$$p_{0,\,wv} = \frac{\mu_b p_{1,B}}{\lambda(1-\alpha)+\gamma} \tag{14}$$

$$p_{0,\,h_1} = \frac{\gamma}{\lambda+\gamma_1} \frac{\mu_b p_{1,B}}{\lambda(1-\alpha)+\gamma} \tag{15}$$

$$p_{0,\,h_j} = \frac{\gamma}{\lambda+\gamma_1} \frac{\mu_b p_{1,B}}{\lambda(1-\alpha)+\gamma} \prod_{j=2}^{m} \left(\frac{\gamma_{j-1}}{(\lambda+\gamma_j)} \right), \quad 2 \le j \le m \tag{16}$$

Taking the derivative of equations (9)–(10) and letting $z \to 1$, we have

$$P'_{wv}(1) = \frac{\lambda(\lambda+\gamma-\mu_{wv})p_{0,\,wv} - \mu_{wv}(\lambda-\mu_{wv})p_{1,\,wv}}{\gamma^2} \tag{17}$$

$$P'_{h_j}(1) = \frac{\lambda(\lambda+\gamma_j)}{\gamma_j^2} p_{0,\,h_j}, \quad 1 \le j \le m \tag{18}$$

From equations (8)–(10), making use of l'Hospital rule, by letting $z \to 1$, we respectively obtain:
The probability that the server is on busy state:

$$P_B = \frac{\mu_b p_{1,B} - (2\lambda p_{0,B} + \gamma(P'_{wv}(1) + P_{wv}(1)) + \sum_{j=1}^{m} \gamma_j (P'_{h_j}(1) + P_{h_j}(1))}{\lambda - \mu_b}.$$

The probability that the server is on working hiatus state:

$$P_{wv} = \frac{\lambda(1-\alpha)}{\gamma} p_{0,\,wv}$$

The probability that the server is on the ith hiatus:

$$P_{h_j} = \frac{\lambda}{\gamma_j} p_{0,\,h_j}, \quad 1 \le j \le m$$

From the total probability, we have

$$p_{1,B} = \frac{1}{\frac{\mu_b}{\lambda(1-\alpha)+\gamma}} \times A \tag{19}$$

where

$$A = \left(\left(\frac{\gamma+\lambda(1-\alpha)}{\gamma} + \frac{\gamma}{\lambda+\gamma_1}\left(1+\frac{\gamma_1}{\gamma_2}\sum_{j=2}^{m}\left(\frac{\gamma_{j-1}}{\gamma_j}\prod_{t=2}^{m-1}\frac{\gamma_{t-1}}{\lambda+\gamma_t}\right)+\frac{\lambda}{\gamma_1}+\frac{\gamma_m}{\lambda}\prod_{t=2}^{m}\frac{\gamma_{t-1}}{\lambda+\gamma_t}\right)\right)+\phi\right)^{-1}$$

and

$$\phi = \frac{\mu_b - \frac{\gamma\mu_b}{(\lambda(1-\alpha)+\gamma)(\lambda+\gamma_1)}\left(2\gamma_m\prod_{j=2}^{m}\frac{\gamma_{j-1}}{\lambda+\gamma_j} - \sum_{j=1}^{m}\left(\frac{\lambda^2+2\lambda\gamma_j}{\gamma_j}\prod_{j=2}^{m}\frac{\gamma_{j-1}}{\lambda+\gamma_j}\right)\right)}{\lambda-\mu_b}$$

$$- \frac{\frac{\mu_b}{\lambda(1-\alpha)+\gamma}\frac{(\lambda^2-\lambda\mu_{wv})(1-\alpha)+\lambda\gamma}{\gamma}}{\lambda-\mu_b}.$$

6. Particular cases

In this section, we present some special models that have been derived by assuming specific values for some parameters involved in the model.

1. ($\gamma \to \infty$ and $\gamma_j \to \infty,\ j = 1, 2, \ldots, m$). The model is reduced into the classical $M/M/1$ queueing model. Here, we have

$$p_{1,B} = \frac{\lambda(\mu_b-\lambda)}{\mu_b{}^2} = \rho(1-\rho)$$

where $\rho = \frac{\lambda}{\mu_b}$. This result coincides with that given in equation (3.2.3) by Medhi in [21].

2. ($\gamma_j \to \infty,\ j = 2, \ldots, m$). The model changes into an $M/M/1$ queueing model with a single vacation followed by a working vacation ($M/M/1 + SWV + SV$). In this case, $p_{1,B}$ becomes:

$$p_{1,B} = \frac{1}{\frac{\mu_b}{\lambda(1-\alpha)+\gamma}\left(1+\frac{\lambda(1-\alpha)}{\gamma}+\frac{\gamma}{\lambda+\gamma_1}\left(1+\frac{\gamma_1}{\lambda}+\frac{\lambda}{\gamma_1}\right)\right)+\phi},$$

where

$$\phi = \frac{\mu_b\left(1-\frac{1}{\lambda(1-\alpha)+\gamma}\left(\frac{\lambda\gamma}{\lambda+\gamma_1}\left(\frac{2\gamma_1}{\lambda}+\frac{\lambda+2\gamma_1}{\lambda+\gamma_1}\right)\right)+\frac{\lambda(2\gamma+\lambda-\mu_{wv})(1-\alpha)}{\gamma}\right)}{\lambda-\mu_b}$$

3. ($\gamma \to \infty$). The model changes into an $M/M/1$ queueing model with multiple adaptive vacations

$$P_{h_1}(z) = \frac{\mu_b}{\lambda(1-z)+\gamma_1} p_{1,B}$$

$$P_{h_j}(z) = \frac{\gamma_{j-1}}{\lambda+\gamma_j-\lambda z} p_{0,\,h_{j-1}}, \;\; 2 \le j \le m$$

$$P_B(z) = \frac{\mu_b z p_{1,B} - \lambda z^2 p_{0,B} - \sum_{j=1}^{m} \gamma_j z P_{h_j}(z)}{\lambda z^2 - (\lambda+\mu_b)z + \mu_b}$$

after transformation we get

$$p_{1,B} = \left(\frac{\mu_b}{\lambda+\gamma_1} \left(1 + \frac{\lambda}{\gamma_1} + \lambda \sum_{j=2}^{m} \left(\frac{1}{\gamma_j} \prod_{j=2}^{m} \frac{\gamma_{t-1}}{\lambda+\gamma_t} \right) + \sum_{j=2}^{m} \prod_{t=2}^{m} \frac{\gamma_{t-1}}{\lambda+\gamma_t} + \frac{\gamma_m}{\lambda} \prod_{j=2}^{m} \frac{\gamma_{j-1}}{\lambda+\gamma_j} \right) + \phi \right)^{-1}$$

where

$$\phi = \frac{\mu_b - \frac{\mu_b}{\lambda+\gamma_1} \left(2\gamma_m \prod_{j=2}^{m} \frac{\gamma_{j-1}}{\lambda+\gamma_j} - \frac{\lambda^2+2\lambda\gamma_1}{\gamma_1} - \sum_{j=2}^{m} \left(\frac{\lambda^2+2\lambda\gamma_j}{\gamma_j} \prod_{j=2}^{m} \frac{\gamma_{j-1}}{\lambda+\gamma_j} \right) \right)}{\lambda - \mu_b}$$

Further, important performance measures are derived. Expected number of beneficiaries in the system when the service provider is busy

$$E(L_B) = \frac{(\mu_b-\lambda)\left(2\lambda p_{0,B} + \frac{2\lambda\mu_b p_{1,B}(\lambda+\gamma_1)}{\gamma_1^2} + \sum_{i=2}^{m} \frac{2\lambda(\lambda+\gamma_i)\gamma_{i-1} p_{0,\,h_{i-1}}}{\gamma_i^2} \right)}{2(\mu_b-\lambda)^2}$$

$$+ \frac{\left(2\lambda p_{0,B} - \mu_b p_{1,B} + \frac{\mu_b(\lambda+\gamma_1)p_{1,B}}{\gamma_1} + \sum_{i=2}^{m} \frac{\gamma_{i-1}(\lambda+\gamma_i)p_{0,\,h_{i-1}}}{\gamma_i} - \sum_{i=1}^{m} \gamma_i p_{0,\,h_i} \right)(2\lambda)}{2(\mu_b-\lambda)^2}$$

Expected number of beneficiaries in the system when the service provider is on jth vacation $(1 \le j \le m)$:

$$E(L_{h_1}) = \frac{\lambda\mu_b p_{1,B}}{\gamma_1^2}, \quad E(L_{h_j}) = \frac{\lambda\gamma_{j-1} p_{0,\,h_{j-1}}}{\gamma_j^2}, \quad 2 \le j \le m.$$

7. Performance measures

In this section, we give some performance measures of the model that will be useful for the numerical part. Expected number of beneficiaries in the system when the service provider is on working vacation is

$$E(L_{wv}) = \frac{\lambda(\gamma + \lambda - \mu_{wv})p_{0,\,wv} - \mu_{wv}(\lambda - \mu_{wv})p_{1,\,wv}}{\gamma^2}.$$

Expected number of beneficiaries in the system when the service provider is on jth vacation:

$$E(L_{h_j}) = \frac{(\lambda + \gamma_j)\lambda p_{0,\,h_j}}{\gamma_j^2}, \quad 1 \le j \le m.$$

Expected number of beneficiaries in the system when the service provider is on is on busy period:

$$E(L_B) = \frac{(\mu_b - \lambda)\left(2\lambda p_{0,B} + \frac{2\lambda\mu_b p_{1,B}(\lambda+\gamma_1)}{\gamma_1^2} + \sum_{i=2}^{m} \frac{2\lambda(\lambda+\gamma_i)\gamma_{i-1}p_{0,\,h_{i-1}}}{\gamma_i^2}\right)}{2(\mu_b - \lambda)^2}$$

$$+ \frac{\left(2\lambda p_{0,B} - \mu_b p_{1,B} + \frac{\mu_b(\lambda+\gamma_1)p_{1,B}}{\gamma_1} + \sum_{i=2}^{m} \frac{\gamma_{i-1}(\lambda+\gamma_i)p_{0,\,h_{i-1}}}{\gamma_i} - \sum_{i=1}^{m} \gamma_i p_{0,\,h_i}\right)(2\lambda)}{2(\mu_b - \lambda)^2}$$

where

$$S = \frac{4\lambda^2\gamma + 2\lambda^2\gamma\alpha - 2\gamma\lambda\mu_{wv} - (1-\alpha)(4\lambda^2 - 2\lambda^3 - 2\lambda\gamma\mu_{wv} - 2\lambda\mu_{wv}^2 + 3\lambda^2\gamma - \lambda\gamma\mu_{wv} + \lambda\gamma^2)}{\gamma^3}$$

Expected number of beneficiaries in the system $E(L)$:

$$E(L) = E(L_B) + E(L_{wv}) + \sum_{j=1}^{m} E(L_{h_j})$$

Expected waiting time: $E(W)$

$$E(W) = E(L)/\lambda$$

8. Cost model

In this section, we develop a cost model by defining the total expected function, in which the service rate is the dominating variable. Our goal is to control these variables in order to minimize the total expected cost function per unit time. We define the following cost elements (per unit time) as follows:

C_N– holding cost for each beneficiary seen in the system,
C_W– waiting cost if one beneficiary is to receive the service,
C_B– cost for the period the server handling service process,
$C_{w\overline{v}}$– cost for the period the server handling working hiatus process,
C_j – cost when the server is on the jth hiatus, $(1 \le j \le m)$
C_{μ_b}– cost for service

The cost function TC is defined as:

$$TC = C_N E(L) + C_W E(W) + C_B P_B + C_{wv} P_{wv} + \sum_{j=1}^{m} C_j P_{h_j} + C_{\mu_b} \mu_b \tag{20}$$

Our aim can be expressed mathematically as:

$$\text{Minimize } TC(\mu_b)$$

9. Numerical discussions

The problem considered in Section 3 can be modeled as an $M/M/1/WV\&MAV$ queueing system. In this section, we present numerically the performance indices and the optimality of the cost model for the considered system. Throughout this section, we assume that the maximum number of hiatuses taking by the service provider is two, that is $m = 2$.

9.1. Analysis on performance indices

Service rate μ_b. The impact of service rate μ_b on the performance indices P_B, $E[L]$, and $E[W]$ are displayed in Figures 2–4, respectively. Obtained results are clearly admitted with our intuition. The increase in the service rate leads to a significant decrease in the considered performance indices.

Arrival rate λ. The effect of an increase in the arrival rate on the performance indices P_B, $E[L]$, and $E[W]$ are presented in Figures 5–7, respectively. It is well observed that a surge in the arrival rate implies a significant decrease in the above three performance indices. Obviously, high number of customers joining the system implies a high probability of busy period. This results in an increase in the average system length and average waiting time of customers.

Working vacation service rate μ_{wv}. The effect of an increase in the arrival rate on the performance indices P_{WV}, $E[L]$, and $E[W]$ are presented in Figures 8–10, respectively. We noticed a considerable decrease in the above three performance indices with μ_{wv}. It is obvious that if the service provider increases the service rate while is on working vacation mode, all customers being in queue will be served. Therefore, P_{h_1} increases accordingly. This leads to unavailability of service to the customers. Hence, a decrease in the average system length and average waiting time of customers is quite reasonable.

Hiatus rates γ_1 and γ_2. The effect of hiatus rate, γ_1 (γ_2) on the performance indices P_{h_1}(resp. P_{h_2}), $E[L]$, and $E[W]$ are studied in Figures 11–13, (Figures 14–16). By increasing the first hiatus rate (γ_1), a considerable decrease in the above three performance indices occurs. It is vivid that by augmenting the first hiatus rate, the probability that the server stays on type-I hiatus becomes small, this implies an increase in P_{h_2}. Further, an increase in the hiatus type 2 rate, results in a considerable decrease in P_{h_2}, $E[L]$, and $E[W]$.

9.2. The cost optimization

The total cost function TC given in section 8 is nonlinear; it is too complex to solve analytically. Then, the optimal value of the service rate can be easily determined by using numerical optimization techniques.

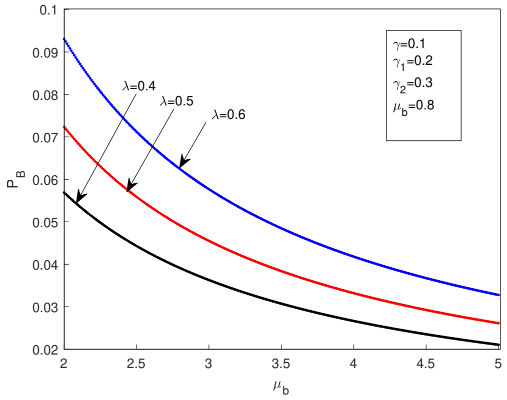

Figure 2. P_B against μ_b

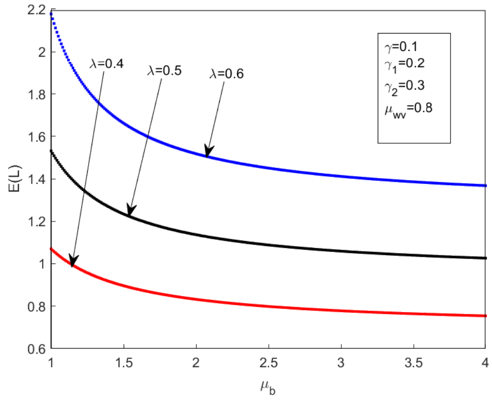

Figure 3. $E[L]$ against μ_b

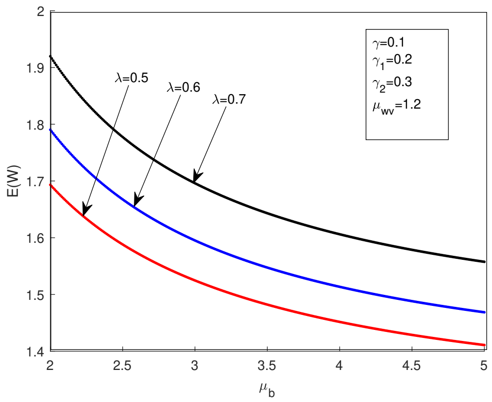

Figure 4. $E(W)$ against μ_b

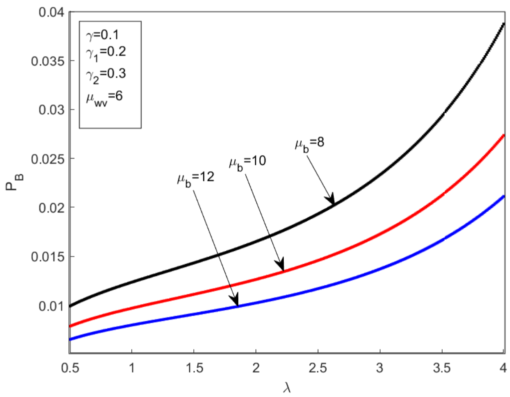

Figure 5. P_B against λ

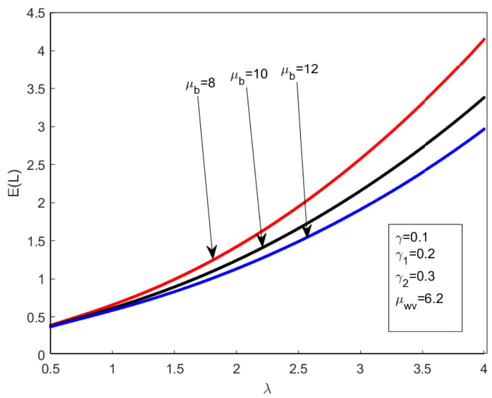

Figure 6. $E(L)$ versus λ

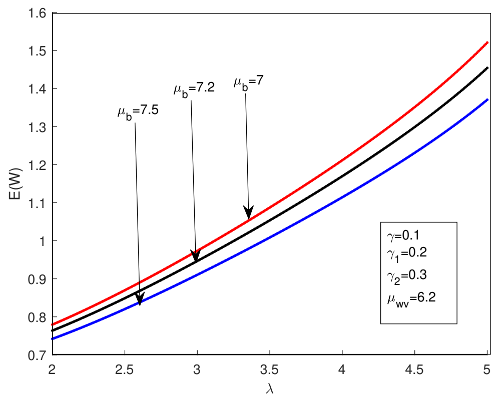

Figure 7. E(W) against λ

9.2.1. The optimal service rate

We apply the PSO algorithm to optimize the expected cost function. For this purpose, we employ $C_N = 10, C_W = 15, C_B = 20, C_{wv} = 20, C_{h_1} = 25, C_{h_2} = 30, C_{\mu_b} = 40$. For three different values of the arrival rate, obtained results are displayed in Figure 17. It is seen that the curves are convex and hence assure the existence of optimal values. In addition, it is observed that with λ the concerned optimal service and the expected cost increase accordingly, as intuitively expected.

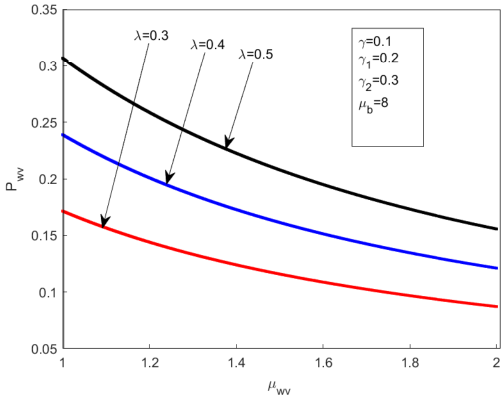

Figure 8. P_{wv} against μ_{wv}

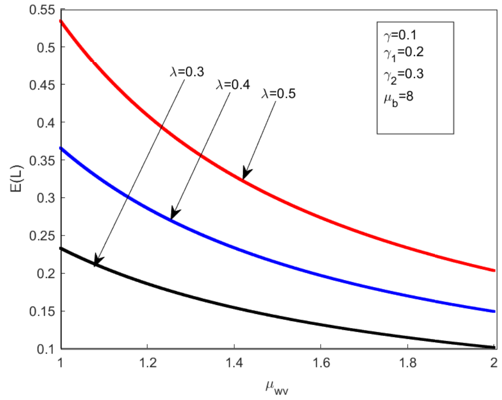

Figure 9. $E(L)$ against μ_{wv}

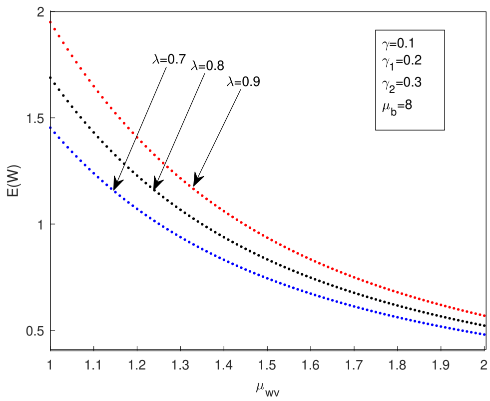

Figure 10. $E(W)$ against μ_{wv}

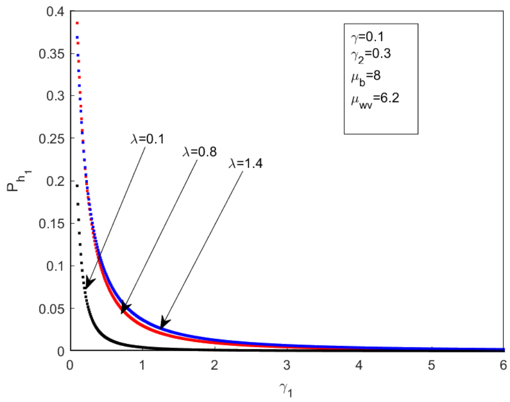

Figure 11. P_{h_1} against γ_1

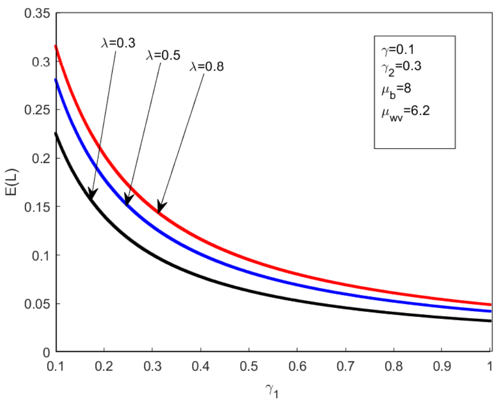

Figure 12. E(L) against γ_1

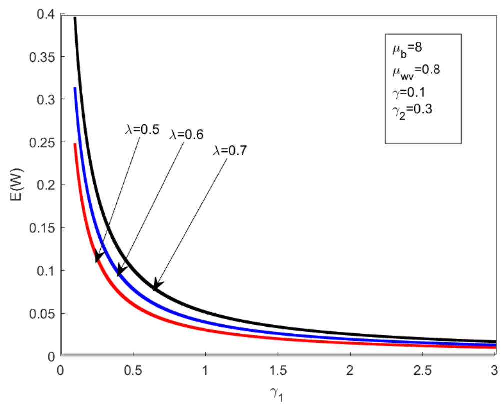

Figure 13. E(W) against γ_1

9.2.2. Comparison of total cost obtained from PSO algorithm and NR method

The Newton–Raphson algorithm and PSO algorithm are applied and the results are computed using an Intel-Pentium 2.4 GHz PC with 1.97 GB RAM. From Table 1, we observe the compliance in the optimal service rate obtained by applying the NR and PSO algorithms from a few experiments. Also, we can see that NR method is better for obtaining the optimal service rate of our model as the running time (in seconds) of the algorithm is less compared with the running time of the PSO algorithm.

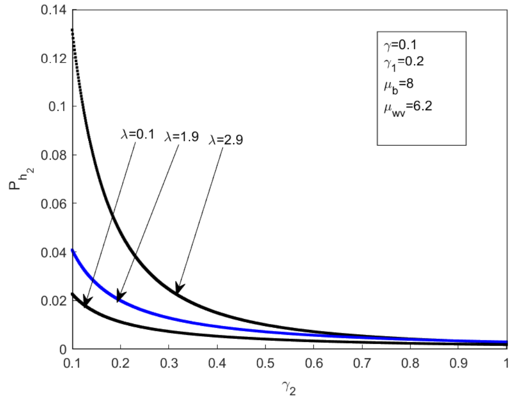

Figure 14. P_{h_2} against γ_2

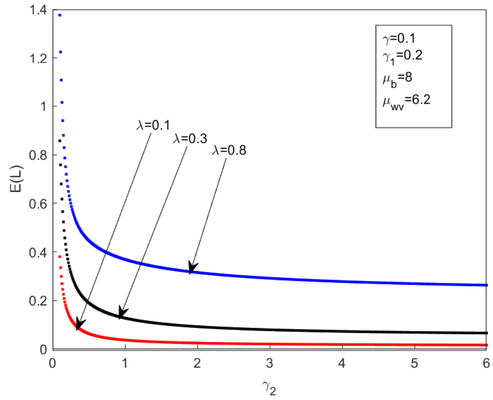

Figure 15. $E(L)$ against γ_2

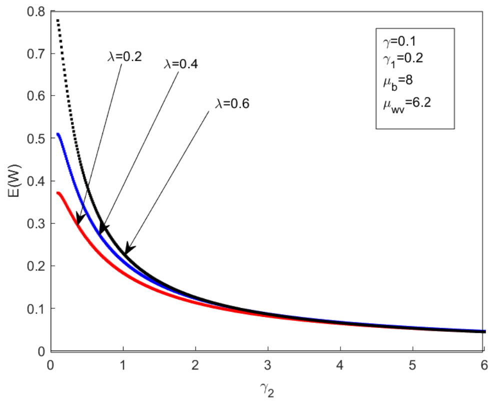

Figure 16. $E(W)$ against γ_2

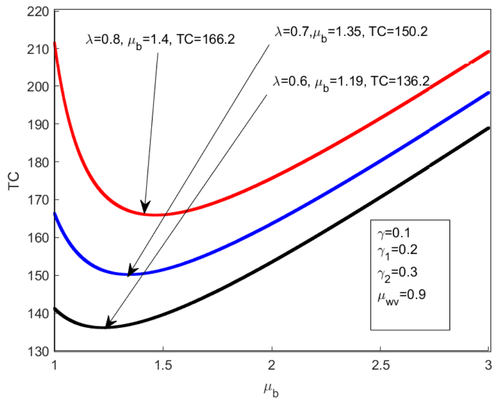

Figure 17. TC against μ_b

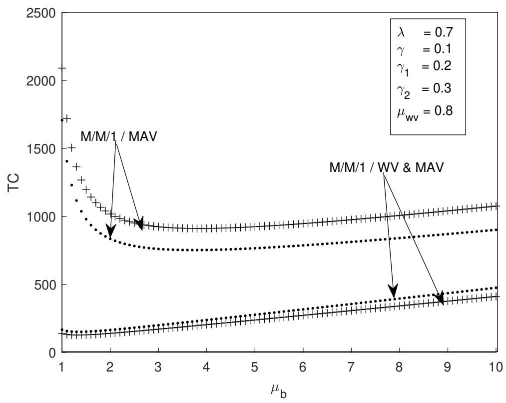

Figure 18. TC against μ_b

9.3. Comparative numerical study on the total cost

In this numerical discussion, we give a comparative study on the optimal service rate and the relative total expected costs of $M/M/1/MAV$ queue, discussed as a third special case and our model $M/M/1/WV\&MAV$ queue. Note that for the $M/M/1/MAV$ queue the cost function is defined as:

$$TC = C_N E(L) + C_W E(W) + C_B P_B + \sum_{j=1}^{m} C_j P_{h_j} + C_{\mu_b}\mu_b \tag{21}$$

Here, we take two set of values as
$C_N = 10; C_W = 20; C_{WV} = 25; C_B = 20; C_1 = 30; C_2 = 35; C_{\mu_b} = 40$ and $C_N = 10; C_W = 15; C_{WV} = 25; C_B = 20; C_1 = 25; C_2 = 30; C_{\mu_b} = 35$ in (20) and (21). The resulting curves to the above two set of values are plotted in figure 18 which shows that the optimal service rate of $M/M/1/WV\&MAV$ queue is less than the optimal service rate of $M/M/1/MAV$ queue. Also, the curve of TC for our model against the service rate is under the curve of TC for the $M/M/1/MAV$ queue. It reveals that our model is better compared with the model $M/M/1/MAV$ queue.

Table 1. Effect of ${\mu_b}^*$ on TC^*, $E(L)$, $E(W)$ for $\gamma = 0.1, \gamma_1 = 0.2, \gamma_2 = 0.3$, and different values of λ

λ	Newton–Raphson optimization					PSO				
	μ_b^*	$E(L)$	$E(W)$	TC^*	Elapsed time [s]	μ_b^*	$E(L)$	$E(W)$	TC^*	Elapsed time [s]
0.1	0.6749	0.2350	2.3501	62.8923	0.023001	0.6749	0.2350	2.3501	62.8923	180.6771
0.2	0.7706	0.4818	2.4091	72.3742	0.015456	0.7706	0.4818	2.4091	72.3742	151.9986
0.3	0.8862	0.7321	2.4402	80.8567	0.01125	0.8862	0.7321	2.4402	80.8567	158.7501
0.4	1.005	1,0024	2.5061	89.6534	0.013178	1.005	1,0024	2.5061	89.6534	114.0655
0.5	1.1246	1.3064	2.6128	99.199	0.14212	1.1246	1.3064	2.6128	99.199	359.2317
0.6	1.2452	1.6579	2.7632	109.7866	0.012472	1.2452	1.6579	2.7632	109.7866	460.2642
0.7	1.3676	2.0717	2.9595	121.6516	0.024794	1.3676	2.0717	2.9595	121.6516	753.9248
0.8	1.4924	2.5611	3.2013	134.93	0.022484	1.4924	2.5611	3.2013	134.93	386.0446

10. Managerial insights and conclusions

This study focuses on optimizing the cost of a Markovian queueing system subjected to multiple adaptive vacation and working vacation policies using both the Newton–Raphson and particle swarm optimization techniques. This system is employed to model a wide variety of real-world situations, including customer service centers, transportation systems, production and manufacturing systems, etc. The multiple adaptive vacation and working vacation policies are a fundamental aspect of this optimization process that can have a strongly positive impact on the system's performance. After presenting the theoretical analysis of the queueing model, the relationship between different system parameters and performance measures, as well as the cost model of the queueing system have been examined. The current study provides valuable insights for managers looking to optimize their queueing systems and reduce costs. These insights can be used by managers to make informed decisions and improve the efficiency of their operations.

The use of the Newton–Raphson and PSO techniques to study the cost optimization of the suggested queueing model is a complex problem that requires careful consideration of multiple factors:

- The impact of vacation policies on system performance should be examined.
- It is important to define the cost function to reflect the objectives of the organization, such as maximizing profit or minimizing cost.
- The optimization problem must be defined. This involves finding the optimal configuration of the system that minimizes the cost function while satisfying performance constraints.
- After implementing the optimization techniques, the results should be evaluated while monitoring the system performance to ensure that the system is operating optimally and considering the impact of variability.

A continuous-time queueing model with working vacation and adaptive vacation policy permits the server to take a random maximum number of vacations as long as the system is empty. With this policy, the queue manager can better load the server than multiple or a single vacation policies. The steady-state solution of the proposed system has been established. Moreover, on the basis of the steady-state probabilities of the system, its performance measures have been derived. Even if the convexity (unimodality) of the expected cost function can not be demonstrated theoretically in this study, we present a finite and quick search for the optimal thresholds using the PSO algorithm and Newton–Raphson method. We also perform a sensitivity analysis among the optimal value of the service rate and cost components. The two method applied were numerically compared. The proposed model is very useful in general situations arising in practical applications.

Acknowledgement

The authors thank the referees for the helpful suggestions and comments to improve the quality of the manuscript.

References

[1] Bouchentouf, A. A., Boualem, M., Yahiaoui, L., and Ahmad, H. A multi–station unreliable machine model with working vacation policy and customers impatience. *Quality Technology and Quantitative Management 19*, 6 (2022), 766–796.

[2] Bouchentouf, A. A., Cherfaoui, M., and Boualem, M. Performance and economic analysis of a single server feedback queueing model with vacation and impatient customers. *Opsearch 56*, 1 (2019), 300–323.

[3] Bouchentouf, A. A., and Guendouzi, A. Sensitivity analysis of multiple vacation feedback queueing system with differentiated vacations, vacation interruptions and impatient customers. *International Journal of Applied Mathematics and Statistics 57*, 6 (2018), 104–121.

[4] Bouchentouf, A. A., Medjehri, L., Boualem, M., and 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*, 1 (2022), 21–38.

[5] Cheng, J., and Tang, Y. Reliability analysis of $M/G/1$ repairable queueing system with multiple adaptive vacations and p-entering disciplin. *Mathematical and Computational Applications 19*, 2 (2014), 105–114.

[6] Doshi, B. T. Queueing systems with vacation-a survey. *Queueing Systems 1*, 1 (1986) 29–66.

[7] Jeyakumar, S., and Rameshkumar, E. Analysis of $M^X/G(a,b)/1$ queue with closedown time with controllable arrivals during multiple adaptive vacations. *International Journal of Pure And Applied Mathematics 106*, 5 (2016), 79–87.

[8] Jeyakumar, S., and Rameshkumar, E. Performance analysis and cost optimization of nonMarkovian bulk queue with 'p'–entering discipline during multiple adaptive vacations international. *Journal of Information and Management Sciences 28*, (2017), 99–111.

[9] Jeyakumar, S., and Rameshkumar, E. Binomial service and multiple adaptive vacation schedules for $M^X/G/1$ queue with control policy on demand for re-service. *Nonlinear Studies 24*, 2 (2017), 417–428.

[10] Jeyakumar, S., and Rameshkumar, E. A study on $M^X/G(a,b)/1$ queue with server breakdown without interruption and controllable arrivals during multiple adaptive vacations *International Journal of Mathematics in Operational Research 15*, 2 (2019), 137–155.

[11] Kalidass, K., and Kasturi, R. A queue with working breakdowns. *Computers and Industrial Engineering 63*, 4 (2012), 779–783.

[12] Kalidass, K., Gnanaraj, J., Gopinath, S., and Kasturi, R. Transient analysis of an $M/M/1$ queue with a repairable server and multiple vacations. *International Journal of Mathematics in Operational Research 6*, 2 (2014), 193–216.

[13] Ke, J.-C, Chang, F.-M., and Liu, T.-H. $M/M/c$ balking retrial queue with vacation. *Quality Technology and Quantitative Management 16*, 1 (2019), 54–66.

[14] Kempa, W. M., and Marjasz, R. Distribution of the time to buffer overflow in the $M/G/1/N$-type queueing model with batch arrivals and multiple vacation policy. *Journal of the Operational Research Society 71*, 3 (2020), 447–455.

[15] Kempa, W. M., Książek, K., and Marjasz, R. On time-dependent queue-size distribution in a model with finite buffer capacity and deterministic multiple vacations with applications to LTE DRX mechanism modeling. *IEEE Access 9*, (2021), 148374–148383.

[16] Kobielnik, M., and Kempa, W. M. On the time to buffer overflow in a queueing model with a general independent input stream and power-saving mechanism based on working vacations *Sensors 21*, 16 (2021), 5507.

[17] Levy, Y., and Yechiali, U. Utilization of idle time in an $M/G/1$ queueing system. *Management Science 22*, 2 (1975), 202–211.

[18] Ma, Z., and Xu , Q. General decrementing service $M/G/1$ queue with multiple adaptive vacations. *Applied Mathematics and Computation 204*, 1 (2008), 478–484.

[19] Ma, Z., Yue, W., and Chen, L. Analysis and performance optimization of a $Geom/G/1$ queue with general limited service and multiple adaptive vacations. *Pacific Journal of Optimization 11*, 1 (2015), 57–78.

[20] Majid, S., Bouchentouf, A. A., and Guendouzi, A. Analysis and optimisation of a $M/M/1/WV$ queue with Bernoulli schedule vacation interruption and customer's impatience. *Acta Universitatis Sapientiae, Mathematica 13*, 2 (2021), 367–395.

[21] Medhi, J. *Stochastic Models in Queueing Theory*. Academic Press, 2003.

[22] Saffer, Z., and Yue, W. $M/G/1$ multiple vacation model with balking for a class of disciplines. *Quality Technology and Quantitative Management 12*, 3 (2015), 383–407.

[23] Seenivasan, M., and Abinaya, R. Markovian queueing model with single working vacation and catastrophic *Materials Today Proceedings 51*, 8 (2022), 2348–2354.

[24] Servi, L. D., and Finn, S. G. $M/M/1$ queue with working vacations $(M/M/1/WV)$. *Performance Evaluation 50*, 1 (2002), 41–52.

[25] Sudhesh, R., Azhagappan, A., and Dharmaraja, S. Transient analysis of $M/M/1$ queue with working vacation, heterogeneous service and customers' impatience, *RAIRO - Operations Research 51*, 3 (2017) 591–606.

[26] Sun, W., Tian, N., and Li, S. Steady state analysis of the batch arrival $Geo/G/1$ queue with multiple adaptive vacations *International Journal of Management Science and Engineering Management 2*, 2 (2007), 83–97.

[27] Takagi, H. *Queueing Analysis: A Foundation of Performance Analysis*, Vol.1. *Vacation and Priority Systems*, Part I, Elsevier, 1991.

[28] Tian, N., Li, Q.-L ., and Gao, J. Conditional stochastic decompositions in the $M/M/c$ queue with server vacation. *Stochastic Models 15*, 2 (1999), 367–377.

[29] Tian, N., and Zhang, Z. G. *Vacation Queueing Models: Theory and Applications*, Springer, New York, 2006.

[30] Vadivukarasi, M., Kalidass, K., and Jayaraman, R. Discussion on the Optimization of Finite Buffer Markovian Queue with Differentiated Vacations In *Soft Computing: Theories and Applications* (Singapure, 2022), T. K. Sharma, C. W. Ahn, O. P. Verma and B. K. Panigrahi, Eds., Springer, pp. 523–534.

[31] Vadivukarasi, M., and Kalidass, K. A. Discussion on the optimality of bulk entry queue with differentiated hiatuses. *Operations Research and Decisions 32*, 2 (2022) 137–150.

[32] Wang, F., Wang, J., and Zhang, F. Equilibrium customer strategies in the $Geo/Geo/1$ queue with single working vacations. *Discrete Dynamics in Nature and Society 2014*, 1 (2014), 309489.

[33] Yang, D. Y., Wang, K. H., and Wu, C. H. Optimization and sensitivity analysis of controlling arrivals in the queueing system with single working vacation. *Journal of Computational and Applied Mathematics 234*, 2 (2010), 545–556.

[34] Yang, D. Y., and Wu, C. H. Performance analysis and optimization of a retrial queue with working vacations and starting failure. *Mathematical and Computer Modelling of Dynamical Systems 25*, 5 (2019), 463–481.

[35] Ye, Q., and Liu, L. Performance Analysis of the $GI/M/1$ Queue with Single Working Vacation and Vacations. *Methodology and Computing in Applied Probability 19* (2016), 685–714.

[36] Zhang, Z. G., and Tian, N. Discrete Time $Geo/G/1$ Queue with Multiple Adaptive Vacations. *Queueing Systems 38*, 4 (2001), 419–429.

[37] Zhang, Z. G., and Tian, N. Analysis on queueing systems with synchronous vacations of partial servers *Performance Evaluation 52*, 4 (2003), 269–282.