OPTIMIZATION ANALYSIS OF UNRELIABLE MULTI-SERVER QUEUEING SYSTEM WITH BERNOULLI SCHEDULE WORKING VACATION, THRESHOLD-BASED RECOVERY POLICY, AND IMPATIENCE

Hayat Ramdani[1], Amina Angelika Bouchentouf[2]
and Lahcene Yahiaoui[3]

•

[1]Laboratory of Mathematics of Sidi Bel Abbes,
Kasdi Merbah University, Ouargla, Algeria.
[2]Djillali Liabes University of Sidi Bel Abbes, Sidi Bel Abbes, Algeria.
[3]Laboratory of Stochastic Models, Statistic and Applications,
University of Saida-Dr. Moulay Tahar, Saida, Algeria.
ramdanihayat30@gmail.com bouchentouf_amina@yahoo.fr lahceneya8@gmail.com

Abstract

This paper analyzes an unreliable multi-server queueing system incorporating working vacations, Bernoulli interruptions, breakdowns with a threshold recovery policy, balking, abandonment, and retention. During the break period, if there are customers in the queue, the servers may either resume normal service or continue their vacation. Customers arriving while the system is saturated are rejected. Failures occur unexpectedly but only when at least one customer is present in the system. Recovery procedures remain in effect until the total number of customers surpasses a predefined threshold. Using matrix-analytic methods, we derive steady-state solutions and explicit formulas for various performance indicators. Further, we explore cost parameter optimization.

Keywords: unreliable queueing systems, threshold-based recovery policy, working vacation, impatience

1. Introduction

With growth of communication systems and networks, manufacturing systems, transportation systems, etc, queueing systems with breakdowns have received growing significance [14, 17, 18].

Queueing models incorporating threshold policies, specifically the N-policy and F-policy, have garnered significant attention in recent years. The former policy dictates that a server activates only when N (where $N \geq 1$) or more customers accumulate in the system [11, 19, 5, 25]. Conversely, the latter policy restricts customer entry into the system once it reaches its capacity. When the queue length decreases to a threshold parameter value F, the server then permits customers to enter [9, 4, 12].

The literature on N and F policies is extensive. However, research on queueing models with breakdowns, repairs, and a threshold-based recovery policy, where the server remains unrepaired until the number of customers in the system reaches a predetermined threshold value, is limited. Notable works include [22, 10, 15].

Vacation queueing models have attracted substantial interest from researchers over the past decades, owing to their ubiquitous applications across diverse fields. These applications span production/manufacturing systems, telecommunication systems and computer networks. Notably, comprehensive surveys on this subject have been conducted by [6, 7, 20, 21].

The concept of working vacations was introduced by [16], proposing a model where the server processes jobs with varying intensities based on the incoming traffic. The primary objective is twofold: better control of queue lengths and reduction of customer loss. Additionally, working vacations enable servers to be strategically redirected for maintenance purposes. As a result, these models have gained significant popularity, leading to a wealth of analytical results in the literature, such as [26, 8, 24, 23].

In recent times, queueing systems that account for customer impatience have garnered significant attention. These models find realistic applications in various service systems and e-commerce domains. For a comprehensive overview of the literature on this theme, readers can refer to studies by [13, 2, 3, 1].

In this paper, we delve into the analysis of a multi-server Markovian queue that integrates several crucial practical features including breakdowns, threshold-based recovery policy, working vacations, Bernoulli interruption schedule, impatient customers, and retention of reneged customers. The contributions and advantages of this paper are as follows:

1. ***The model.*** Unlike existing literature that predominantly focuses on single-server queueing models, our study embraces a multi-server queue. By incorporating the diverse features mentioned above, our proposed model offers greater flexibility in characterizing complex stochastic phenomena within multi-server machining systems.

2. ***Methodology and results.*** Leveraging the Q-matrix method, we provide a detailed theoretical analysis. We derive steady-state probabilities and various performance measures. Our chosen method is well-suited for analyzing quasi-birth-and-death (QBD) processes in steady-state.

3. ***Numerical illustrations.*** We develop a cost function to optimize service rates during both working vacation and normal busy periods. Additionally, we determine the optimal number of servers and explore threshold-based recovery policies. These insights empower system managers and decision-makers to regulate the system economically.

The manuscript is structured concisely in the following manner: Section 2 presents the main motivation and practical applications of the current research work. Section 3 briefly describes the model under consideration. Section 4 comprises the analysis of the model in the stationary state. Section 5 enlists important performance measures. Section 6 develops a cost model for the proposed system and introduces cost optimization methods, namely, the direct search method and the quasi-Newton method. Section 7 deals with a cost optimization problem and provides numerical examples to illustrate the effects of different system parameters on performance measures, total expected cost, and total expected profit. Section 8 presents the conclusions of the study.

2. Main motivation and practical application

The motivating context for our model is analysis of automated teller machine (ATM) manufacturing systems. Such facilities commonly face machine failures and repairs, congestion issues, operator unavailability, impatient customers, and more that can significantly hamper production efficiency.

Specifically, we consider a production system with c parallel machines and finite finished goods capacity. Upon arrival of failed parts/subassemblies for repair, they immediately occupy any available operator. Otherwise failed units wait in queue for a random duration. Once all repairs are completed, operators take group vacations, relying on substitutes with slower service rates, and may have their breaks interrupted if failures resume.

Moreover, operators undergo their own failures following a breakdown process. Repairs only initiate after M failed machines have accumulated via a threshold policy. Newly arriving failures may balk from the repair queue or later renege after prolonged waits.

All such issues–breakdowns, vacations, congestion, balking and reneging–are commonly faced by real ATM manufacturers. By mathematically capturing these dynamics in a closed-form queueing model, we aim to evaluate the complex tradeoffs between maintainability, throughput, and customer impatience. The model can help optimize the number of machines, the threshold-based recovery policy, and service rates, to control costs in ATM production systems through resilience to inevitable disruptions.

3. Model description

Consider an Automated Manufacturing System modeled as an unreliable $M/M/c/L$ queueing system. The model formulation necessitates several distinct assumptions, which can be summarized as follows:

(i) Arrival process: Customers arrive following Poisson process with parameter α.

(ii) Service and working vacation processes:

  (a) Upon arrival, customers are served if any servers are available.
  (b) After serving all existing customers, servers synchronously switch to a vacation period.
  (c) Upon returning from vacation, if the system remains empty, servers immediately begin another synchronous vacation.
  (d) The vacation duration follows an exponential distribution with parameter τ.
  (e) During vacation, substitute servers take over from the main servers to serve new customers.
  (f) Service times during regular busy periods (RBP) and vacations follow exponential distributions with parameters μ and ν, respectively. We assume that $\nu < \mu$.
  (g) If a customer arrives and finds any of the c servers free (during busy or working vacation), they immediately occupy that server. If all servers are busy, the customer joins the end of the queue in the buffer and is served later according to the First-Come-First-Served (FCFS) discipline.

(iii) Bernoulli interruption scheme:

  (a) During the working vacation period (WVP), the server operates under the Bernoulli rule. Specifically, at the instant of service completion during this period, if there are customers in the system:
    - With probability β, the server interrupts the vacation and switches to the regular working period.
    - With probability $\beta' = 1 - \beta$, the server continues the vacation.
  (b) Notably, the service during WVP is applied only to the first customer who arrives during this period.

Then, we can write

$$\delta_n = n\beta\nu \mathbb{1}_{2\leq n\leq c-1} + c\beta\nu \mathbb{1}_{c\leq L}.$$

(iv) Breakdown process: The system is susceptible to unreliability at any given time. During regular busy periods, servers are vulnerable to breakdowns. Specifically, a server break down only if there is at least one customer in the system. The occurrence of breakdowns follows a stationary Poisson process with parameter φ. Importantly, during repair periods (RP), customers cannot be served.

(v) The threshold-based recovery policy and repair process: The recovery can be performed when M $(1 \leq M \leq c-1)$ or more customers are present. The repair period has exponential distribution with parameter γ. Customers arriving during the repair time are ignored by the system.

(vi) Balking: When a customer arrives, their actions depend on server availability:

- - If some servers are working and others are free, the customer is directly served.
- - Otherwise, during working vacation, regular busy, or repair periods:
  (a) The customer may join the queue with probability θ_n.
  (b) Customers faced with joining a queue have an alternative: they may balk, choosing not to enter. The balking probability is denoted as: $\theta'_n = 1 - \theta_n$, where in the case of working vacation/regular busy period, we have : $0 \leq \theta_{n+1} \leq \theta_n \leq 1$. Consider the following scenarios:
    i. For working vacation/regular busy period case, we have:
      – $0 \leq \theta_{n+1} \leq \theta_n \leq 1$ for $c \leq n \leq L-1$;
      – $\theta_0 = 1, \ldots, \theta_{c-1} = 1$.
    ii. For repair period, we observe:
      – $0 \leq \theta_{n+1} \leq \theta_n \leq 1$ for $1 \leq n \leq L-1$.
      – $\theta_0 = 1$ (no balking when the system is empty).
    iii. In both cases, we have: $\theta_L = 0$ (no entering when the system is at full capacity).

Shortly, we have for working vacation and regular normal busy:

$$\alpha_n = \alpha \mathbb{1}_{n<c} + \theta_n \alpha \mathbb{1}_{c \leq n \leq L},$$

and for breakdown period: $\alpha_n = \theta_n \alpha, 1 \leq n \leq L$.

(vii) Reneging and retention:

(a) Upon arrival, customers exhibit different behaviors based on the server status:
  - If servers are in regular working mode or working vacation period:
    – The customer activates an impatience timer T_1 (for regular working) or T_0 (for working vacation). If the customer's service is not completed before the timer expires, they may abandon the system.
  - During the reparation period:
    – A new arrival activates its own timer T_2. If service is unavailable before the expiration of the impatience timer, the customer may give up.

(b) The impatience time T_j follows an exponentially distributed random variable with rates $\varsigma_j > 0$ (where $j = 0, 1, 2$).

(c) Impatient customers have two options:
  - They may quit the system without receiving service with probability κ.
  - Alternatively, they may be kept in the system with probability $\kappa' = 1 - \kappa$.

Then, we can put:

$$\epsilon_{n,j} = n\kappa\varsigma_0 \mathbb{1}_{j=0} + n\kappa\varsigma_1 \mathbb{1}_{j=1} + n\kappa\varsigma_2 \mathbb{1}_{j=2},$$
$$v_n = (n\mu + \epsilon_{n,1})\mathbb{1}_{1 \leq n \leq c-1} + (c\mu + \epsilon_{n,1})\mathbb{1}_{c \leq n \leq L}$$

and

$$\zeta_n = (\nu + \epsilon_{1,0})\mathbb{1}_{n=1} + (n\beta'\nu + \epsilon_{n,0})\mathbb{1}_{2 \leq n \leq c-1} + (c\beta'\nu + \epsilon_{n,0})\mathbb{1}_{c \leq n \leq L}.$$

The customers timers are independent and identically distributed (i.i.d.) random variables and independent of the number of customers currently waiting.

(viii) The various stochastic processes within the system are assumed to be mutually independent.

4. Equilibrium probability analysis

We employ the Markov process approach, utilizing the Q-matrix, to establish the steady-state distribution for our proposed queueing model. Our primary focus lies in deriving the steady-state probabilities of the system, specifically as a function of the probability $\pi_{1,1}$, rather than relying on $\pi_{0,j}$ or $\pi_{L,j}$ for $j = 0, 1, 2$.

The system under consideration can be modeled as a continuous-time Markov process, denoted by $\{\mathfrak{X}(t), \mathfrak{Y}(t); t \geq 0\}$, where $\mathfrak{X}(t)$ represents the number of customers present in the system at time t, and $\mathfrak{Y}(t)$ characterizes the operational state of the servers at time t. The possible states for $\mathfrak{Y}(t)$ are as follows:

$$\mathfrak{Y}(t) = \begin{cases} 0, & \text{Servers are in a WVP} \\ 1, & \text{Servers are in a RBP} \\ 2, & \text{Servers are in a RP} \end{cases}$$

Let $\pi_{n,j}$ denote the steady-state probability that the system has n customers and the servers are in state j, such that: $\pi_{n,j} = \lim_{t \to \infty} P\{\mathfrak{X}(t) = n, \mathfrak{Y}(t) = j\}$, where $(n, j) \in \{\{(n, 0) : n = 0, 1, ..., L\} \cup \{(n, 1) : n = 1, 2, ..., L\} \cup \{(n, 2) : n = 1, 2, ..., L\}\}$.The state transition rate diagram is depicted in Figure 1.

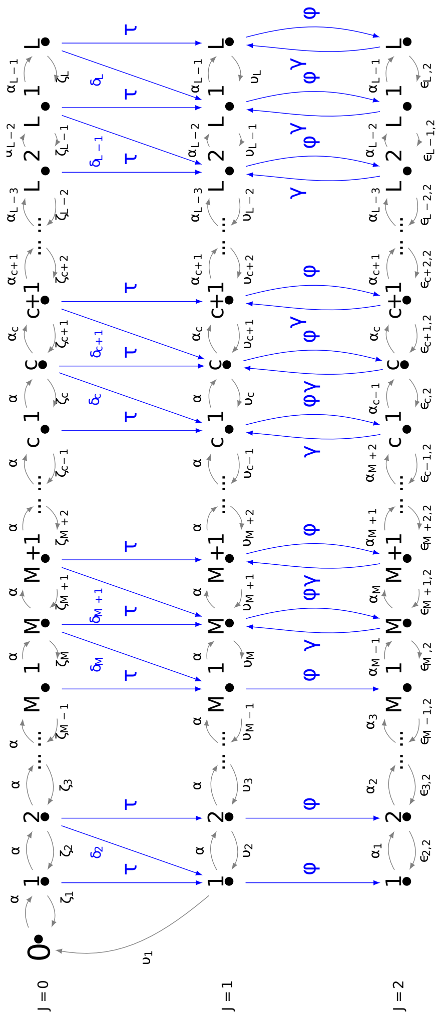

Figure 1: *State transition diagram for the proposed model*

4.1. Governing equations

The steady-state balance equations that govern our system are expressed as follows:

$$
\begin{aligned}
\alpha\pi_{0,0} &= (\mu+\kappa\varsigma_1)\pi_{1,1} + (\nu+\kappa\varsigma_0)\pi_{1,0}, \;\; n=0,\\
(\alpha + n(\nu+\kappa\varsigma_0)+\tau)\pi_{n,0} &= \alpha\pi_{n-1,0} + (n+1)(\beta'\nu+\kappa\varsigma_0)\pi_{n+1,0},\\
&\quad 1\le n\le c-1,\\
(\alpha\theta_n + c\nu + n\kappa\varsigma_0+\tau)\pi_{n,0} &= \alpha\pi_{n-1,0} + (c\beta'\nu + (n+1)\kappa\varsigma_0)\pi_{n+1,0},\\
&\quad n=c,\\
(\alpha\theta_n + c\nu + n\kappa\varsigma_0+\tau)\pi_{n,0} &= \alpha\theta_{n-1}\pi_{n-1,0} + (c\beta'\nu + (n+1)\kappa\varsigma_0)\pi_{n+1,0},\\
&\quad c+1\le n\le L-1,\\
(\tau + c\nu + L\kappa\varsigma_0)\pi_{L,0} &= \alpha\theta_{L-1}\pi_{n-1,0}, \;\; n=L,\\
(\alpha+\mu+\kappa\varsigma_1+\varphi)\pi_{1,1} &= \tau\pi_{1,0} + 2\beta\nu\pi_{2,0} + 2(\mu+\kappa\varsigma_1)\pi_{2,1}, \;\; n=1,
\end{aligned}
$$

$$
\begin{aligned}
(\alpha + n(\mu+\kappa\varsigma_1)+\varphi)\pi_{n,1} &= \alpha\pi_{n-1,1} + (n+1)\beta\nu\pi_{n+1,0} + (n+1)(\mu+\kappa\varsigma_1)\pi_{n+1,1}\\
&\quad +\tau\pi_{n,0}, \;\; 2\le n\le M-1,\\
(\alpha + n(\mu+\kappa\varsigma_1)+\varphi)\pi_{n,1} &= \alpha\pi_{n-1,1} + (n+1)\beta\nu\pi_{n+1,0} + (n+1)(\mu+\kappa\varsigma_1)\pi_{n+1,1}\\
&\quad +\tau\pi_{n,0} + \gamma\pi_{n,2}, \;\; M\le n\le c-1,\\
(\theta_n\alpha + n(\mu+\kappa\varsigma_1)+\varphi)\pi_{n,1} &= \alpha\pi_{n-1,1} + c\beta\nu\pi_{n+1,0} + (c\mu + (n+1)\kappa\varsigma_1)\pi_{n+1,1}\\
&\quad +\tau\pi_{n,0} + \gamma\pi_{n,2}, \;\; n=c,\\
(\alpha\theta_n + n(\mu+\kappa\varsigma_1)+\varphi)\pi_{n,1} &= \alpha\theta_{n-1}\pi_{n-1,1} + c\beta\nu\pi_{n+1,0} + (c\mu + (n+1)\kappa\varsigma_1)\pi_{n+1,1}\\
&\quad +\tau\pi_{n,0} + \gamma\pi_{n,2}, \;\; c+1\le n\le L-1,\\
(c\mu + L\kappa\varsigma_1+\varphi)\pi_{L,1} &= \alpha\theta_{L-1}\pi_{L-1,1} + \tau\pi_{L,0} + \gamma\pi_{L,2}, \;\; n=L,\\
\theta_1\alpha\pi_{1,2} &= \varphi\pi_{1,1} + 2\kappa\varsigma_2\pi_{2,2}, \;\; n=1,\\
(\alpha\theta_n + n\kappa\varsigma_2)\pi_{n,2} &= \alpha\theta_{n-1}\pi_{n-1,2} + (n+1)\kappa\varsigma_2\pi_{n+1,2} + \varphi\pi_{n,1},\\
&\quad 2\le n\le M-1,\\
(\alpha\theta_n + n\kappa\varsigma_2+\gamma)\pi_{n,2} &= \alpha\theta_{n-1}\pi_{n-1,2} + (n+1)\kappa\varsigma_2\pi_{n+1,2} + \varphi\pi_{n,1},\\
&\quad M\le n\le L-1,\\
(L\kappa\varsigma_2+\gamma)\pi_{L,2} &= \alpha\theta_{L-1}\pi_{L-1,2} + \varphi\pi_{L,1}, \;\; n=L.
\end{aligned}
$$

The normalizing condition is expressed as:

$$
\sum_{n=0}^{L}\pi_{n,0} + \sum_{n=1}^{L}\pi_{n,1} + \sum_{n=1}^{L}\pi_{n,2} = 1. \tag{1}
$$

Let's introduce the necessary notations for the subsequent sections of the paper:

$$
\dot{\zeta}_n = \begin{cases} n\nu + \epsilon_{n,0}, & 1\le n\le c-1,\\ c\nu+\epsilon_{n,0}, & c\le n\le L,\end{cases}
$$

$$
\varrho_n = \begin{cases} -(\alpha + \upsilon_n + \varphi), & 1\le n\le c-1,\\ -(\alpha_c + \upsilon_n+\varphi), & n=c,\\ -(\alpha_n+\upsilon_n+\varphi), & c+1\le n\le L-1,\\ -(\upsilon_L+\varphi), & n=L,\end{cases}
$$

$$
\vartheta_n = \begin{cases} -(\alpha + \dot{\zeta}_n + \tau), & 1\le n\le c-1,\\ -(\alpha_c + \dot{\zeta}_n+\tau), & n=c,\\ -(\alpha_n+\dot{\zeta}_n+\tau), & c+1\le n\le L-1,\\ -(\dot{\zeta}_L+\tau), & n=L,\end{cases}
$$

$$\sigma_n = \begin{cases} -\alpha_1, & n=1, \\ -(\alpha_n+\epsilon_{n,2}), & 2\leq n\leq M-1, \\ -(\alpha_n+\epsilon_{n,2}+\gamma), & M\leq n\leq L-1, \\ -(\epsilon_{L,2}+\gamma), & n=L. \end{cases}$$

4.2. Matrix solution

To obtain the steady-state solution, let the steady-state probability vector of the infinitesimal generator Q be denoted as $P=(\pi_0,\pi_1,\pi_2)$, where $\pi_0=(\pi_{0,0},\pi_{1,0},...,\pi_{L,0})$, $\pi_1=(\pi_{1,1},\pi_{2,1},...,\pi_{L,1})$, and $\pi_2=(\pi_{1,2},\pi_{2,2},...,\pi_{L,2})$. The steady-state equations $PQ=0$ must be satisfied by P along with the normalization condition $Pe=1$, where 0 is a zero row vector, and $e=(e_0,e_1,e_2)$ is a $(3L+1)$ column vector of ones, with e_0 being an $(L+1)$ column vector and e_1 and e_2 being L column vectors. The block structure of the infinitesimal generator Q is as follows:

$$Q=\begin{pmatrix} \mathcal{A}_1 & \mathcal{A}_2 & \mathcal{A}_3 \\ \mathcal{E}_1 & \mathcal{E}_2 & \mathcal{E}_3 \\ \mathcal{D}_1 & \mathcal{D}_2 & \mathcal{D}_3 \end{pmatrix},$$

where

$$\mathcal{A}_1=\begin{pmatrix} -\alpha & \alpha & & & & & & & \\ \zeta_1 & \vartheta_1 & \alpha & & & & & & \\ & \zeta_2 & \vartheta_2 & \alpha & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & \zeta_{c-1} & \vartheta_{c-1} & \alpha & & & \\ & & & & \zeta_c & \vartheta_c & \alpha_c & & \\ & & & & & \ddots & \ddots & \ddots & \\ & & & & & & \zeta_{L-1} & \vartheta_{L-1} & \alpha_{L-1} \\ & & & & & & 0 & \zeta_L & \vartheta_L \end{pmatrix},$$

$$\mathcal{A}_2=\begin{pmatrix} 0 & 0 & ... & ... & 0 \\ \tau & 0 & ... & ... & 0 \\ 0 & \tau & ... & ... & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & ... & ... & \tau \end{pmatrix},$$

$$\mathcal{E}_2=\begin{pmatrix} \varrho_1 & \alpha & & & & & & & \\ \upsilon_2 & \varrho_2 & \alpha & & & & & & \\ & \upsilon_3 & \varrho_3 & \alpha & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & \upsilon_{c-1} & \varrho_{c-1} & \alpha & & & \\ & & & & \upsilon_c & \varrho_c & \alpha_c & & \\ & & & & & \ddots & \ddots & \ddots & \\ & & & & & & \upsilon_{L-1} & \varrho_{L-1} & \alpha_{L-1} \\ & & & & & & 0 & \upsilon_L & \varrho_L \end{pmatrix},$$

$$\mathcal{D}_3=\begin{pmatrix} \sigma_1 & \alpha_1 & & & & \\ \epsilon_{2,2} & \sigma_2 & \alpha_2 & & & \\ & \epsilon_{3,2} & \sigma_3 & \alpha_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \epsilon_{L-1,2} & \sigma_{L-1} & \alpha_{L-1} \\ & & & 0 & \epsilon_{L,2} & \sigma_L \end{pmatrix},$$

$$\mathcal{E}_1 = \begin{bmatrix} v_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}, \ \mathcal{E}_3 = \begin{bmatrix} \varphi & & \\ & \ddots & \\ & & \varphi \end{bmatrix}, \ \text{and} \ \mathcal{D}_2 = \begin{bmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \gamma & & \\ & & & & \ddots & \\ & & & & & \gamma \end{bmatrix}.$$

Note that $\mathcal{A}_1$ and $\mathcal{A}_2$ are matrices of dimensions $(L+1) \times (L+1)$ and $(L+1) \times L$, respectively. $\mathcal{A}_3$ is a zero matrix with dimensions $(L+1) \times L$. $\mathcal{E}_1$ is an $L \times (L+1)$ matrix. $\mathcal{E}_2$, $\mathcal{E}_3$, $\mathcal{D}_2$, and $\mathcal{D}_3$ are square matrices of order $L \times L$. Additionally, $\mathcal{D}_1$ is a zero matrix with dimensions $L \times (L+1)$.

4.3. System state probabilities

We present the steady state equations $PQ = 0$ and the normalization condition $Pe = 1$, as

$$\begin{cases} \pi_0 \mathcal{A}_1 + \pi_1 \mathcal{E}_1 + \pi_2 \mathcal{D}_1 = 0, \\ \pi_0 \mathcal{A}_2 + \pi_1 \mathcal{E}_2 + \pi_2 \mathcal{D}_2 = 0, \\ \pi_0 \mathcal{A}_3 + \pi_1 \mathcal{E}_3 + \pi_2 \mathcal{D}_3 = 0, \\ \pi_0 e_1 + \pi_1 e_2 + \pi_2 e_3 = 1. \end{cases} \tag{2}$$

As $\mathcal{A}_3$ and $\mathcal{D}_1$ are null matrices, then Equation (2) can be rewritten as

$$\pi_0 \mathcal{A}_1 + \pi_1 \mathcal{E}_1 = 0, \tag{3}$$
$$\pi_0 \mathcal{A}_2 + \pi_1 \mathcal{E}_2 + \pi_2 \mathcal{D}_2 = 0, \tag{4}$$
$$\pi_1 \mathcal{E}_3 + \pi_2 \mathcal{D}_3 = 0, \tag{5}$$
$$\pi_0 e_1 + \pi_1 e_2 + \pi_2 e_3 = 1. \tag{6}$$

Next, put $\mathcal{A}_2 = \begin{pmatrix} O_1 \\ \tau I_L \end{pmatrix}$, $\mathcal{E}_1 = \begin{pmatrix} v_1 & O_1 \\ O_2 & O_3 \end{pmatrix}$, $\mathcal{E}_3 = (\varphi I_n)$, $\mathcal{D}_2 = \begin{pmatrix} O_4 & \\ \dot{O}_4 & \gamma I_{L-M+1} \end{pmatrix}$, with I_L denotes the identity matrix. Further, O_1 is a $1 \times L$ matrix. O_2 and O_3 are both of order $(L-1) \times 1$ and $(L-1) \times L$, respectively. O_4 has dimensions $(M-1) \times L$. $\dot{O}_4$ is of order $(L-M+1) \times (M-1)$. I_{L-M+1} represents the identity matrix of order $L-M+1$.

Let $\mathcal{A}1^{-1}$ and $\mathcal{D}3^{-1}$ denote the inverse matrices of $\mathcal{A}1$ and $\mathcal{D}3$, respectively. By referring to Eq. (3), we obtain the following result:

$$\begin{aligned} \pi_0 &= -\pi_1 \mathcal{E}_1 \mathcal{A}_1^{-1} \\ &= -\pi_1 \begin{pmatrix} v_1 o \\ O_5 \end{pmatrix} \\ &= -\pi_{1,1} v_1 o, \end{aligned} \tag{7}$$

where $o = (o_0, \tilde{o})$, such that $\tilde{o} = (o_1, ..., o_L)$ be an L row vector of the matrix $\mathcal{A}_1^{-1}$, and O_5 is $(L-1) \times (L+1)$. From Eq. (5), we have

$$\pi_2 = -\pi_1 \mathcal{E}_3 \mathcal{D}_3^{-1} = -\pi_1 \varphi \mathcal{D}_3^{-1}. \tag{8}$$

Substituting Eqs. (7) and (8) into Eq. (4), obtain

$$-\pi_{1,1} v_1 \tilde{o} \tau + \pi_1 (\mathcal{E}_2 - \varphi \mathcal{D}_3^{-1} \mathcal{D}_2) = 0. \tag{9}$$

As $\mathcal{E}_2$ and $\mathcal{D}_3^{-1}$ are both square matrices of order L, we can affirm the existence of the matrix:

$$\tilde{\mathcal{E}} = (\mathcal{E}_2 - \varphi \mathcal{D}_3^{-1} \mathcal{D}_2)^{-1}.$$

Thus

$$\pi_1 = (v_1 \tilde{o} \tau \tilde{\mathcal{E}}) \pi_{1,1}. \tag{10}$$

Consequently, we can deduce easily

$$\pi_2 = -(v_1 \tilde{o} \tau \tilde{\mathcal{E}} \varphi \mathcal{D}_3^{-1}) \pi_{1,1}. \tag{11}$$

Then, using Eqs. (7)–(11), we get:

$$\begin{cases} \pi_{n,0} = -v_1 o_n \pi_{1,1}, \\ \pi_{n,1} = (v_1 \tau \sum_{i=1}^{L} o_{i+1} \tilde{\psi}_{in}) \pi_{1,1}, \\ \pi_{n,2} = -\left(v_1 \tau \varphi \sum_{j=1}^{L} \sum_{i=1}^{L} o_{i+1} \tilde{\psi}_{ij} \tilde{\omega}_{jn} \right) \pi_{1,1}, \end{cases}$$

where $\tilde{\omega}_{ij}$ are the elements of matrix $\tilde{\mathcal{D}} = \mathcal{D}_3^{-1}$, and $\tilde{\psi}_{ij}$ are the elements of the matrix $\tilde{\mathcal{E}} = (\mathcal{E}_2 - \varphi \mathcal{D}_3^{-1} \mathcal{D}_2)^{-1}$. Finally, to determine $\pi_{1,1}$, we apply the normalizing condition (as described in Equation (1)):

$$\pi_{1,1} = \left(-v_1 \sum_{n=0}^{L} o_n + v_1 \tau \sum_{n=1}^{L} \sum_{i=1}^{L} o_{i+1} \tilde{\psi}_{in} - v_1 \tau \varphi \sum_{n=1}^{L} \sum_{j=1}^{L} \sum_{i=1}^{L} o_{i+1} \tilde{\psi}_{ij} \tilde{\omega}_{jn} \right)^{-1}.$$

5. Performance measures

In this section, we delve into the derivation of crucial system indices, leveraging the probabilities associated with the system distribution.

Result 1: The servers are in busy period with probability

$$P_{busy} = \sum_{n=1}^{L} \pi_{n,1} = v_1 \tau \sum_{n=1}^{L} \sum_{i=1}^{L} o_{i+1} \tilde{\psi}_{in} \pi_{1,1}. \tag{12}$$

Result 2: The servers are in working vacation period with probability

$$P_{wv} = \sum_{n=1}^{L} \pi_{n,0} = -\left(v_1 \sum_{n=0}^{L} o_n \right) \pi_{1,1}. \tag{13}$$

Result 3: The servers are in breakdown period with probability

$$P_{bp} = \sum_{n=1}^{L} \pi_{n,2} = -v_1 \tau \varphi \left(\sum_{n=1}^{L} \sum_{j=1}^{L} \sum_{i=1}^{L} o_{i+1} \tilde{\psi}_{ij} \tilde{\omega}_{jn} \right) \pi_{1,1}. \tag{14}$$

Result 3: The probability of system reliability

$$P_{re} = 1 - \pi_{pb}.$$

Result 4: The mean system size is

$$\begin{aligned} E_s &= \sum_{n=1}^{L} n(\pi_{n,0} + \pi_{n,1} + \pi_{n,2}) \qquad (15) \\ &= v_1 \left(-\sum_{n=1}^{L} n o_n + \tau \sum_{n=1}^{L} \sum_{i=1}^{L} n o_{i+1} \tilde{\psi}_{in} - \tau \varphi \sum_{n=1}^{L} \sum_{j=1}^{L} \sum_{i=1}^{L} n o_{i+1} \tilde{\psi}_{ij} \tilde{\omega}_{jn} \right) \pi_{1,1}. \end{aligned}$$

Result 5: The effective arrival rate

$$\alpha' = \alpha \pi_{0,0} + \sum_{n=1}^{L} \alpha_n \pi_{n,0} + \sum_{n=1}^{L} \alpha_n \pi_{n,1} + \sum_{n=1}^{L} \alpha_n \pi_{n,2}.$$

Result 6: The mean waiting time of customers in the system

$$W_s = E_s/\alpha^{'}.$$

Result 7: The average balking rate

$$R_{balk} = \alpha - \alpha^{'}. \tag{16}$$

Result 8: The average reneging rate

$$\begin{aligned} R_{ren} &= \kappa\varsigma_0 \sum_{n=1}^{L} n\pi_{n,0} + \kappa\varsigma_1 \sum_{n=1}^{L} n\pi_{n,1} + \kappa\varsigma_2 \sum_{n=2}^{L} n\pi_{n,2} \\ &= v_1\kappa \left(-\varsigma_0 \sum_{n=1}^{L} no_n + \tau\varsigma_1 \sum_{n=1}^{L}\sum_{i=1}^{L} no_{i+1}\tilde{\psi}_{in} - \tau\varphi\varsigma_2 \sum_{n=2}^{L}\sum_{j=1}^{L}\sum_{i=1}^{L} no_{i+1}\tilde{\psi}_{ij}\tilde{\omega}_{jn} \right) \pi_{1,1}. \end{aligned} \tag{17}$$

Result 9: The average retention rate

$$\begin{aligned} R_{ret} &= \kappa^{'}\varsigma_0 \sum_{n=1}^{L} n\pi_{n,0} + \kappa^{'}\varsigma_1 \sum_{n=1}^{L} n\pi_{n,1} + \kappa^{'}\varsigma_2 \sum_{n=2}^{L} n\pi_{n,2} \\ &= v_1\kappa^{'} \left(-\varsigma_0 \sum_{n=1}^{L} no_n + \tau\varsigma_1 \sum_{n=1}^{L}\sum_{i=1}^{L} no_{i+1}\tilde{\psi}_{in} - \tau\varphi\varsigma_2 \sum_{n=2}^{L}\sum_{j=1}^{L}\sum_{i=1}^{L} no_{i+1}\tilde{\psi}_{ij}\tilde{\omega}_{jn} \right) \pi_{1,1}. \end{aligned} \tag{18}$$

Result 10: The mean number of customers served per unit time

$$C_s = \nu \sum_{n=1}^{L} n\pi_{n,0} + \mu \sum_{n=1}^{L} n\pi_{n,1}.$$

6. Cost model and optimization

For our queueing model, we consider he cost components as outlined below:

1. C_{busy}: unit time cost for system being in busy period.
2. C_{wv}: unit time cost for system being is in working vacation.
3. C_{break}: unit time cost for system being in breakdown period.
4. C_{sq}: Holding unit time cost when a customer enters the queue.
5. C_{s_1}: Cost per service per unit time in regular working period.
6. C_{s_2}: Cost per service per unit time in working vacation period.
7. C_l: unit time cost when a customer is lost.
8. C_t: unit time cost when the system retains a customer.
9. C_f: Fixed purchase cost of the server per unit.

The formulation of the cost per unit time function for the queueing system is as follows:

$$\begin{aligned} \mathcal{T}_c &= C_{busy}P_{busy} + C_{wv}P_{wv} + C_{break}P_{pb} + C_{sq}E_s + C_l(R_{ren} + R_{balk}) \\ &\quad + C_tR_{ret} + c(\mu C_{s_1} + \nu C_{s_2}) + cC_f. \end{aligned} \tag{19}$$

Expressing the expected cost function $\mathcal{T}_c$ explicitly by substituting Equations (12)-(18) into (19) would result in an extremely complex formulation. Consequently, studying the analytical behavior of $\mathcal{T}_c$ becomes a big challenge. Furthermore, due to the nonlinearity and intricacy of the expected cost function, deriving the optimal solution (c^*, M^*, μ^*, ν^*) in closed form would be an arduous task.

To circumvent these difficulties and perform the optimization analysis, we employ direct search and Newton's methods as numerical optimization techniques to search for the optimal solution (c^*, M^*, μ^*, ν^*). Initially, the direct search method is utilized to determine the optimal values of the variables (c^*, M^*). Subsequently, with these variables fixed, Newton's method is applied to find the optimal values of the variables (μ^*, ν^*).

6.1. Numerical cost optimum parameter

We consider a practical problem concerning the automated teller machine (ATM) production facility mentioned in Section 2. In the context of the considered practical example, the system parameters are delineated as follows: failed machines arrive according to a Poisson process with $\alpha = 7$. The system capacity is considered finite with $L = 20$. If the system is in operation, the failure occurs in which the breakdown times are exponential distribution with $\varphi = 0.1$. The service and repair times of the machines obey exponential distributions with parameters $\mu = 3.0$, $\nu = 0.9$, and $\gamma = 0.3$, respectively. Once the system gets empty, it goes on vacation period, the vacation period follows exceptional distribution with parameter $\tau = 0.4$. The failed machines during both period may get impatient and leave the system with being served. The impatience timers follow exponential distribution with $\varsigma_0 = 0.5$, $\varsigma_1 = 0.3$, $\varsigma_2 = 0.9$. Further, during working vacation period, the failed machines service may be continue their service during working vacation period with probability $\beta' = 0.6$, and they can leave the system with probability $\kappa = 0.7$. The joining probability is taken as $\theta_n = 1 - \frac{n}{L}$.

An efficient algorithm based on the direct search method is employed to determine the optimal discrete values (c^*, M^*) that optimize the expected cost function. The effectiveness of this approach hinges on the convexity (or unimodality) of the cost function. Throughout the numerical analysis, the following cost elements are considered: $C_{busy} = \$20$, $C_{wv} = \$20$, $C_{bp} = \$50$, $C_{sq} = \$10$, $C_{s1} = \$5$, $C_{s2} = \$5$, $C_l = \$30$, $C_t = \$25$, $C_f = \$1$ and $R = 50$.

Figure 2 illustrates the behavior of the expected cost function $\mathcal{T}_c(c^*, M^*)$ for varying values of c and M. The plotted curve exhibits a convex shape, indicating the existence of a single relative minimum. Consulting Table 1, it is evident that the minimum expected cost per unit time, which amounts to **182.5710**, is attained when $c^* = 6$ and $M^* = 1$.

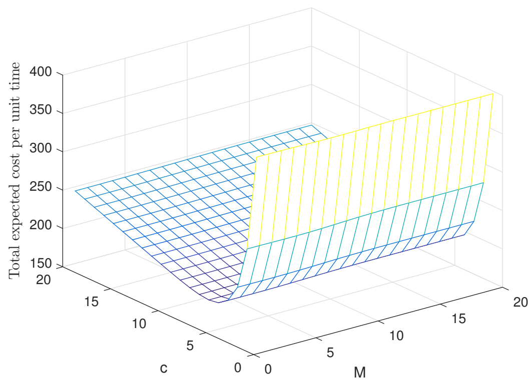

Figure 2: *The expected cost $\mathcal{T}_c$ for different values of c and M.*

Table 1: *c and M vs.* $\mathcal{T}_c(c, M)$

c / M	1	2	3	4	5	6	7	8
2	271.9823	-	-	-	-	-	-	-
3	216.2269	216.2388	-	-	-	-	-	-
4	193.3445	193.3762	193.3707	-	-	-	-	-
5	184.5919	184.6346	184.6462	184.7100	-	-	-	-
6	**182.5710**	182.6202	182.6410	182.7123	182.8470	-	-	-
7	183.9711	184.0241	184.0501	184.1262	184.2599	184.4512	-	-
8	187.1491	187.2043	187.2331	187.3123	187.4466	187.6304	187.8610	-
9	191.2406	191.2970	191.3274	191.4083	191.5437	191.7250	191.9432	192.1971

Once the optimal values (c^*, M^*) are determined, Newton's method is employed to locate the minimum value of $\mathcal{T}_c(c^*, M^*, \mu^*, \nu^*)$ by iteratively optimizing the continuous variables μ and ν. Newton's method is an efficient iterative technique for finding the optimum of a nonlinear function by computing the search direction at each iteration.

The Quasi-Newton method, a variant of Newton's method, is utilized to numerically determine μ^* and ν^*. Numerical results obtained through this optimization process are presented in Tables 2-7 for various system parameter settings.

Table 2: $\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$ *and* (M^*, c^*, μ^*, ν^*) *while adjusting* τ *and* α *(*$\varsigma_0 = 0.5$, $\varsigma_1 = 0.3$, $\varsigma_2 = 0.9$, $L = 20$, $\kappa = 0.7$, $\beta' = 0.6$, $\varphi = 0.1$, $\gamma = 0.3$*)*

(τ , α)	(2,5)	(2,8)	(2.5,5)	(2.5,8)	(3.0,5)	(3.0,8)
(M^* ,c^*)	(1,2)	(4,7)	(1,2)	(4,7)	(1,2)	(4,7)
μ^*	3.3953	1.4294	3.5066	1.4697	3.5893	1.4991
ν^*	1.7352	0.4679	1.3760	0.3242	1.0157	0.1853
$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$	125.1117	182.8660	124.4980	180.6144	123.3919	177.7679

Table 3: $\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$ *and* (M^*, c^*, μ^*, ν^*) *while adjusting* β' *(*$\varsigma_0 = 0.5$, $\varsigma_1 = 0.3$, $\varsigma_2 = 0.9$,$L = 20$, $\kappa = 0.7$, $\alpha = 7$, $\varphi = 0.1$, $\tau = 0.4$, $\gamma = 0.3$*)*

β'	0.75	0.8	0.85	0.9	0.95
(M^* , c^*)	(3,4)	(4,5)	(4,6)	(5,7)	(6,8)
μ^*	1.6362	1.3007	1.0810	0.9233	0.8065
ν^*	1.5967	1.2149	0.9672	0.7970	0.6748
$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$	147.4804	150.3079	153.4443	156.8410	160.5053

Table 4: $\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$ *and* (M^*, c^*, μ^*, ν^*) *while adjusting* L *and* κ *(*$\varsigma_0 = 0.5$, $\varsigma_1 = 0.3$, $\varsigma_2 = 0.9$, $\alpha = 7$, $\tau = 0.4$, $\beta' = 0.6$, $\varphi = 0.1$ $\gamma = 0.3$*)*

(L,κ)	(10,0.6)	(40,0.6)	(10,0.7)	((40,0.7)	(10,0.9)	(40,0.9)
(c^* ,M^*)	(7,1)	(7,4)	(7,1)	(7,4)	(7,1)	(7,4)
μ^*	0.9190	1.0397	0.8876	0.9821	0.8333	0.8757
ν^*	0.9180	0.8733	0.8866	0.8594	0.8323	0.8333
$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$	167.7709	167.6681	165.5115	165.0619	161.7871	160.6461

Table 5: *$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$ and (M^*, c^*, μ^*, ν^*) while adjusting φ and γ ($\varsigma_0 = 0.5$, $\varsigma_1 = 0.3$, $\varsigma_2 = 0.9$, $\kappa = 0.7$, $\beta' = 0.6$, $\tau = 0.4$, $\alpha = 7$, $L = 20$)*

$(\varphi\ ,\ \gamma)$	(0.4 1)	(0.4 5)	(0.6 1)	(0.6 5)	(0.8 1)	(0.8 5)
(7,3)						
μ^*	0.9571	0.9275	0.9299	0.8955	0.9691	0.9342
ν^*	0.8955	0.8907	0.9178	0.8945	0.9075	0.9030
$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$	170.3847	162.9654	172.9348	163.4767	177.8815	165.2887

Table 6: *$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$ and (M^*, c^*, μ^*, ν^*) while adjusting ς_0 , ς_1 and $\varsigma_2 = 1.9$ ($\kappa = 0.7$, $L = 20$, $\alpha = 7$, $\tau = 0.4$, $\varphi = 0.1$, $\beta' = 0.6$, $\gamma = 0.3$)*

$(\varsigma_1\ \varsigma_0)$	(0.2,0.6)	(0.4,0.6)	(0.2,0.8)	(0.4,0.8)	(0.2,1.0)	(0.4,1.0)
$(c^*\ ,M^*)$	(8,2)	(8,4)	(8,2)	(8,3)	(8,5)	(8,7)
μ^*	0.8287	0.7616	0.8079	0.7640	0.8041	0.7644
ν^*	0.7697	0.7606	0.8056	0.7630	0.8031	0.7634
$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$	171.4859	172.4115	179.1606	180.3308	186.5172	188.2080

Table 7: *$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$ and (M^*, c^*, μ^*, ν^*) while adjusting ς_0 , ς_1 and $\varsigma_2 = 3.1$ ($\kappa = 0.7$, $L = 20$, $\alpha = 7$, $\tau = 0.4$, $\varphi = 0.1$, $\beta' = 0.6$, $\gamma = 0.3$)*

$(\varsigma_1\ \varsigma_0)$	(0.2,0.6)	(0.4,0.6)	(0.2,0.8)	(0.4,0.8)	(0.2,1.0)	(0.4,1.0)
$(c^*\ ,M^*)$	(8,2)	(8,4)	(8,2)	(8,3)	(8,5)	(8,7)
μ^*	0.8211	0.7583	0.8032	0.7607	0.8006	0.7621
ν^*	0.7712	0.7573	0.8022	0.7597	0.7996	0.7611
$\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$	170.7178	171.7406	178.4237	179.6992	185.8408	187.5095

Tables 2-7 illustrate the relationships between various system parameters and the optimal service rates (μ^*, ν^*) that minimize the expected cost $\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$:

- As the arrival rate of failed machines (α) increases, the expected cost $\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$ rises substantially. Similarity for β'. This is understandable, as a higher influx of failures naturally strains the service system, leading to longer queues, more congestion, and ultimately increased costs.
- Conversely, higher operator vacation rate (τ) and greater customer non-retention probability (κ) decrease the expected cost. Obviously, more frequent vacations provide more opportunities to serve customers during vacation periods, alleviating congestion. Likewise, allowing more customers to renege without service reduces the queue length and wait times.
- The positive effect of operator breakdown rate (φ) on expected cost is expected, since more breakdowns directly degrade service capability and capacity. In contrast and as anticipated, faster operator repair rate (γ) significantly improves system performance and reduces costs by restoring capacity quicker after failures.
- Larger system capacity (L) and impatience rates ($\varsigma_j, j = 0, 1, 2$) increase service abandonments, lowering congestion and $\mathcal{T}_c(M^*, c^*, \mu^*, \nu^*)$. However, excessive abandonments negatively impact customer service. An optimal balance is required.

7. Conclusion

In this paper, based on the characteristics of the repair machine, we presented a $M/M/c/L$ queue with breakdowns, repairs, threshold-based recovery policy, working vacation, Bernoulli interruption, balking, reneging, and retention. We established the steady-state solution of the system using Q-matrix. Then, we studied important system characteristics based on the steady-state probabilities. Finally, we presented the sensitivity and cost optimization analysis; we discussed an economic analysis as well as the optimal threshold, the optimal number of servers as well as the service rates μ and ν under a given cost assumption because determining these parameters to achieve the minimum cost is very important in queueing theory. As further potential future study, we can generalize this queueing model with to some different cases, as follows:

- (i) Considering the feedback phenomenon within the queueing systems, it is pertinent to examine the scenario involving feedback customers in the proposed queueing model.
- (ii) It will be interesting to incorporate retrial policy and preemptive resume priority, this makes the system closer to real-life congestion scenarios and the study can provide potentially practical application in flexible manufacturing systems, transportation system, telecommunication systems, and so on.
- (iii) One could also extend the present study by considering multi-optional services.

Declarations

Ethics Approval Not applicable.
Consent to Participate Not applicable.
Consent for Publication Not applicable.
Conflict of Interest The authors declare no competing interests.
Author Contribution The authors have contributed equally to the manuscript.
Funding Not applicable.
Availability of Data and Materials Not applicable.

References

[1] Bouchentouf, A. A., Boualem , M. , Yahiaoui, L. and Ahmad, H. (2022). A multi-station unreliable machine model with working vacation policy and customers' impatience. *Quality Technology & Quantitative Management*, 1-31. doi.org/10.1080/16843703.2022.2054088.
[2] Bouchentouf, A. A., Cherfaoui, M. and Boualem, M. (2019). Performance and economic analysis of a single server feedback queueing model with vacation and impatient customers. *Opsearch*, 56(1):300–323.
[3] Bouchentouf, A. A., Cherfaoui, M. and Boualem, M. (2021). Analysis and performance evaluation of Markovian feedback multi-server queueing model with vacation and impatience. *American Journal of Mathematical and Management Sciences*, 40(3):261–282.
[4] Chang, F. M., Lee , Y. T. , Chang, C. J. and Yeh, C. (2018). Analysis of a standby redundant system with controlled arrival of failed machines. *International Journal of Industrial and Systems Engineering*, 28(1):117–134.
[5] Chen, W. L. and Wang, K. H. (2018). Reliability analysis of a retrial machine repair problem with warm standbys and a single server with N-policy. *Reliab. Eng. Syst. Saf*, 180:476–486.
[6] B. T. Doshi. (1986). Queueing systems with vacations a survey. *Queueing Systems*, 1(1):29–66.
[7] Doshi , B. T. :(1990). Single sever queues with vacations. *In Stochastic analysis of computer and communication systems, ed. H. Takagi, Amsterdam: North Holland*, 217–65.
[8] Gao, S. and Liu, Z. (2013). An $M/G/1$ queue with single working vacation and vacation interruption under Bernoulli schedule. *Appl. Math. Model*, 37:1564–1579.

[9] V. Goswami.(2016). Relationship between randomized F-policy and randomized N-policy in discrete-time queues. *Opsearch*, 53(1):131–150.

[10] Jain, M. (2017). Priority queue with batch arrival, balking, threshold recovery, unreliable server and optimal service. *Rairo-Oper. Res*, 51:417–432. DOI: 10.1051/ro/2016032 www.rairo-ro.org.

[11] Jayachitra, P. and Albert, A. J. (2014). Recent developments in queueing models under N-policy, A short survey. *Int. J. Math. Arch*, 5:227–233.

[12] Kumari, U. and Sharma, D. C. (2021). Performance analysis of a warm standby machine repair problem with servers vacation, impatient and controlling F-policy. *Mathematics in Engineering, science and aerospace*, 12(2):1–21.

[13] Lee, D. H. (2017). Optimal pricing strategies and customers' equilibrium behavior in an unobservable $M/M/1$ queueing system with negative customers and repair. *Mathematical Problems in Engineering*, 2017: Article ID 8910819–11 pages. https://doi.org/10.1155/2017/8910819

[14] Lv, S. and Li, J. (2013). The $M/M/N$ repairable queueing system with variable breakdown rates. *Discrete Dynamics in Nature and Society*, Article ID 173407: 10 pages. https://doi.org/10.1155/2013/173407.

[15] Ojobor, S. A. and Ogini, N. O. (2022). Threshold recovery policy for the machine interference repair problem with server vacations. *Journal of Physics: Conference Series*, 2199:012020. doi:10.1088/1742-6596/2199/1/012020.

[16] Servi, L. D. and Finn, S. G. (2002). $M/M/1$ queues with working vacations $(M/M/1/WV)$. *Perform. Eval*, 50:41–52.

[17] Shekhar, C. , Raina, A. A., Kumar, A. and Iqbal, J. (2017). A survey on queues in machining system progress from 2010 to 2017. *Yugoslav Journal of Operations Research*, Vol.27, No.4, pp. 391–413. https://doi.org/10.2298/YJOR161117006R

[18] Shoukry, E., Salwa, M. A. and Boshra, A. S. (2018). Matrix geometric method for $M/M/1$ queueing model with and without breakdown ATM machines. *American Journal of Engineering Research*, 7(1):246–252.

[19] Sun, W. , Li, S. and Guo, E. C. (2016). Equilibrium and optimal balking strategies of customers in Markovian queues with multiple vacations and N-policy. *Applied Mathematical Modelling*, 40:284–301.

[20] H. Takagi. (1991). Queueing analysis, Vacation and priority systems, Vol. 1. Amsterdam: North Holland .

[21] Tian, N. and Zhang, Z. G. (2006). Vacation queueing models, Theory and applications, New York: Springer .

[22] Yang, D. Y., Chiang, Y. C. and C. S. Tsou. (2013). Cost analysis of a finite capacity queue with server breakdowns and threshold-based recovery policy. *Journal of Manufacturing Systems*, 32:174-179.

[23] Yang, D. Y. , Chung, C. H. and Wu, C. H. (2021). Sojourn times in a Markovian queue with working breakdowns and delayed working vacations. *Comput. Ind. Eng*, 156:107–239.

[24] Ye, Q. and Liu, L. (2017). Performance analysis of the $GI/M/1$ queue with single working vacation and vacations. *Methods Comp. Appl. Prob*, 19:685–714.

[25] Yen, T. C., Wang, K. H. and Chen, J. Y.(2020). Optimization Analysis of the N Policy $M/G/1$ Queue with working breakdowns. *Symmetry*, 12: 583. doi:10.3390/sym12040583.

[26] Zhang, M. and Hou, Z. (2012). $M/G/1$ queue with single working vacation. *J. Appl. Math. Comput*, 39: 221–234.