Methodology and Computing in Applied Probability (2024) 26:21
https://doi.org/10.1007/s11009-024-10090-1

RESEARCH

Performance and Optimization Analysis of a Queue with Delayed Uninterrupted Multiple Vacation and *N*-Policy

Yaxing He[1] · Yinghui Tang[1] · Miaomiao Yu[1] · Wenqing Wu[2]

Received: 14 July 2022 / Revised: 3 February 2024 / Accepted: 14 May 2024 /
Published online: 13 June 2024

Abstract
This paper considers an $M/G/1$ queue with delayed uninterrupted multiple vacation and N-policy, in which (i) the server remains dormant from vacation to a non-empty system with no more than N customers, and (ii) when the system becomes exhausted, the server waits for a random time instead of immediately going on vacation (applicable to many scenarios and common to human behavior). We first study the transient queue length distribution from an arbitrary initial state and obtain its Laplace transform expressions. Then, the recursive formulas of the stationary queue length distribution are developed. Meanwhile, some crucial performance measures are presented. Finally, the cost optimization problems are discussed with and without the average waiting time constraint. Four numerical examples are illustrated to determine the optimal control policy for minimizing cost under the conditions that the random variables obey different phase-type (PH) distributions.

Keywords N-policy · Delayed period · Uninterrupted multiple vacation · Queue length distribution · Optimal control policy

Mathematics Subject Classification 60K25 · 90B22

1 Introduction

The purpose of studying the queueing system is to adjust and control the system so that it operates optimally. It is well known that in some practical systems, there is a significant setup cost for each busy period. Reducing setup frequency has an economic benefit. Thus, finding the cost-minimization control policy and vacation policy is a typical control problem in queueing theory. The N-policy (Yadin and Naor 1963), T-policy (Heyman 1977), and D-policy (Balachandran 1973) were the earliest contributions to the control strategies of queueing systems. In the last two decades, there has been an increasing interest in investigating

✉ Yinghui Tang
tangyh@sicnu.edu.cn

[1] School of Mathematical Sciences, Sichuan Normal University, Chengdu 610068, China

[2] School of Science, Southwest University of Science and Technology, Mianyang 621000, China

some related queueing systems with joint control policies or unreliable servers. Until now, an enormous number of works on threshold policies and vacation queues have been published. For example, Wang et al. (2006) and Sethi et al. (2020) studied the unreliable queues under N-policy control. Artalejo (2002) discussed the optimality of the N-policy and D-policy for the $M/G/1$ queue. Hur et al. (2003) studied an $M/G/1$ queue with N-policy and T-policy. Kuang et al. (2022) proposed an $M/G/1$ queue with bi-level randomized $(p, N1, N2)$-policy. Readers interested in joint control policies can refer to relevant literature (e.g., Lee and Seo et al. 2008; Lee et al. 2010; Wei et al. 2020).

Many scholars have studied the single and multiple vacation models as two essential generalizations of the classical $M/G/1$ queue. Some excellent studies on vacation models have been reported (e.g., Levy and Yechiali 1975; Doshi 1986; Tian and Zhang 2008). Utilizing the supplement variable technique, Lee et al. (1994a) considered an $M^x/G/1$ queueing model with N-policy and single vacation related to a manufacturing system. Kella (1989) first studied an $M/G/1$ queue by combining multiple vacation and N-policy. Lee et al. (1994b) further studied an $M^x/G/1$ queue N-policy and multiple vacation, in which multiple vacation mean that as soon as the system empties, the server leaves for a vacation of random length. If the server returns and finds the queue size is not less than the threshold N, the server immediately begins to serve the waiting customers. Otherwise, the server leaves for another vacation, and so on, until he finally finds at least N customers.

Also worth mentioning is that the above models with multiple vacation and N-policy are different from the models with multiple vacation and $\min(N, V)$-policy (e.g., Wu et al. 2014; Lan and Tang 2019; Luo et al. 2023), i.e., the above multiple vacation cannot be interrupted immediately after the threshold is reached, whereas vacations are interrupted in the queueing model with $\min(N, V)$ policy. In some real-world scenarios, the auxiliary work of the server cannot be interrupted immediately. Customers must wait for the server to complete the auxiliary work before restarting the service. In the past two decades, queueing models with N-policy and uninterrupted vacations have also received much attention. Baba (2004) analyzed a $GI/M/1$ queue with multiple working vacations. Ayyappan and Nirmala (2020) consider a repairable queue with N-policy and multiple vacation. More literature on uninterrupted vacations can be found in references like Agarwal and Dshalalow (2005), Li and Tian (2010), Luo et al. (2013) and Ayyappan and Karpagam (2019).

In the previously studied models with uninterrupted multiple vacation, the server takes another vacation if the number of waiting customers does not reach N. In contrast to these existing literature, this paper will propose a new uninterrupted multiple vacation mechanism, i.e., when less than N customers are waiting for service in the system, the server waits for the number of customers to reach N and then provides service immediately instead of taking another vacation. When the vacation time is too long, or the arrival rate is too high, this flexible vacation mechanism can reduce the waiting time for existing customers and improve customer satisfaction. Additionally, the server cannot go on vacation immediately after the system is empty. He/She has to experience a delay before going on vacation, similar to a bank employee needs to sort out accounts before leaving work. Readers interested in delayed vacations can refer to papers by Tang et al. (2008) and Luo et al. (2023).

The prime objective of our research is to achieve cost minimization by determining the control threshold and adjusting vacation time. This paper will provide a reference and theoretical basis for decision-makers to manage such systems. Using an analysis technique different from the traditional analysis methods (e.g., embedded Markov chain and supplementary variable technique), we perform transient and steady-state analysis on the queue size employing the renewal process theory, total probability decomposition technique and Laplace transform. Furthermore, to minimize customer waiting times and trade off costs between customers and

system managers, the cost optimization problems with and without the mean waiting time constraint are studied when the random variables obey different PH distributions.

The remainder of this paper is organized as follows. The next section describes the mathematical model under consideration and some notations adopted in this paper. Section 3 provides two lemmas and then presents the expressions of the Laplace transform of the transient queue size distribution concerning time t. Section 4 states some important queueing performance indices when the system is in equilibrium. Section 5 offers four numerical examples to determine the optimal control policy N^* and the optimal two-dimensional control policy (N^*, T^*) for minimizing the system cost. At last, Section 6 draws conclusions and puts forward some ideas for future work.

2 Model Description and Related Preparation

The queueing system with delayed uninterrupted multiple vacation and N-policy considered in this paper is described as follows:

1. The inter-arrival times $\tau_n, n \geq 1$ are independent identically distributed random variables each with distribution $F_\tau(t) = 1 - e^{-\lambda t}, t \geq 0, \lambda > 0$. The service times $\chi_n, n \geq 1$ are independent identically distributed random variables each with distribution $F_\chi(t), t \geq 0$, which is supposed to have a finite mean $E[\chi]$.
2. When the system becomes empty, the server experiences a delay period Y with an arbitrary distribution $F_Y(t)$ before taking a vacation. Once a customer arrives at the system within Y, the server starts serving the customer until the system becomes empty again and Y is restarted. If no customer arrives during Y, the server takes a vacation with a random length V that obeys an arbitrary distribution $F_V(t)$. After returning from vacation, the server will respond in three different ways depending on the state of the system: (i)If the number of customers in the system is greater than or equal to the control threshold value N (≥ 1) which is set before, the server starts its service immediately; (ii)If there are less than N customers but at least one customer in the system, the server stays idle until there are N customers and starts its service at once; (iii)If no customers are waiting to be served, the server immediately goes on another vacation.
3. The inter-arrival time τ, the service time χ, the delay period Y, and the vacation time V are mutually independent. In addition, if j (≥ 1) customers are waiting at the initial time $t = 0$, the service begins at once. If the system is empty at $t = 0$, the server keeps idle until the next customer arrives and provides service immediately.

Remark 1 In our model formulation, to minimize waiting times for customers who are sensitive to delays, we assume that if the number of customers is less than N but there is at least one customer in the system, the server will stay idle until N customers are present, at which point the service will commence immediately. The assumptions here are quite different from those in existing literature (e.g., Kella 1989; Lee et al. (1994a, b); Ayyappan and Karpagam 2019; Ayyappan and Nirmala 2020), which assumes that the server remains on vacation if there are fewer than N customers in the waiting line. The above assumption will undoubtedly increase the customer's waiting time because the server's vacation is still ongoing when N customers are accumulated in the system. At the same time, as we mentioned in the introduction, the multiple vacation model we present in this paper can reduce waiting time for existing customers. When the vacation time is too long or the arrival rate is too high, this vacation mechanism can increase customer satisfaction.

Remark 2 The queueing model considered in this paper extends previous models studied by other authors as shown below.

- When $P\{Y = \infty\} = 1$, the queueing model considered in this paper is equivalent to the classic $M/G/1$ queueing model that has been investigated (see e.g., Cohen 1982).
- When $P\{Y = 0\} = 1$ and $N = 1$, the queueing model considered in this paper can be simplified to the standard multiple vacation queueing model that has been studied (see e.g., Tian and Zhang 2008).

Remark 3 The notations used throughout this paper are listed as follow for reference.

$N(t)$: The number of customers at time t;
$f_\tau(s)$: Laplace-Stieltjes transform for $F_\tau(t)$, i.e., $f_\tau(s) = \int_0^\infty e^{-st} dF_\tau(t)$;
$f_\tau^*(s)$: Laplace transform for $F_\tau(t)$, i.e., $f_\tau^*(s) = \int_0^\infty e^{-st} F_\tau(t)dt$;
$F_\chi^{(k)}(t)$: k-fold convolution of $F_\chi(t)$, i.e., $F_\chi^{(k)}(t) = \int_0^t F_\chi^{(k-1)}(t-x)dF_\chi(x)$, $k \geq 1$, and $F_\chi^{(0)}(t) = 1$;
$F_\tau(t) * F_\chi(t)$: $F_\tau(t) * F_\chi(t) = \int_0^t F_\tau(t-x)dF_\chi(x) = \int_0^t F_\chi(t-x)dF_\tau(x)$;
$\bar{F}_\chi(t)$, $\bar{f}_\chi(s)$: $\bar{F}_\chi(t) = 1 - F_\chi(t)$ and $\bar{f}_\chi(s) = 1 - f_\chi(s)$;
$E[\chi]$, $E[\chi^2]$: First two moments of χ;
$\rho = \lambda E[\chi]$: Traffic intensity of the system;
$\Re(s)$: Real part of the complex variable s;
v_m: Probability of m customers arriving at the system during vacation time, i.e., $v_m = \int_0^\infty \frac{(\lambda t)^m}{m!} e^{-\lambda t} dF_V(t), m \geq 0$;
y_n: Probability of n customers arriving at the system during delay period, i.e., $y_n = \int_0^\infty \frac{(\lambda t)^n}{n!} e^{-\lambda t} dF_Y(t), n \geq 0$.

3 The Transient Queue Length Distribution and its Laplace Transform Expression

Firstly, the definition of the server's busy period and two lemmas are introduced as follows.

Server's busy period: It is the period from when the server begins to provide service until the end of all services. As a result, the server's busy period here is identical to the system busy period of the classic $M/G/1$ queueing system.

Let b denote the duration of the server's busy period which begins with just one customer, and $F_B(t) = P\{b \leq t\}, t \geq 0$, $f_B(s) = \int_0^\infty e^{-st} dF_B(t)$. Similar to the discussing in Cohen (1982), we have the following Lemma 1.

Lemma 1 *For $\Re(s) > 0$, $f_B(s)$ is the unique solution of the equation $z = f_\chi(s + \lambda(1-z)) = \int_0^\infty e^{-(s+\lambda(1-z))} dF_\chi(t)$ in $|z| < 1$, and $F_B(t)$ is given by*

$$F_B(t) = \sum_{k=1}^{\infty} \int_0^t \frac{(\lambda x)^{k-1}}{k!} e^{-\lambda x} dF_\chi^{(k)}(t)(x), t \geq 0,$$

and

$$\lim_{t\to\infty} F_B(t) = \lim_{s\to 0^+} f_B(s) = \begin{cases} 1, & \rho \leq 1, \\ \omega < 1, & \rho > 1, \end{cases} \quad E[b] = \begin{cases} \frac{\rho}{\lambda(1-\rho)}, & \rho < 1, \\ \infty, & \rho \geq 1, \end{cases}$$

where $\omega (0 < \omega < 1)$ is the root of the equation $z = f_\chi(\lambda - \lambda z)$ in $(0, 1)$.

Further, let $b^{<i>}$ be the duration of the server's busy period which begins with i (≥ 1) customers. Due to a Poisson arrival process of customers, the distribution function of $b^{<i>}$ can be expressed as $P\left\{b^{<i>} \leq t\right\} = F_B^{(i)}(t), t \geq 0, i \geq 1$.

Next, we introduce the joint probability distribution of the queue size and the server's busy period by $Q_j(t) = P\{0 \leq t < b; N(t) = j\}$. Thus, $Q_j(t)$ can be interpreted as the transient probability of queue size being j at time point t during the period $(0, b]$. It implies that $N(t) = j$ in $Q_j(t)$ requires the initial state $N(0) = 1$, $(0, t] \subset (0, b]$ and the time point $t = 0$ is the beginning moment of b. Then, the boundary condition is given by $Q_1(0) = 1$, $Q_j(0) = 0$, $j > 1$.

Lemma 2 *Let $q_j^*(s) = \int_0^\infty e^{-st} Q_j(t)\, dt$ denote the Laplace transform of $Q_j(t)$. For $\Re(s) > 0$ and $j \geq 1$, we have the recursive expression of $q_j^*(s)$ as follows:*

$$q_1^*(s) = \frac{f_B(s)\, \bar{f}_\chi(s+\lambda)}{(s+\lambda)\, f_\chi(s+\lambda)},$$

$$q_j^*(s) = \frac{f_B(s)}{f_\chi(s+\lambda)} \int_0^\infty e^{-st} \bar{F}_\chi(t) \frac{(\lambda t)^{j-1}}{(j-1)!} e^{-\lambda t} dt + \frac{1}{f_\chi(s+\lambda)} \sum_{k=1}^{j-1} \frac{q_{j-k}^*(s)}{f_B^k(s)} \left\{ f_B(s) - \sum_{i=0}^{k} \int_0^\infty e^{-(s+\lambda)t} \frac{[\lambda f_B(s)\, t]^i}{i!} dF_\chi(t) \right\}, j > 1,$$

where $f_B(s)$ is defined as in Lemma 1. $\sum_{k=i}^{j} = 0$ if $j < i$.

Proof See Kuang et al. (2022). □

Now, we analyze the transient queue length distribution of the system with an arbitrary initial state. Let $p_{ij}(t) = P\{N(t) = j | N(0) = i\}$ be the conditional probability that queue size is j at any time t with initial state $N(0) = i$ $(i = 0, 1, 2, \cdots)$, and its Laplace transform can be given by $p_{ij}^*(s) = \int_0^\infty e^{-st} p_{ij}(t) dt$, $i, j = 0, 1, 2, \cdots$.

Theorem 1 *For $\Re(s) > 0$, we have*

$$p_{00}^*(s) = \frac{\bar{f}_\tau(s)}{s} \left\{ 1 + \frac{f_\tau(s) f_B(s) \bar{f}_V(s+\lambda)}{\bar{f}_V(s+\lambda)[1 - \bar{f}_Y(s+\lambda) f_\tau(s) f_B(s)] - \Delta_N(s)} \right\}, \quad (1)$$

$$p_{i0}^*(s) = \frac{\bar{f}_\tau(s)}{s} \cdot \frac{f_B^i(s) \bar{f}_V(s+\lambda)}{\bar{f}_V(s+\lambda)[1 - \bar{f}_Y(s+\lambda) f_\tau(s) f_B(s)] - \Delta_N(s)}, i \geq 1, \quad (2)$$

where

$$\Delta_N(s) = f_Y(s+\lambda) \left[f_B^N(s) \Phi_N(s) + f_V(s+\lambda - \lambda f_B(s)) - \Psi_N(s) \right],$$

$$f_V(s+\lambda) = \int_0^\infty e^{-(s+\lambda)t} dF_V(t),\ f_Y(s+\lambda) = \int_0^\infty e^{-(s+\lambda)t} dF_Y(t),$$

$$\Phi_N(s) = \sum_{n=1}^{N-1} \int_0^\infty f_\tau^{N-n}(s) \frac{[f_B(s)\lambda t]^n}{n!} e^{-(s+\lambda)t} dF_V(t),$$

$$\Psi_N(s) = \sum_{n=0}^{N-1} \int_0^\infty \frac{[f_B(s)\lambda t]^n}{n!} e^{-(s+\lambda)t} dF_V(t).$$

Proof See Appendix. □

Theorem 2 *For $\Re(s) > 0$, we have*

1. *If $j = 1, 2, \cdots, N-1$, then*

$$p_{0j}^*(s) = f_\tau(s) q_j^*(s) + f_\tau(s) f_B(s) \cdot \frac{s f_\tau(s) q_j^*(s) \bar{f}_Y(s+\lambda) \bar{f}_V(s+\lambda) + f_Y(s+\lambda) \left[f_\tau^j(s) \bar{f}_\tau(s) \bar{f}_V(s+\lambda) + s\theta_j(s) \right]}{s \left\{ \bar{f}_V(s+\lambda) \left[1 - \bar{f}_Y(s+\lambda) f_\tau(s) f_B(s) \right] - \Delta_N(s) \right\}}, \tag{3}$$

$$p_{ij}^*(s) = \sum_{k=1}^{i} q_{j-i+k}^*(s) f_B^{k-1}(s) + f_B^i(s) \cdot \frac{s f_\tau(s) q_j^*(s) \bar{f}_Y(s+\lambda) \bar{f}_V(s+\lambda) + f_Y(s+\lambda) \left[f_\tau^j(s) \bar{f}_\tau(s) \bar{f}_V(s+\lambda) + s\theta_j(s) \right]}{s \left\{ \bar{f}_V(s+\lambda) \left[1 - \bar{f}_Y(s+\lambda) f_\tau(s) f_B(s) \right] - \Delta_N(s) \right\}}. \tag{4}$$

2. *If $j \geq N$, then*

$$p_{0j}^*(s) = f_\tau q_j^*(s) + f_\tau(s) f_B(s) \cdot \frac{f_\tau(s) q_j^*(s) \bar{f}_Y(s+\lambda) \bar{f}_V(s+\lambda) + f_Y(s+\lambda) \sigma_j(s)}{\bar{f}_V(s+\lambda) \left[1 - \bar{f}_Y(s+\lambda) f_\tau(s) f_B(s) \right] - \Delta_N(s)}, \tag{5}$$

$$p_{ij}^*(s) = \sum_{k=1}^{i} q_{j-i+k}^*(s) f_B^{k-1}(s) + f_B^i(s) \cdot \frac{f_\tau(s) q_j^*(s) \bar{f}_Y(s+\lambda) \bar{f}_V(s+\lambda) + f_Y(s+\lambda) \sigma_j(s)}{\bar{f}_V(s+\lambda) \left[1 - \bar{f}_Y(s+\lambda) f_\tau(s) f_B(s) \right] - \Delta_N(s)}, \tag{6}$$

where

$$\sigma_j(s) = \theta_j(s) + \int_0^\infty \bar{F}_V(t) e^{-(s+\lambda)t} \frac{(\lambda t)^j}{j!} dt,$$

$$\theta_j(s) = \Phi_N(s) \sum_{k=1}^{N} q_{j-N+k}^*(s) f_B^{k-1}(s) + \sum_{n=N}^{\infty} \sum_{k=1}^{n} q_{j-n+k}^*(s) f_B^{k-1}(s) \int_0^\infty \frac{(\lambda t)^n}{n!} e^{-(s+\lambda)t} dF_V(t),$$

$\Delta_N(s)$ and $\Phi_N(s)$ are stated as in Theorem 1.

Proof See Appendix. □

4 The Recursive Solution of the Steady-State Queue Length Distribution

In this section, based on the transient results presented in Theorems 1 and 2 above, the explicit recursive formulas for the stationary queue-length distribution are obtained by applying

L'Hospital's rule. Furthermore, other critical queueing performance metrics of importance are derived by some algebraic manipulation.

Theorem 3 *Let* $p_j = \lim_{t\to\infty} P\{N(t) = j\}$, $j = 0, 1, 2, \cdots$, *we have*

1. *When* $\rho \geq 1$, $p_j = 0$, $j = 0, 1, 2, \cdots$.
2. *When* $\rho < 1$, *the recursive formulas of* $\{p_j, j = 0, 1, 2, \cdots\}$ *are listed as follows,*

$$p_0 = (1-\rho)\frac{1-v_0}{M_N}, \tag{7}$$

$$p_j = \lambda(1-\rho)\frac{(1-v_0)\,[\bar{f}_Y(\lambda)q_j + \frac{1}{\lambda}f_Y(\lambda)] + f_Y(\lambda)\,\theta_j}{M_N},\ j = 1, 2, \cdots, N-1, \tag{8}$$

$$p_j = \lambda(1-\rho)\frac{(1-v_0)\,\bar{f}_Y q_j + f_Y(\lambda)\,\sigma_j}{M_N},\ j = N, N+1, \cdots, \tag{9}$$

and $\{p_j, j = 0, 1, 2, \cdots\}$ *forms a probability distribution, where*

$$\begin{aligned}
v_0 &= \int_0^\infty e^{-\lambda t} dF_V(t),\ f_Y(\lambda) = \int_0^\infty e^{-\lambda t} dF_Y(t),\\
M_N &= (1-v_0)\bar{f}_Y(\lambda) + f_Y(\lambda)\left[N\sum_{m=1}^{N-1} v_m + \lambda\int_0^\infty t F_\tau^{(N-1)}(t) dF_V(t),\right]\\
q_j &= \lim_{s\to 0^+} q_j^*(s)\\
&= \frac{1}{f_\chi(\lambda)}\int_0^\infty \bar{F}_\chi(t)\frac{(\lambda t)^{j-1}}{(j-1)!}e^{-\lambda t}dt\\
&\quad + \frac{1}{f_\chi(\lambda)}\sum_{k=1}^{j-1} q_{j-k}\left[1 - \sum_{i=0}^{k}\int_0^\infty \frac{(\lambda t)^i}{i!}e^{-\lambda t} dF_\chi(t)\right],\\
\theta_j &= \sum_{k=1}^{N} q_{j-N+k}\sum_{n=1}^{N-1} v_n + \sum_{n=N}^{\infty}\sum_{k=1}^{n} q_{j-n+k}v_n,\\
\sigma_j &= \theta_j + \int_0^\infty \bar{F}_V(t)\frac{(\lambda t)^j}{j!}e^{-\lambda t}dt.
\end{aligned}$$

Proof It is noted that the stationary probability of queue length can be calculated by

$$p_j = \lim_{t\to\infty}\sum_{i=0}^{\infty} P\{N(0) = i\}p_{ij}(t) = \sum_{i=0}^{\infty} P\{N(0) = i\}\lim_{t\to\infty} p_{ij}(t)$$

and

$$\lim_{t\to\infty} p_{ij}(t) = \lim_{s\to 0^+} s p_{ij}^*(s).$$

Hence, we have to compute $\lim_{s\to 0^+} s p_{ij}^*(s)$.

1. When $\rho > 1$, due to $\lim_{s\to 0^+} f_B(s) = \omega(0 < \omega < 1)$, it yields

$$\begin{aligned}
&\lim_{s\to 0^+} \bar{f}_V(s+\lambda)\left[1-\bar{f}_Y(s+\lambda) f_\tau(s) f_B(s)\right] - \Delta_N(s) \\
&= (1-v_0)\{1-[1-f_Y(\lambda)]\omega\} \\
&\quad - f_Y(\lambda)\left[\omega^N \sum_{n=1}^{N-1} \int_0^\infty \frac{(\lambda\omega t)^n}{n!} e^{-\lambda t} dF_V(t) + f_V(\lambda-\lambda\omega)\right. \\
&\quad \left. - \sum_{n=0}^{N-1} \int_0^\infty \frac{(\lambda\omega t)^n}{n!} e^{-\lambda t} dF_V(t)\right] \\
&= (1-v_0)\{1-[1-f_Y(\lambda)]\omega\} - f_Y(\lambda)[f_V(\lambda-\lambda\omega)-v_0] \\
&\quad + f_Y(\lambda)(1-\omega^N) \sum_{n=1}^{N-1} \int_0^\infty \frac{(\lambda\omega t)^n}{n!} e^{-\lambda t} dF_V(t).
\end{aligned} \tag{10}$$

We noted that

$$1 - v_0 > f_V(\lambda-\lambda\omega) - v_0 > 0, 1-[1-f_Y(\lambda)]\omega > 1-[1-f_Y(\lambda)] = f_Y(\lambda) > 0,$$

and

$$f_Y(\lambda)\left(1-\omega^N\right) \sum_{n=1}^{N-1} \int_0^\infty \frac{(\lambda\omega t)^n}{n!} e^{-\lambda t} dF_V(t) \geq 0.$$

Hence,

$$\begin{aligned}
&\lim_{s\to 0^+} \bar{f}_V(s+\lambda)\left[1-\bar{f}_Y(s+\lambda) f_\tau(s) f_B(s)\right] - \Delta_N(s) \\
&= \{(1-v_0)[1-[1-f_Y(\lambda)]\omega] - f_Y(\lambda)[f_V(\lambda-\lambda\omega)-v_0]\} \\
&\quad + f_Y(\lambda)\left(1\text{-}\omega^N\right) \sum_{n=1}^{N-1} \int_0^\infty \frac{(\lambda\omega t)^n}{n!} e^{-\lambda t} dF_V(t) \\
&\neq 0.
\end{aligned}$$

Combining the expressions presented in Theorems 1 and 2, we can obtain $\lim_{t\to\infty} p_{ij}(t) = \lim_{s\to 0^+} sp_{ij}^*(s) = 0$ by a direct calculation, $i, j = 0, 1, 2, \cdots$.
When $\rho = 1$, due to $\lim_{s\to 0^+} f_B(s) = 1$ and $E[b] = \infty$, we have

$$\begin{aligned}
&\lim_{s\to 0^+} \bar{f}_V(s+\lambda)\left[1-\bar{f}_Y(s+\lambda) f_\tau(s) f_B(s)\right] - \Delta_N(s) \\
&= (1-v_0) f_Y(\lambda) - f_Y(\lambda)\left(\sum_{n=1}^{N-1} v_n + 1 - \sum_{n=0}^{N-1} v_n\right) \\
&= (1-v_0) f_Y(\lambda) - f_Y(\lambda)(1-v_0) \\
&= 0,
\end{aligned} \tag{11}$$

and

$$\begin{aligned}
&\lim_{s\to 0^+}\frac{d}{ds}\left\{\bar{f}_V(s+\lambda)\left[1-\bar{f}_Y(s+\lambda)f_\tau(s)\,f_B(s)\right]-\Delta_N(s)\right\}\\
&=E[b]\left\{1-v_0-f_Y(\lambda)\left[1-\lambda E[V]-\sum_{n=1}^{N-1}(N-1)v_n\right]\right\}\\
&\quad+\frac{(1-v_0)[1-f_Y(\lambda)]}{\lambda}+f_Y(\lambda)\left\{E[V]+\frac{1}{\lambda}\sum_{n=1}^{N-1}(N-n)v_n\right\}\\
&=\infty
\end{aligned}\tag{12}$$

holds. Applying L'Hospital's rule leads to $\lim_{s\to 0^+} sp_{ij}^*(s)=0, i,j=0,1,2,\cdots$. Therefore, for $\rho\geq 1$, we have $p_j=0,\ j=0,1,2\cdots$.

2. When $\rho<1$, because of $\lim_{s\to 0^+} f_B(s)=1$ and $E[b]=\frac{\rho}{\lambda(1-\rho)}$, we get

$$\begin{aligned}
&\lim_{s\to 0^+}\frac{d}{ds}\left\{\bar{f}_V(s+\lambda)\left[1-\bar{f}_Y(s+\lambda)f_\tau(s)\,f_B(s)\right]-\Delta_N(s)\right\}\\
&=\frac{(1-v_0)[1-f_Y(\lambda)]}{\lambda(1-\rho)}+\frac{Nf_Y(\lambda)}{\lambda(1-\rho)}\sum_{m=1}^{N-1}v_m\\
&\quad+\frac{\lambda f_Y(\lambda)}{\lambda(1-\rho)}\int_0^\infty tF_\tau^{(N-1)}(t)dF_V(t).
\end{aligned}\tag{13}$$

By using L'Hospital's rule again and directly calculating, we can obtain the recursive expressions of $\{p_j, j=0,1,2,\cdots\}$.

For $\rho<1$, it is easy to see that

$$\begin{aligned}
\sum_{j=0}^{\infty}p_j&=\frac{(1-\rho)(1-v_0)}{M_N}+\frac{\lambda(1-\rho)(1-v_0)\bar{F}_Y(\lambda)}{M_N}\sum_{j=1}^{\infty}q_j\\
&\quad+\frac{(N-1)(1-\rho)(1-v_0)f_Y(\lambda)}{M_N}\\
&\quad+\frac{\lambda(1-\rho)f_Y(\lambda)}{M_N}\left[\sum_{j=1}^{\infty}\theta_j+\int_0^\infty\bar{F}_V(t)F_\tau^{(N)}(t)dt\right].
\end{aligned}\tag{14}$$

After calculation and simplification, it follows that

$$\begin{aligned}
\sum_{j=1}^{\infty}\theta_j&=\sum_{j=1}^{\infty}\left(\sum_{k=1}^{N}q_{j-N+k}\sum_{n=1}^{N-1}v_n+\sum_{n=N}^{\infty}\sum_{k=1}^{n}q_{j-n+k}\cdot v_n\right)\\
&=\sum_{j=1}^{\infty}\sum_{k=1}^{N}q_{j-N+k}\sum_{n=1}^{N-1}v_n+\sum_{j=1}^{\infty}\sum_{n=N}^{\infty}\sum_{k=1}^{n}q_{j-n+k}\cdot v_n\\
&=(\sum_{j=1}^{\infty}q_j)\left(N\sum_{n=1}^{N-1}v_n+\sum_{n=N}^{\infty}n\cdot v_n\right)\\
&=(\sum_{j=1}^{\infty}q_j)\left[N\sum_{n=1}^{N-1}v_n+\lambda\int_0^\infty tF_\tau^{(N-1)}(t)dF_V(t)\right],
\end{aligned}\tag{15}$$

$$\text{and} \quad \sum_{j=1}^{\infty} q_j = E[b] = \frac{\rho}{\lambda(1-\rho)} \tag{16}$$

holds. Then, substituting Eqs. (15) and (16) into Eq. (14), $\sum_{j=0}^{\infty} p_j = 1$ can be easily obtained. That is, $\{p_j, j = 0, 1, 2, \cdots\}$ which satisfies the stationary condition $\rho < 1$ forms a probability distribution. □

Theorem 4 *When $\rho < 1$ and $|z| < 1$, let $P(z)$ be the probability generating function (p.g.f.) of $\{p_j, j = 0, 1, 2, \cdots\}$, then*

$$P(z) = \frac{(1-\rho)(1-z) f_\chi(\lambda(1-z))}{f_\chi(\lambda(1-z)) - z} \cdot \frac{(1-z)(1-v_0) + f_Y(\lambda)\left[z(1-v_0) - f_V(\lambda(1-z)) + v_0 + \sum_{m=1}^{N-1}(z^m - z^N)v_m\right]}{(1-z)M_N}, \tag{17}$$

and the mean of steady-state queue size, denoted by $E[L]$, is presented by

$$E[L] = \rho + \frac{\lambda^2 E[\chi^2]}{2(1-\rho)} + \frac{f_Y(\lambda)\left\{\sum_{m=1}^{N-1}[N(N-1) - m(m-1)]v_m + \lambda^2 E[V^2]\right\}}{2M_N}, \tag{18}$$

where M_N is determined in Theorem 3, $f_\chi(\lambda(1-z)) = \int_0^\infty e^{-\lambda(1-z)t} dF_\chi(t)$, $f_V(\lambda(1-z)) = \int_0^\infty e^{-\lambda(1-z)t} dF_V(t)$.

Proof According to the definition of $p.g.f.$ and the expression of p_j given in Theorem 3, it yields

$$\begin{aligned} P(z) = \sum_{j=0}^{\infty} z^j p_j &= \frac{(1-\rho)(1-v_0)}{M_N} + \frac{\lambda(1-\rho)\bar{f}_Y(\lambda)(1-v_0)}{M_N}\sum_{j=1}^{\infty} z^j q_j \\ &+ \frac{\lambda(1-\rho) f_Y(\lambda)}{M_N}\sum_{j=1}^{\infty} z^j \theta_j + \frac{(1-\rho)(1-v_0) f_Y(\lambda)}{M_N}\cdot\frac{z - z^N}{1-z} \\ &+ \frac{\lambda(1-\rho) f_Y(\lambda)}{M_N}\sum_{j=N}^{\infty} z^j \int_0^\infty \bar{F}_V(t)\frac{(\lambda t)^j}{j!}e^{-\lambda t} dt. \end{aligned} \tag{19}$$

Through calculation and simplification, we get the following equations,

$$\sum_{j=1}^{\infty} z^j q_j = \frac{z\left[1 - f_\chi(\lambda(1-z))\right]}{\lambda\left[f_\chi(\lambda(1-z)) - z\right]}, \tag{20}$$

$$\sum_{j=1}^{\infty} z^j \theta_j = \left(\sum_{j=1}^{\infty} z^j q_j\right)\left[\frac{1-z^N}{1-z}\sum_{m=1}^{N-1} v_m + \sum_{m=N}^{\infty}\frac{1-z^m}{1-z}v_m\right], \tag{21}$$

$$\begin{aligned}&\lambda \sum_{j=N}^{\infty} z^j \int_0^{\infty} \bar{F}_V(t) \frac{(\lambda t)^j}{j!} e^{-\lambda t} dt \\ &= \frac{1}{1-z} \left\{ v_0 - f_V(\lambda(1-z)) + z^N (1-v_0) + \sum_{m=1}^{N-1} \left(z^m - z^N \right) v_m \right\}.\end{aligned} \tag{22}$$

Then, substituting Eqs. (20)–(22) into (19) leads to Eq. (17), and Eq. (18) can be derived by using $E[L] = \frac{d}{dz}[P(z)]|_{z=1}$. □

Theorem 5 ***(Stochastic decomposition structure of the steady-state queue size)*** *For $\rho < 1$, the steady-state queue size studied in this paper can be decomposed into the sum of two independent parts: one is the steady-state queue size of the classic $M/G/1$ queue that has been studied (see e.g., Cohen 1982), and the other is the additional queue size L_d caused by N-policy and delayed uninterrupted multiple vacation, which is distributed as*

$$P\{L_d = 0\} = \frac{1-v_0}{M_N}, \tag{23}$$

$$P\{L_d = j\} = \frac{f_Y(\lambda)(1-v_0)}{M_N}, \; j = 1, 2, \cdots, N-1, \tag{24}$$

$$P\{L_d = j\} = \frac{f_Y(\lambda) \int_0^{\infty} F_{\tau}^{(j+1)}(t) dF_V(t)}{M_N}, \; j = N, N+1, \cdots, \tag{25}$$

where M_N is determined in Theorem 3.

Proof From the Eq. (17) above, it can be seen that the stochastic decomposition property of the steady-state queue size holds. In order to obtain Eqs. (23)–(25), let

$$\begin{aligned}P_V(z) &= \frac{(1-z)(1-v_0) + f_Y(\lambda) \left[z(1-v_0) - f_V(\lambda(1-z)) + v_0 + \sum_{m=1}^{N-1} (z^m - z^N) v_m \right]}{(1-z) M_N} \\ &= \frac{1-v_0}{M_N} + \frac{f_Y(\lambda)}{M_N} H(z) \cdot I(z),\end{aligned} \tag{26}$$

where $H(z) = z(1-v_0) - f_V(\lambda(1-z)) + v_0 + \sum_{m=1}^{N-1} (z^m - z^N) v_m$, and $I(z) = \frac{1}{1-z}$.

Using a straightforward calculation, the following results are easily obtained.

$$\begin{aligned}&H^{(0)}(z)|_{z=0} = H(z)|_{z=0} = 0, \; H^{(1)}(z)|_{z=0} = 1 - v_0, \\ &H^{(j)}(z)|_{z=0} = 0, \; 2 \le j \le N-1, \\ &H^{(N)}(z)|_{z=0} = -N! \sum_{m=1}^{N-1} v_m - \int_0^{\infty} (\lambda t)^N e^{-\lambda t} dF_V(t), \\ &H^{(j)}(z)|_{z=0} = -\int_0^{\infty} (\lambda t)^j e^{-\lambda t} dF_V(t), \; j \ge N+1, \\ &I^{(j)}(z)|_{z=0} = j!, \; j \ge 0.\end{aligned}$$

Then, employing

$$P\{L_d = j\} = \frac{1}{j!} \cdot \frac{d^j}{dz^j} [P_V(z)]|_{z=0}$$

and

$$[H(z) \cdot I(z)]^{(j)} = \sum_{k=0}^{j} C_j^k H^{(k)}(z) I^{(j-k)}(z),$$

we can get Eqs. (23)–(25) by carrying some algebraic simplifications, where $H^{(k)}(z)$ represents the k-th order derivative of $H(z)$ with respect to z, and $C_m^k = \frac{m!}{k!(m-k)!}$. □

Theorem 6 *Let $E[W]$ denote the average waiting time of an arbitrary customer, then for $\rho < 1$, we have*

$$E[W] = \frac{\lambda E\left[\chi^2\right]}{2(1-\rho)} + \frac{f_Y(\lambda)\left\{\sum_{m=1}^{N-1}[N(N-1) - m(m-1)]\, v_m + \lambda^2 E[V^2]\right\}}{2\lambda M_N}, \tag{27}$$

where M_N is given in Theorem 3.

Proof According to the model description, we can see that customers are served in a first-come-first-served manner, and the arrival process is a Poisson process with rate $\lambda(>0)$. It is also observed that in equilibrium, the number of customers in the queue at a departure is equal to the number of customers that arrived during the departing customer's sojourn time. Therefore, Little's law holds, and the Eq. (27) can be derived by $E[W] = \frac{E[L]}{\lambda} - E[\chi]$. □

5 The Optimal Control Policy under the Cost Model

5.1 The Cost Model and Cost Function

This section offers four numerical examples to determine the optimal control policy N^* and the optimal two-dimensional control policy (N^*, T^*) for minimizing the system cost. Considering actual operating costs of the system, we first establish the cost structure as follows:

1. $R \equiv$ fixed setup cost of the system for each time (this cost is due to the consumption of switching on the server each time);
2. $h \equiv$ fixed holding cost for each customer per unit time in the system (the cost arises from the customer's sojourn time, including the waiting and service times).

Let $C(N)$ donate the long-run expected cost of the system per unit of time. Applying the renewal reward theorem, we get

$$\begin{aligned} C(N) &= \frac{\text{Expected cost within a busy cycle}}{\text{Expected length of a busy cycle}} \\ &= hE[L] + \frac{R}{E[I] + E[Z]}, \end{aligned} \tag{28}$$

where $E[L]$ is given by Eq. (18) as mentioned above, I denotes the length of a server's non-busy period (the time interval from the moment the system becomes empty until the moment the server returns to the system from vacation and provides service), which is similar to the definition in the paper by Luo et al. (2023). Let Z denote the length of a server's busy period

and U_b be the queue size at the beginning of Z, and its probability distribution is given by

$$P\{U_b = 1\} = 1 - f_Y(\lambda),$$
$$P\{U_b = N\} = \frac{f_Y(\lambda)}{1 - v_0} \sum_{n=1}^{N} v_n,$$
$$P\{U_b = n\} = \frac{f_Y(\lambda)}{1 - v_0} v_n, n = N+1, N+2, \cdots.$$

Therefore, we can obtain the average value of U_b as $E[U_b] = \frac{M_N}{1 - f_V(\lambda)}$.

Using the above Lemma 1, we can calculate the average value of Z as follows:

$$E[Z] = E[b] E[U_b] = \frac{\rho M_N}{\lambda(1-\rho)(1-v_0)}, \rho < 1. \tag{29}$$

Since the queue size at the beginning of Z is equal to the number of customers arriving during I and customers arrival process is a Poisson with rate λ (> 0), the average length of the server's non-busy period can be represented by $E[I] = \frac{E[U_b]}{\lambda}$. Then,

$$E[I] + E[Z] = \frac{M_N}{\lambda(1-\rho)(1-v_0)}, \rho < 1. \tag{30}$$

Substituting Eqs. (18) and (30) into Eq. (28), we have

$$C(N) = h\left\{\rho + \frac{\lambda^2 E[\chi^2]}{2(1-\rho)} + \frac{f_Y(\lambda)\left\{\sum_{m=1}^{N-1}[N(N-1) - m(m-1)]v_m + \lambda^2 E[V^2]\right\}}{2M_N}\right\} + \frac{R\lambda(1-\rho)(1-v_0)}{M_N}, \tag{31}$$

where M_N is given in Theorem 3.

5.2 Optimal Control Policy without the Average Waiting Time Constraint

From the above Eq. (31), we can see that the cost function $C(N)$ is a non-linear function that depends on the decision variable N. It is too difficult to find the optimal solution of Eq. (31) in an analytical manner. Therefore, we next search for the optimal solution through the following numerical examples.

Example 1 We consider a practical situation related to a maintenance system—an auto repair shop to illustrate the theoretical results of the model. The manager of an auto repair shop invited a repair expert with excellent skills to be responsible for repairing a particular car brand. Considering the needs of the business, the manager asked the expert to wait for a random time Y instead of leaving immediately for a vacation V after all repair work is finished. Once a car needs repair during Y, the repair expert provides service immediately until no more cars need repair. After that, the expert restarts another delay period Y. In order to save costs (including costs of the expert's appearance and starting repairs), the manager designs the following repair strategy according to the number of arriving cars during V: (i)If no cars are waiting for repair after V expires, the maintenance expert will continue to be on

vacation; (ii)If the number of arrivals is greater than or equal to N; the expert immediately repair these cars at the end of V until there is no car again; (iii)If the number of arrivals is less than N but greater than 0, the expert waits for N customers to accumulate in the system before providing repair.

Assume that cars arrive at the auto repair shop according to a Poisson process with the parameter $\lambda(>0)$. The repair time of each car, the delay period Y and the time V for the repair expert to continuously repair other brands of cars obey different PH distributions with representation $(\boldsymbol{\alpha}, \boldsymbol{L})$, $(\boldsymbol{\beta}, \boldsymbol{K})$ and $(\boldsymbol{\gamma}, \boldsymbol{S})$ of orders m, n and i, respectively. Vectors $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots \alpha_m)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots \beta_n)$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots \gamma_i)$ can be interpreted as the initial probability vectors among the transient states, where $\boldsymbol{\alpha e}_m = 1$, $\boldsymbol{\beta e}_n = 1$ and $\boldsymbol{\gamma e}_i = 1$. $\boldsymbol{e}_m$, $\boldsymbol{e}_n$ and $\boldsymbol{e}_i$ represent column vectors of dimensions m, n and i with all elements equal to one, respectively. Column vectors $\boldsymbol{L}^0$, $\boldsymbol{K}^0$ and $\boldsymbol{S}^0$ meet the conditions $\boldsymbol{Le_m} + \boldsymbol{L}^0 = 0$, $\boldsymbol{Ke_n} + \boldsymbol{K}^0 = 0$ and $\boldsymbol{Se_i} + \boldsymbol{S}^0 = 0$, respectively. According to Eq. (31), the cost function $C(N)$ without the average waiting time constraint can be expressed as

$$\begin{aligned} C(N) = h \Bigg\{ & -\lambda \boldsymbol{\alpha} \boldsymbol{L}^{-1} \boldsymbol{e}_m + \frac{\lambda^2 \boldsymbol{\alpha} \boldsymbol{L}^{-2} \boldsymbol{e}_m}{1 + \lambda \boldsymbol{\alpha} \boldsymbol{L}^{-1} \boldsymbol{e}_m} \\ & + \frac{\zeta(\psi_N + 2\lambda^2 \boldsymbol{\gamma} \boldsymbol{S}^{-2} \boldsymbol{e}_i)}{2\left[(1-\eta)(1-\zeta) + \zeta(\varphi_N - \lambda \boldsymbol{\gamma} \boldsymbol{S}^{-1} \boldsymbol{e}_i)\right]} \Bigg\} \\ & + \frac{R\lambda \left(1 + \lambda \boldsymbol{\alpha} \boldsymbol{L}^{-1} \boldsymbol{e}_m\right)(1-\eta)}{(1-\eta)(1-\zeta) + \zeta(\varphi_N - \lambda \boldsymbol{\gamma} \boldsymbol{S}^{-1} \boldsymbol{e}_i)}, \end{aligned} \tag{32}$$

where $\boldsymbol{I}_n$ represents the identity matrix of dimension n,

$$\begin{aligned} \zeta &= \boldsymbol{\beta}(\lambda \boldsymbol{I}_n - \boldsymbol{K})^{-1} \boldsymbol{K}^0, \eta = \boldsymbol{\gamma}(\lambda \boldsymbol{I}_i - \boldsymbol{S})^{-1} \boldsymbol{S}^0, \\ \varphi_N &= \boldsymbol{\gamma} \sum_{k=1}^{N-1} (N-k)\, \lambda^k (\lambda \boldsymbol{I}_i - \boldsymbol{S})^{-k-1} \boldsymbol{S}^0, \\ \psi_N &= \sum_{k=1}^{N-1} [N(N-1) - k(k-1)] \lambda^k \boldsymbol{\gamma} (\lambda \boldsymbol{I}_i - \boldsymbol{S})^{-k-1} \boldsymbol{S}^0. \end{aligned}$$

We set $R = 200$, $h = 3$, $\lambda = 0.8$, $\boldsymbol{L} = \begin{pmatrix} -15 & 1.5 & 0.1 \\ 0.2 & -1 & 0.2 \\ 0.3 & 0.1 & -4 \end{pmatrix}$, $\boldsymbol{\alpha} = \begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \end{pmatrix}^\top$, $\boldsymbol{L}^0 = \begin{pmatrix} 13.4 \\ 0.6 \\ 3.6 \end{pmatrix}$, $\boldsymbol{K} = \begin{pmatrix} -5 & 2 & 2 \\ 0.5 & -12 & 3 \\ 0.8 & 0.2 & -20 \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} 0.1 \\ 0.8 \\ 0.1 \end{pmatrix}^\top$, $\boldsymbol{K}^0 = \begin{pmatrix} 1 \\ 8.5 \\ 19 \end{pmatrix}$, $\boldsymbol{S} = \begin{pmatrix} -2 & 1 & 0.5 \\ 0.2 & -1 & 0.4 \\ 0.6 & 0.2 & -1.2 \end{pmatrix}$, $\boldsymbol{\gamma} = \begin{pmatrix} 0.3 \\ 0.3 \\ 0.4 \end{pmatrix}^\top$, $\boldsymbol{S}^0 = \begin{pmatrix} 0.5 \\ 0.4 \\ 0.4 \end{pmatrix}$. Then, we obtain the value of $C(N)$ for different values of decision variable N and plot the figure via the MATLAB software (hold 4 digits after the decimal point). The numerical results are displayed in Table 1. Figure 1 plots the curve of $C(N)$ with respect to N.

From Table 1 and Fig. 1, the optimal value of N is achieved at $N^* = 8$ and its corresponding minimum function value $C(N^*) = 42.5158$.

Example 2 If the random variable V in Example 1 is a fixed length $T(\geq 0)$, that is, $P\{V = T\} = 1$. The two-dimensional cost function, denoted by $C(N, T)$, can be expressed

Table 1 Numerical results for $C(N)$ under different N

N	$C(N)$	N	$C(N)$	N	$C(N)$	N	$C(N)$	N	$C(N)$
1	60.4075	7	42.8907	13	44.8299	19	51.2040	25	58.8405
2	56.4088	8	**42.5158**	14	45.7350	20	52.4200	26	60.1743
3	51.7797	9	42.5298	15	46.7199	21	53.6630	27	61.5204
4	48.0382	10	42.8305	16	47.7696	22	54.9294	28	62.8775
5	45.4361	11	43.3458	17	48.8724	23	56.2160	29	64.2444
6	43.7989	12	44.0245	18	50.0196	24	57.5204	30	65.6201

as

$$C(N,T) = h\left\{-\lambda\boldsymbol{\alpha}\boldsymbol{L}^{-1}\boldsymbol{e}_m + \frac{\lambda^2\boldsymbol{\alpha}\boldsymbol{L}^{-2}\boldsymbol{e}_m}{1+\lambda\boldsymbol{\alpha}\boldsymbol{L}^{-1}\boldsymbol{e}_m} + \frac{\zeta[\phi_N+(\lambda T)^2]}{2\left[\left(1-e^{-\lambda T}\right)(1-\zeta)+\zeta(\vartheta_N+\lambda T)\right]}\right\} + \frac{R\lambda\left(1+\lambda\boldsymbol{\alpha}\boldsymbol{L}^{-1}\boldsymbol{e}_m\right)\left(1-e^{-\lambda T}\right)}{\left(1-e^{-\lambda T}\right)(1-\zeta)+\zeta(\vartheta_N+\lambda T)}, \tag{33}$$

where $\boldsymbol{I}_n$ and ζ are determined in Example 1,

$$\phi_N = \sum_{k=1}^{N-1}[N(N-1)-k(k-1)]\frac{(\lambda T)^k}{k!}e^{-\lambda T},\ \vartheta_N = \sum_{k=1}^{N-1}(N-k)\frac{(\lambda T)^k}{k!}e^{-\lambda T}.$$

Since N is a discrete variable, we will calculate the optimum values of N and T for the cost function $C(N,T)$ through the following algorithm:

Step 1: Substituting the system parameters into Eq. (33), and initialize the iteration counter $N=1$.

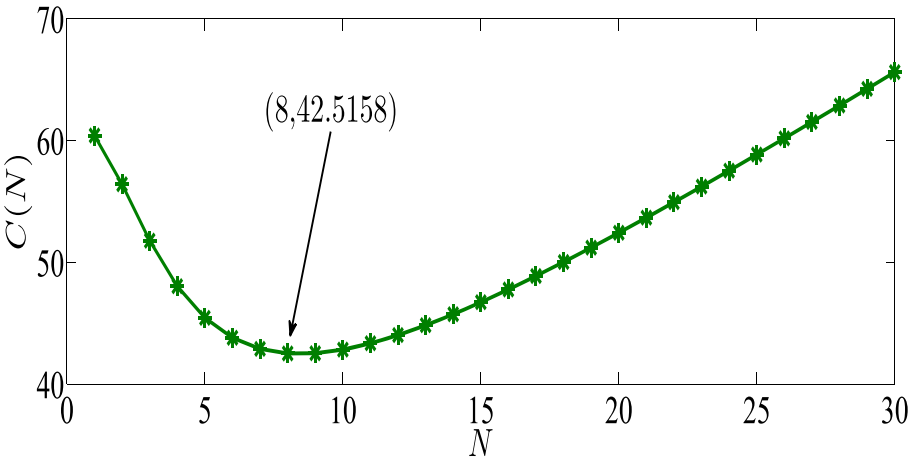

Fig. 1 $C(N)$ varies with different N

Table 2 Numerical results for $T_N{}^*$ and $C\left(N, T_N^*\right)$ under different N

N	$T_N{}^*$	$C\left(N, T_N^*\right)$	N	$T_N{}^*$	$C\left(N, T_N^*\right)$
1	10.5320	26.4114	11	2.0557	27.4416
2	10.5207	26.4066	12	2.1056	28.0545
3	10.4589	26.3898	13	2.1516	28.8006
4	10.4066	**26.3750**	14	2.1881	29.6518
5	10.5558	26.4499	15	2.2294	30.5873
6	11.1142	26.7598	16	2.2819	31.5918
7	12.0732	27.3818	17	2.3049	32.6530
8	1.8981	26.8846	18	2.3416	33.7618
9	1.9538	26.7880	19	2.3691	34.9107
10	2.0103	27.0009	20	2.3870	36.0938

Step 2: Use MATLAB program to search the optimal numerical solution T_N^* of decision variable T so as to minimize the cost function $C\ (N, T)$. The minimum value of the cost function $C\ (N, T)$ can be denoted as $C\left(N, T_N^*\right)$.
Step 3: Set $N = N + 1$ and repeat the **step 2**.
Step 4: If $C\left(N+1, T_{N+1}^*\right) - C\left(N, T_N^*\right) > 0$, then stop and report $(N^*, T^*) = \left(N^*, T_N^*\right)$ and $C\ (N, T^*) = C\left(N, T_N^*\right)$. Otherwise, go to **Step 3** and **Step 4**.

We first choose the cost parameters R, h and the system parameters λ, $\boldsymbol{L}$, $\boldsymbol{\alpha}$, $\boldsymbol{K}$, and $\boldsymbol{\beta}$ to have the same values as in Example 1. Next, we program in MATLAB by using the above algorithm. The obtained numerical results are reported in Table 2. Figure 2 shows the change of $C(N, T)$ for different N and T.

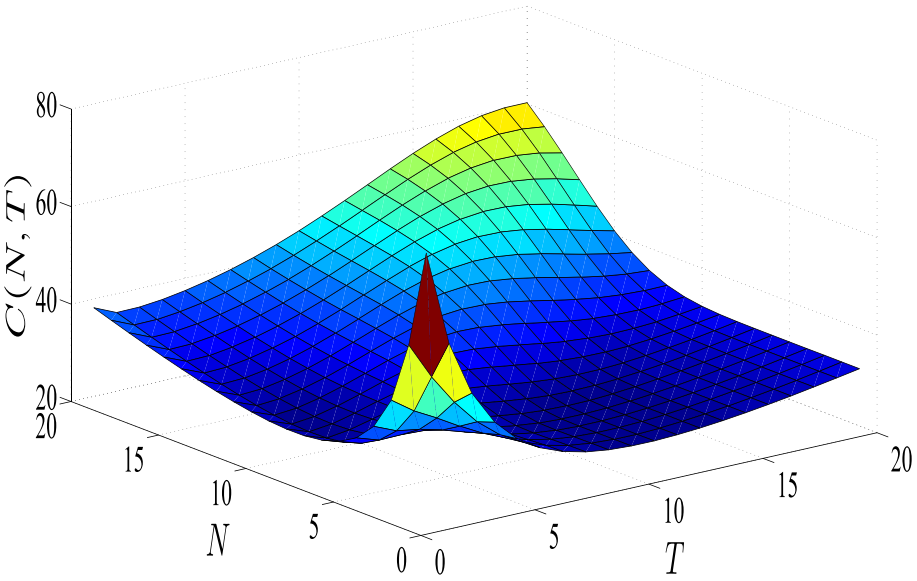

Fig. 2 The change of $C(N, T)$ for differen N and T

We can see from Table 2 that the two-dimensional optimal solution of Eq. (33) $(N^*, T^*) = (4, 10.4066)$ and its corresponding minimum function value $C(N^*, T^*) = 26.3750$. In addition, we can also check the correctness of the results by $\frac{dC(N,T)}{dT}|_{(N=4,T=T_4^*)} \approx 0$.

5.3 Optimal Control Policy with the Average Waiting Time Constraint

Scholars rarely considered the influence of the length of the customers' waiting time on the optimal control policy in previous optimization problems. By reducing the number of system startups, the N-policy can save running costs. Nevertheless, it will also increase customers' waiting times, which will result in lost customers or system congestion if waiting times are too long. It is therefore necessary to examine the optimal control policy under the premise that the average waiting time of an arbitrary customer is less than the threshold $E[W_0]$ in order to minimize customer waiting times and trade off costs between customers and system managers. Next, we will discuss the optimal control policy for economizing the system cost under the same conditions as in Example 1 and Example 2.

Example 3 Let the distributions of all random variables be the same as in Example 1. The cost function $C(N)$ with the average waiting time constraint can be expressed as

$$\begin{cases} \min C(N), \\ s.t. E[W] \leq E[W_0], \end{cases} \tag{34}$$

where $C(N)$ is given by the above Eq. (32), and

$$E[W] = \frac{\lambda \boldsymbol{\alpha} \boldsymbol{L}^{-2} \boldsymbol{e}_m}{1 + \lambda \boldsymbol{\alpha} \boldsymbol{L}^{-1} \boldsymbol{e}_m} + \frac{\zeta(\psi_N + 2\lambda^2 \boldsymbol{\gamma} \boldsymbol{S}^{-2} \boldsymbol{e}_i)}{2\lambda \left[(1-\eta)(1-\zeta) + \zeta(\varphi_N - \lambda \boldsymbol{\gamma} \boldsymbol{S}^{-1} \boldsymbol{e}_i)\right]}.$$

Let the values of all parameters be the same as that in Example 1. We use the MATLAB program to calculate the values of $C(N)$ and $E[W]$ under different values of N and $E[W_0]$. All numerical results are reported in Table 3.

From Table 3, we can reveal that

- If $E[W_0] = 6$, the optimal control threshold value $N^* = 8$ and its corresponding minimum cost value $C(N^*) = 42.5158$ under the constraint condition $E[W] \leq 6$.

Table 3 Numerical results for $C(N)$ and $E[W]$ under different N and $E[W_0]$

	Under $E[W] \leq E[W_0] = 6$		Under $E[W] \leq E[W_0] = 4.5$		Under $E[W] \leq E[W_0] = 3$	
N	$C(N)$	$E[W]$	$C(N)$	$E[W]$	$C(N)$	$E[W]$
1	60.4075	2.8589	60.4075	2.8589	60.4075	2.8589
2	56.4088	2.7499	56.4088	2.7499	56.4088	2.7499
3	51.7797	2.7940	51.7797	2.7940	**51.7797**	2.7940
4	48.0382	3.0278	48.0382	3.0278	____	____
5	45.4361	3.4056	45.4361	3.4056	____	____
6	43.7989	3.8774	**43.7989**	3.8774	____	____
7	42.8907	4.4076	42.8907	4.4076	____	____
8	**42.5158**	4.9737	____	____	____	____
9	42.5298	5.5618	____	____	____	____

Table 4 Numerical results for ${T_N}^*$, $C\left(N, T_N^*\right)$ and $E[W]$ under different N and $E[W_0]$

	Under $E[W] \le E[W_0] = 4.5$			Under $E[W] \le E[W_0] = 3$			Under $E[W] \le E[W_0] = 1.5$		
N	${T_N}^*$	$C\left(N, T_N^*\right)$	$E[W]$	${T_N}^*$	$C\left(N, T_N^*\right)$	$E[W]$	${T_N}^*$	$C\left(N, T_N^*\right)$	$E[W]$
1	8.5652	26,9397	4.5000	5.5631	31.4015	3.0000	2.5474	49.1008	1.5000
2	8.5714	26.9125	4.5000	5.5960	31.0127	3.0000	2.5408	**44.1927**	1.5000
3	8.5695	26.8445	4.5000	5.4827	**30.5611**	3.0000	____	____	____
4	8.4525	**26.8370**	4.5000	4.0851	31.8226	3.0000	____	____	____
5	7.7392	27.4087	4.5000	0.9964	32.1159	3.0000	____	____	____
6	1.8593	28.5561	4.5000	____	____	____	____	____	____
7	1.8644	27.4125	4.5000	____	____	____	____	____	____
8	____	____	____	____	____	____	____	____	____

- $E[W_0] = 4.5$, the optimal control threshold value $N^* = 6$ and its corresponding minimum cost value $C(N^*) = 43.7989$ under the constraint condition $E[W] \le 4.5$.
- If $E[W_0] = 3$, the optimal control threshold value $N^* = 3$ and its corresponding minimum cost value $C(N^*) = 51.7797$ under the constraint condition $E[W] \le 3$.

Example 4 Let the distributions of all random variables be the same as in Example 2. The cost function $C(N, T)$ with the average waiting time constraint can be expressed as

$$\begin{cases} \min C(N, T), \\ s.t. E[W] \le E[W_0], \end{cases} \tag{35}$$

where $C(N, T)$ is given by the above Eq. (33), and

$$E[W] = \frac{\lambda \boldsymbol{\alpha} \boldsymbol{L}^{-2} \boldsymbol{e}_m}{1 + \lambda \boldsymbol{\alpha} \boldsymbol{L}^{-1} \boldsymbol{e}_m} + \frac{\zeta(\phi_N + (\lambda T)^2)}{2\lambda\left[\left(1 - e^{-\lambda T}\right)(1 - \zeta) + \zeta(\vartheta_N + \lambda T)\right]}.$$

Let the values of all parameters be the same as that in Example 2. We use the algorithm given in Example 2 to calculate Eq. (35). The numerical results are shown in Table 4.

By Table 4, we can obtain the following conclusions:

- If $E[W_0] = 4.5$, the optimal control policy $(N^*, T^*) = (4, 8.4525)$ and its corresponding minimum cost value $C(N^*, T^*) = 26.8370$ under the constraint condition $E[W] \le 4.5$.
- If the $E[W_0] = 3$, the optimal control policy $(N^*, T^*) = (3, 5.4827)$ and its corresponding minimum cost value $C(N^*, T^*) = 30.5611$ under the constraint condition $E[W] \le 3$.
- If $E[W_0] = 1.5$, the optimal control policy $(N^*, T^*) = (2, 2.5408)$ and its corresponding minimum cost value $C(N^*, T^*) = 44.1927$ under the constraint condition $E[W] \le 1.5$.

6 Conclusions

In this paper, we proposed a new $M/G/1$ queueing model with delayed uninterrupted multiple vacation under N-policy control and discussed the transient and steady-state properties of the queue size. Our analysis technique differed from traditional analysis methods (e.g., embedded Markov chain and supplementary variable technique). The expressions of the

Laplace transform of the transient queue-length distribution with respect to time t were presented. Then, based on the transient analysis, some crucial queueing performance indices were derived. Finally, we set several numerical examples to discuss the optimal control policy to minimize the expected cost when the random variables obey different PH distributions. Our analysis can help policy-makers set a reasonable threshold for cost optimization in many applications, such as maintenance and order-based systems. The system studied in this paper can be extended to more complex queueing models, such as the non-Markovian arrival process of customers and unreliable servers, which can be the future research directions.

Appendix

Proof of Theorem 1

Proof Let $l_0 = S_0 = 0$, $l_n = \sum_{i=1}^{n} \tau_i$, $S_n = \sum_{i=1}^{n} V_i$, $n \geq 1$. The necessary and sufficient condition for the queue size to be zero is that the time point t is in the system idle period (no customer period), and the i-th system idle period is denoted by $\hat{\tau}_i$ with distribution function $F_\tau(t) = 1 - e^{-\lambda t}$, $i \geq 1$. Since each state change moment of the system is the renewal point, applying the renewal process theory and the law of total probability decomposition, we get

$$\begin{aligned} p_{00}(t) &= P\{0 \leq t < \hat{\tau}_1\} + P\{\hat{\tau}_1 + b_1 \leq t < \hat{\tau}_1 + b_1 + \hat{\tau}_2\} \\ &\quad + P\{\hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = 0\} \\ &= \bar{F}_\tau(t) + \int_0^t \bar{F}_\tau(t-x) d\left[F_\tau(x) * F_B(x)\right] \\ &\quad + P\{\hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = 0\}. \end{aligned} \tag{36}$$

Based on the model description given in Section 2, the third term of Eq. (36) can be further divided into the following two situations:

$$\begin{aligned} &P\{\hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = 0\} \\ &= P\{0 < \hat{\tau}_2 \leq Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = 0\} \\ &\quad + P\{\hat{\tau}_2 > Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = 0\} \\ &= \int_0^t \int_0^{t-x} p_{10}(t-x-y) \bar{F}_Y(y) d F_\tau(y) d\left[F_\tau(x) * F_B(x)\right] \\ &\quad + P\{\hat{\tau}_2 > Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = 0\}. \end{aligned} \tag{37}$$

If there is no arrival during the delay period Y, the server immediately takes a vacation after Y expires. According to the model description, it can be seen that after returning from vacation, the server will respond in three different ways depending on the number of waiting customers: (i)If the number of waiting customers is greater than or equal to the control threshold value N (≥ 1), the server starts its service immediately until the system becomes empty again; (ii)If there are less than N customers but at least one customer in the system, the server stays idle until N customers are accumulated in the system and then starts providing service; (iii)If there is no arrival during the vacation, the server takes another vacation immediately. Therefore, based on the number of customers arriving in the k-th vacation $V_k (k = 1, 2, \cdots)$, the second term of Eq. (37) can be divided into the following two cases:

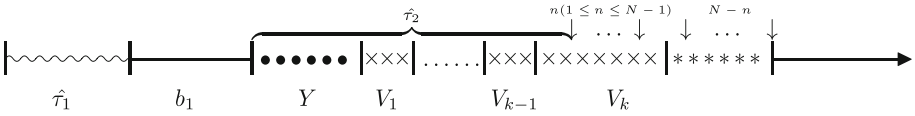

Fig. 3 There are n $(1 \le n \le N-1)$ arrivals during V_k

Case 1. There are n $(1 \le n \le N-1)$ customers arriving in the system during V_k, the server stays idle after V_k expires until N customers are accumulated in the system, and then immediately starts serving customers (see Fig. 3).

Case 2. There are more than or equal to the control threshold value N arrivals during V_k. The server starts providing service immediately after V_k expires (see Fig. 4).

Thus, the second term of Eq. (37) is given by

$$
\begin{aligned}
&P\{\hat{\tau}_2 > Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \le t; N(t) = 0\} \\
&= \sum_{k=1}^{\infty} \sum_{n=1}^{N-1} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{n-1} \le Y \\
&\quad + S_k < \hat{\tau}_2 + l_n; \hat{\tau}_1 + b_1 + Y + S_k + l_{N-n} \le t; N(t) = 0\} \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=N}^{\infty} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{n-1} \le Y \\
&\quad + S_k < \hat{\tau}_2 + l_n; \hat{\tau}_1 + b_1 + Y + S_{k-1} + V_k \le t; N(t) = 0\} \\
&= \sum_{k=1}^{\infty} \sum_{n=1}^{N-1} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} \int_0^{t-x-y-z-w} \int_{y+z}^{y+z+w} p_{N0}(t-x-y-z-w-u) \\
&\quad \times \frac{[\lambda(y+z+w-g)]^{n-1}}{(n-1)!} e^{-\lambda(y+z+w-g)} dF_\tau(g)\, dF_\tau^{(N-n)}(u)\, dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) \\
&\quad \times d[F_\tau(x) * F_B(x)] \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=N}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} \int_{y+z}^{y+z+w} p_{n0}(t-x-yz-w) \frac{[\lambda(y+z+w-u)]^{n-1}}{(n-1)!} \\
&\quad \times e^{-\lambda(y+z+w-u)} dF_\tau(u)\, dF_V(w)\, dV^{(k-1)}(z)\, dF_Y(y) d[F_\tau(x) * F_B(x)] \\
&= \sum_{k=1}^{\infty} \sum_{n=1}^{N-1} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} \int_0^{t-x-y-z-w} p_{N0}(t-x-y-z-w-u) \\
&\quad \times \frac{(\lambda w)^n}{n!} e^{-\lambda(y+z+w)} dF_\tau^{(N-n)}(u)\, dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) d[F_\tau(x) * F_B(x)] \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=N}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} p_{n0}(t-x-y-z-w) \frac{(\lambda w)^n}{n!} e^{-\lambda(y+z+w)} \\
&\quad \times dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) d[F_\tau(x) * F_B(x)].
\end{aligned}
\tag{38}
$$

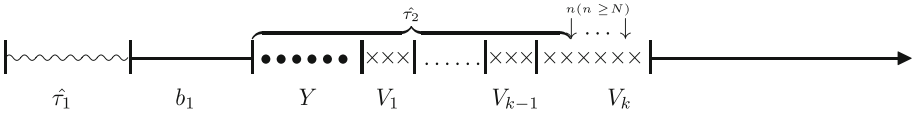

Fig. 4 There are more than or equal to N arrivals during V_k

Combining Eqs. (36)–(38) leads to

$$
\begin{aligned}
p_{00}(t) = & \bar{F}_\tau(t) + \int_0^t \bar{F}_\tau(t-x) d\left[F_\tau(x) * F_B(x)\right] \\
& + \int_0^t \int_0^{t-x} p_{10}(t-x-y)\bar{F}_Y(y) dF_\tau(y) d\left[F_\tau(x) * F_B(x)\right] \\
& + \sum_{k=1}^{\infty}\sum_{n=1}^{N-1} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} \int_0^{t-x-y-z-w} p_{N0}\left(t-x-y-z-w-u\right) \\
& \times \frac{(\lambda w)^n}{n!} e^{-\lambda(y+z+w)} dF_\tau^{(N-n)}(u)\, dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) d\left[F_\tau(x) * F_B(x)\right] \\
& + \sum_{k=1}^{\infty}\sum_{n=N}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} p_{n0}\left(t-x-y-z-w\right) \frac{(\lambda w)^n}{n!} e^{-\lambda(y+z+w)} \\
& \times dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) d\left[F_\tau(x) * F_B(x)\right].
\end{aligned}
\tag{39}
$$

For $i \geq 1$, similar to the discussion of Eq. (39), $p_{i0}(t)$ is given by

$$
\begin{aligned}
p_{i0}(t) = & \int_0^t \bar{F}_\tau(t-x) dF_B^{(i)}(x) + \int_0^t \int_0^{t-x} p_{10}(t-x-y)\bar{F}_Y(y) dF_\tau(y) dF_B^{(i)}(x) \\
& + \sum_{k=1}^{\infty}\sum_{n=1}^{N-1} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} \int_0^{t-x-y-z-w} p_{N0}\left(t-x-y-z-w-u\right) \\
& \times \frac{(\lambda w)^n}{n!} e^{-\lambda(y+z+w)} dF_\tau^{(N-n)}(u)\, dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) dF_B^{(i)}(x) \\
& + \sum_{k=1}^{\infty}\sum_{n=N}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} p_{n0}\left(t-x-y-z-w\right) \frac{(\lambda w)^n}{n!} e^{-\lambda(y+z+w)} \\
& \times dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) dF_B^{(i)}(x).
\end{aligned}
\tag{40}
$$

Taking the Laplace transform of Eqs. (39)-(40), we get

$$
\begin{aligned}
p_{00}^*(s) = & \frac{\bar{f}_\tau(s)\left[1 + f_\tau(s) f_B(s)\right]}{s} + p_{10}^*(s) f_\tau^2(s) b(s) \bar{f}_Y(s+\lambda) \\
& + p_{N0}^*(s) f_Y(s+\lambda) f_\tau(s) f_B(s) \frac{1}{\bar{f}_V(s+\lambda)} \sum_{n=1}^{N-1} \int_0^\infty e^{-(s+\lambda)t} \frac{(\lambda t)^n}{n!} f_\tau^{N-n}(s)\, dF_V(t) \\
& + f_Y(s+\lambda) f_\tau(s) f_B(s) \frac{1}{\bar{f}_V(s+\lambda)} \sum_{n=N}^{\infty} p_{n0}^*(s) \int_0^\infty e^{-(s+\lambda)t} \frac{(\lambda t)^n}{n!} dF_V(t),
\end{aligned}
\tag{41}
$$

$$
\begin{aligned}
p_{i0}^{*}(s) = & \frac{\bar{f}_{\tau}(s) f_{B}^{i}(s)}{s} + p_{10}^{*}(s) f_{B}^{i}(s) f_{\tau}(s) \bar{f}_{Y}(s+\lambda) \\
& + p_{N0}^{*}(s) f_{Y}(s+\lambda) f_{B}^{i}(s) \frac{1}{\bar{f}_{V}(s+\lambda)} \sum_{n=1}^{N-1} \int_{0}^{\infty} e^{-(s+\lambda)t} \frac{(\lambda t)^{n}}{n!} f_{\tau}^{N-n}(s) \, dF_{V}(t) \\
& + f_{Y}(s+\lambda) f_{B}^{i}(s) \frac{1}{\bar{f}_{V}(s+\lambda)} \sum_{n=N}^{\infty} p_{n0}^{*}(s) \int_{0}^{\infty} e^{-(s+\lambda)t} \frac{(\lambda t)^{n}}{n!} dF_{V}(t), \, i \geq 1.
\end{aligned} \tag{42}
$$

From Eqs. (41)–(42), the relation between $p_{00}^{*}(s)$ and $p_{i0}^{*}(s)$ can be expressed as

$$
p_{i0}^{*}(s) = \frac{f_{B}^{i-1}(s)}{f_{\tau}(s)} \left\{ p_{00}^{*}(s) - \frac{\bar{f}_{\tau}(s)}{s} \right\}, i = 1, 2, \cdots. \tag{43}
$$

Then, substituting Eq. (43) into (41) and solving Eq. (41) yields Eq. (1). Substituting Eq. (1) into (43) gives Eq. (2).

□

Proof of Theorem 2

Proof (1) For $j = 1, 2, \cdots, N-1$, the queue size is j at time point t only when that t is located in the server's non-busy period or the server's busy period with j customers. Using the probabilistic argument as in the analysis of $p_{00}(t)$ above, it yields

$$
\begin{aligned}
p_{0j}(t) = & P\{\hat{\tau}_1 \leq t < \hat{\tau}_1 + b_1; N(t) = j\} + P\{0 < \hat{\tau}_2 \leq Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = j\} \\
& + P\{\hat{\tau}_2 > Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = j\} \\
= & \int_{0}^{t} Q_j(t-x) dF_{\tau}(x) + \int_{0}^{t} \int_{0}^{t-x} p_{1j}(t-x-y) \bar{F}_{Y}(y) dF_{\tau}(y) d\left[F_{\tau}(x) * F_{B}(x)\right] \\
& + P\{\hat{\tau}_2 > Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \leq t; N(t) = j\},
\end{aligned} \tag{44}
$$

where the first term of Eq. (44) represents the probability that t is located in the first server's busy period with j customers, the second term of Eq. (44) represents the joint probability that t is located in after the second system idle period with j customers and one customer arrives during Y, and the third term Eq. (44) denotes the joint probability that there is no arrival during Y and t is located after the second system idle period with j customers.

According to the number of customers arriving during the k-th vacation V_k ($k = 1, 2, \cdots$), the third term of Eq. (44) can be decomposed into the following four cases:

Case 1: At least j customers arrive during V_k and the time point t is located in V_k with j customers (see Fig. 5).
Case 2: There are n ($1 \leq n \leq N-1$) arrivals and the time point t is located after the end of V_k but in the server's non-busy period with j customers (see Fig. 6).
Case 3: There are n ($1 \leq n \leq N-1$) arrivals and the time point t is located after the beginning of the second server's busy period with j customers (see Fig. 7).
Case 4: There are more than or equal to N arrivals during V_k and the time point t is located after the beginning of the second server's busy period with j customers (see Fig. 8).

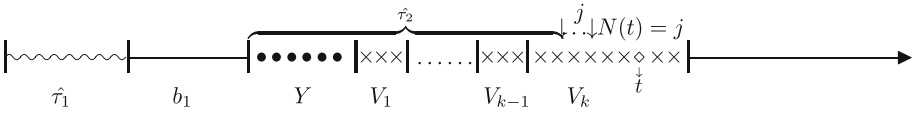

Fig. 5 The time point t is located in V_k with j customers

Then, the third term of Eq. (44) is given by

$$
\begin{aligned}
&P\{\hat{\tau}_2 > Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \le t; N(t) = j\}. \\
&= \sum_{k=1}^{\infty} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{j-1} \le Y \\
&\quad + S_k; \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_{j-1} \le t < \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_j\} \\
&\quad + \sum_{k=1}^{\infty} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{j-1} > Y \\
&\quad + S_k; \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_{j-1} \le t < \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_j\} \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=1}^{N-1} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{n-1} \le Y \\
&\quad + S_{k-1} + V_k < \hat{\tau}_2 + l_n; \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_{N-1} \le t; N(t) = j\} \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=N}^{\infty} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{n-1} \le Y \\
&\quad + S_{k-1} + V_k < \hat{\tau}_2 + l_n; \hat{\tau}_1 + b_1 + Y + S_{k-1} + V_k \le t; N(t) = j\} \\
&= \sum_{k=1}^{\infty} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \\
&\quad + l_{j-1} \le t < \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_j\} \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=1}^{N-1} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{n-1} \le Y \\
&\quad + S_{k-1} + V_k < \hat{\tau}_2 + l_n; \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_{N-1} \le t; N(t) = j\} \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=N}^{\infty} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{n-1} \le Y \\
&\quad + S_{k-1} + V_k < \hat{\tau}_2 + l_n; \hat{\tau}_1 + b_1 + Y + S_{k-1} + V_k \le t; N(t) = j\} \\
&= \sum_{k=1}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} \left[F_\tau^{(j-1)}(t-x-y-z-w) - F_\tau^{(j)}(t-x-y-z-w) \right] \\
&\quad \times \bar{F}_V(w)\, \bar{F}_\tau(y+z)\, dF_\tau(w)\, dF_V^{(k-1)}(z)\, dF_Y(y)\, d[F_\tau(x) * F_B(x)]
\end{aligned}
$$

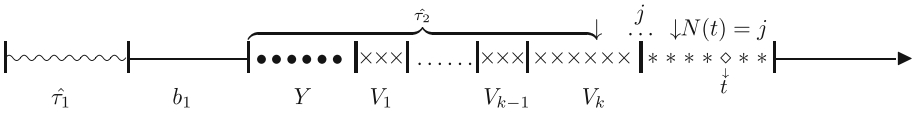

Fig. 6 The time point t is located after the end of V_k but in the server's non-busy period with j customers

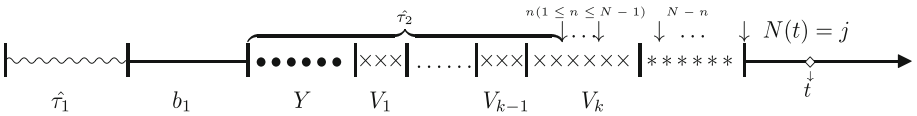

Fig. 7 The time point t is located after the beginning of the second server's busy period with j customers

$$
\begin{aligned}
&+\sum_{k=1}^{\infty}\sum_{n=1}^{N-1}\int_0^t\int_0^{t-x}\int_0^{t-x-y}\int_0^{t-x-y-z}\int_0^{t-x-y-z-w} p_{Nj}(t-x-y-z-w-u)\\
&\times\frac{(\lambda w)^n}{n!}e^{-\lambda w}\bar{F}_\tau(y+z)\,dF_\tau^{(N-n)}(u)\,dF_V(w)\,dF_V^{(k-1)}(z)\,dF_Y(y)\,d[F_\tau(x)*F_B(x)]\\
&+\sum_{k=1}^{\infty}\sum_{n=N}^{\infty}\int_0^t\int_0^{t-x}\int_0^{t-x-y}\int_0^{t-x-y-z} p_{nj}(t-x-y-z-w)\frac{(\lambda w)^n}{n!}e^{-\lambda w}\bar{F}_\tau(y+z)\\
&\times dF_V(w)\,dF_V^{(k-1)}(z)\,dF_Y(y)d[F_\tau(x)*F_B(x)]. \qquad (45)
\end{aligned}
$$

Substituting Eq. (45) into (44) leads to

$$
\begin{aligned}
p_{0j}(t) &= \int_0^t Q_j(t-x)dF_\tau(x)+\int_0^t\int_0^{t-x} p_{1j}(t-x-y)\bar{F}_Y(y)dF_\tau(y)d[F_\tau(x)*F_B(x)]\\
&+\sum_{k=1}^{\infty}\int_0^t\int_0^{t-x}\int_0^{t-x-y}\int_0^{t-x-y-z}\left[F_\tau^{(j-1)}(t-x-y-z-w)-F_\tau^{(j)}(t-x-y-z-w)\right]\\
&\times\bar{F}_V(w)\,\bar{F}_\tau(y+z)\,dF_\tau(w)\,dF_V^{(k-1)}(z)\,dF_Y(y)\,d[F_\tau(x)*F_B(x)]\\
&+\sum_{k=1}^{\infty}\sum_{n=1}^{N-1}\int_0^t\int_0^{t-x}\int_0^{t-x-y}\int_0^{t-x-y-z}\int_0^{t-x-y-z-w} p_{Nj}(t-x-y-z-w-u)\\
&\times\frac{(\lambda w)^n}{n!}e^{-\lambda w}\bar{F}_\tau(y+z)\,dF_\tau^{(N-n)}(u)\,dF_V(w)\,dF_V^{(k-1)}(z)\,dF_Y(y)\,d[F_\tau(x)*F_B(x)]\\
&+\sum_{k=1}^{\infty}\sum_{n=N}^{\infty}\int_0^t\int_0^{t-x}\int_0^{t-x-y}\int_0^{t-x-y-z} p_{nj}(t-x-y-z-w)\frac{(\lambda w)^n}{n!}e^{-\lambda w}\bar{F}_\tau(y+z)\\
&\times dF_V(w)\,dF_V^{(k-1)}(z)\,dF_Y(y)d[F_\tau(x)*F_B(x)]. \qquad (46)
\end{aligned}
$$

Similarly, we can obtain the expression of $p_{ij}(t)$ as follows,

$$
\begin{aligned}
p_{ij}(t) &= \sum_{k=1}^{i} Q_{j-i+k}(t)*F_B^{(k-1)}(t)+\int_0^t\int_0^{t-x} p_{1j}(t-x-y)\bar{F}_Y(y)dF_\tau(y)dF_B^{(i)}(x)\\
&+\sum_{k=1}^{\infty}\int_0^t\int_0^{t-x}\int_0^{t-x-y}\int_0^{t-x-y-z}\left[F_\tau^{(j-1)}(t-x-y-z-w)-F_\tau^{(j)}(t-x-y-z-w)\right]\\
&\times\bar{F}_V(w)\,\bar{F}_\tau(y+z)\,dF_\tau(w)dF_V^{(k-1)}(z)dF_Y(y)dF_B^{(i)}(x)
\end{aligned}
$$

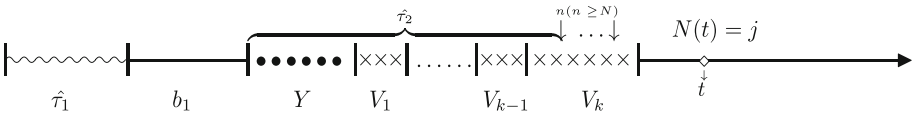

Fig. 8 The time point t is located after the beginning of the second server's busy period with j customers

$$
\begin{aligned}
&+\sum_{k=1}^{\infty}\sum_{n=1}^{N-1}\int_0^t\int_0^{t-x}\int_0^{t-x-y}\int_0^{t-x-y-z}\int_0^{t-x-y-z-w} p_{Nj}(t-x-y-z-w-u)\\
&\times\frac{(\lambda w)^n}{n!}e^{-\lambda w}\bar{F}_\tau(y+z)\,dF_\tau^{(N-n)}(u)\,dF_V(w)\,dF_V^{(k-1)}(z)\,dF_Y(y)\,dF_B^{(i)}(x)\\
&+\sum_{k=1}^{\infty}\sum_{n=N}^{\infty}\int_0^t\int_0^{t-x}\int_0^{t-x-y}\int_0^{t-x-y-z} p_{nj}(t-x-y-z-w)\frac{(\lambda w)^n}{n!}e^{-\lambda w}\bar{F}_\tau(y+z)\\
&\times dF_V(w)\,dF_V^{(k-1)}(z)\,dF_Y(y)dF_B^{(i)}(x), i\geq 1. \qquad (47)
\end{aligned}
$$

Taking the Laplace transform of Eqs. (46)–(47), $p_{0j}^*(s)$ and $p_{ij}^*(s)$ are given by

$$
\begin{aligned}
p_{0j}^*(s) &= f_\tau(s)q_j^*(s)+p_{1j}^*(s)f_\tau^2(s)f_B(s)\bar{f}_Y(s+\lambda)\\
&+f_Y(s+\lambda)f_\tau(s)f_B(s)\frac{f_\tau^j(s)-f_\tau^{j+1}(s)}{s}\\
&+p_{Nj}^*(s)f_Y(s+\lambda)f_\tau(s)f_B(s)\frac{1}{\bar{f}_V(s+\lambda)}\sum_{n=1}^{N-1}\int_0^\infty e^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}f_\tau^{N-n}(s)\,dF_V(t)\\
&+f_Y(s+\lambda)f_\tau(s)f_B(s)\frac{1}{\bar{f}_V(s+\lambda)}\sum_{n=N}^{\infty}p_{nj}^*(s)\int_0^\infty e^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}dF_V(t),
\end{aligned}
\qquad (48)
$$

$$
\begin{aligned}
p_{ij}^*(s) &= \sum_{k=1}^{i}q_{j-i+k}^*(s)f_B^{k-1}(s)+p_{1j}^*(s)f_B^i(s)f_B(s)f_Y(s+\lambda)\\
&+f_Y(s+\lambda)f_B^i(s)\frac{f_\tau^j(s)-f_\tau^{j+1}(s)}{s}\\
&+p_{Nj}^*(s)f_Y(s+\lambda)f_B^i(s)\frac{1}{\bar{f}_V(s+\lambda)}\sum_{n=1}^{N-1}\int_0^\infty e^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}f_\tau^{N-n}(s)\,dF_V(t)\\
&+f_Y(s+\lambda)f_B^i(s)\frac{1}{\bar{f}_V(s+\lambda)}\sum_{n=N}^{\infty}p_{nj}^*(s)\int_0^\infty e^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}dF_V(t), i\geq 1.
\end{aligned}
\qquad (49)
$$

Thus, the relation between $p_{0j}^*(s)$ and $p_{ij}^*(s)$ can be expressed as

$$
p_{ij}^*(s)=\sum_{k=1}^{i}q_{j-i+k}^*(s)f_B^{k-1}(s)+\frac{f_B^{i-1}(s)}{f_\tau(s)}\left[p_{0j}^*(s)-f_\tau(s)q_j^*(s)\right], i\geq 1. \qquad (50)
$$

Then, substituting Eq. (50) into (48) and solving Eq. (48) gets Eq. (3). Substituting Eq. (3) into (50) again leads to Eq. (4).

(2) For $j\geq N$, it is noted that the queue size is j at time point t only when that t is located in the server's busy period or the server's vacation with j customers. Applying the same arguments as in the above analysis, and similar to the decomposition of Eq. (46), we can get $p_{0j}(t)$ and $p_{ij}(t)$ as follows:

$$
\begin{aligned}
p_{0j}(t) &= P\{\hat{\tau}_1 \le t < \hat{\tau}_1 + b_1; N(t) = j\} + P\{0 < \hat{\tau}_2 \le Y; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \le t; N(t) = j\} \\
&\quad + \sum_{k=1}^{\infty} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_1 + b_1 + \hat{\tau}_2 \\
&\quad + l_{j-1} \le t < \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_j; \hat{\tau}_1 + b_1 + Y + S_{k-1} + V_k > t\} \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=1}^{N-1} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{n-1} \le Y \\
&\quad + S_{k-1} + V_k < \hat{\tau}_2 + l_n; \hat{\tau}_1 + b_1 + \hat{\tau}_2 + l_{N-1} \le t; N(t) = j\} \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=N}^{\infty} P\{\hat{\tau}_2 > Y; Y + S_{k-1} \le \hat{\tau}_2 < Y + S_k; \hat{\tau}_2 + l_{n-1} \le Y \\
&\quad + S_{k-1} + V_k < \hat{\tau}_2 + l_n; \hat{\tau}_1 + b_1 + Y + S_{k-1} + V_k \le t; N(t) = j\} \\
&= \int_0^t Q_j(t-x) dF_\tau(x) + \int_0^t \int_0^{t-x} p_{1j}(t-x-y) \bar{F}_Y(y) dF_\tau(y) d\left[F_\tau(x) * F_B(x)\right] \\
&\quad + \sum_{k=1}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \left[F_\tau^{(j)}(t-x-y-z) - F_\tau^{(j+1)}(t-x-y-z)\right] \bar{F}_\tau(y+z) \\
&\quad \times \bar{F}_V(t-x-y-z)\, dF_V^{(k-1)}(z)\, dF_Y(y)\, d\left[F_\tau(x) * F_B(x)\right] \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=1}^{N-1} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} \int_0^{t-x-y-z-w} p_{Nj}(t-x-y-z-w-u) \\
&\quad \times \frac{(\lambda w)^n}{n!} e^{-\lambda w} \bar{F}_\tau(y+z)\, dF_\tau^{(N-n)}(u)\, dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y)\, d\left[F_\tau(x) * F_B(x)\right] \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=N}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} p_{nj}(t-x-y-z-w) \frac{(\lambda w)^n}{n!} e^{-\lambda w} \bar{F}_\tau(y+z) \\
&\quad \times dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) d\left[F_\tau(x) * F_B(x)\right].
\end{aligned} \tag{51}
$$

$$
\begin{aligned}
p_{ij}(t) &= \sum_{k=1}^{i} Q_{j-i+k}(t) * F_B^{(k-1)}(t) + \int_0^t \int_0^{t-x} p_{1j}(t-x-y) \bar{F}_Y(y) dF_\tau(y) dF_B^{(i)}(x) \\
&\quad + \sum_{k=1}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \left[F_\tau^{(j)}(t-x-y-z) - F_\tau^{(j+1)}(t-x-y-z)\right] \bar{F}_\tau(y+z) \\
&\quad \times \bar{F}_V(t-x-y-z)\, dF_V^{(k-1)}(z)\, dF_Y(y)\, dF_B^{(i)}(x) \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=1}^{N-1} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} \int_0^{t-x-y-z-w} p_{Nj}(t-x-y-z-w-u) \\
&\quad \times \frac{(\lambda w)^n}{n!} e^{-\lambda w} \bar{F}_\tau(y+z)\, dF_\tau^{(N-n)}(u)\, dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y)\, dF_B^{(i)}(x) \\
&\quad + \sum_{k=1}^{\infty} \sum_{n=N}^{\infty} \int_0^t \int_0^{t-x} \int_0^{t-x-y} \int_0^{t-x-y-z} p_{nj}(t-x-y-z-w) \frac{(\lambda w)^n}{n!} e^{-\lambda w} \bar{F}(y+z) \\
&\quad \times dF_V(w)\, dF_V^{(k-1)}(z)\, dF_Y(y) dF_B^{(i)}(x), i \ge 1.
\end{aligned} \tag{52}
$$

The Laplace transform of Eqs. (51) and (52) are given by

$$\begin{aligned} p_{0j}^*(s) = & f_\tau(s)q_j^*(s) + +p_{1j}^*(s) f_\tau^2(s) f_B(s) \bar{f}_Y(s+\lambda) \\ & + f_Y(s+\lambda) f_\tau(s) f_B(s) \frac{1}{\bar{f}_V(s+\lambda)} \int_0^\infty e^{-st} \bar{F}_V(t) \left[F_\tau^{(j)}(t) - F_\tau^{(j+1)}(t) \right] dt \\ & + p_{Nj}^*(s) f_Y(s+\lambda) f_\tau(s) f_B(s) \frac{1}{\bar{f}_V(s+\lambda)} \sum_{n=1}^{N-1} \int_0^\infty e^{-(s+\lambda)t} \frac{(\lambda t)^n}{n!} f_\tau^{N-n}(s) \, dF_V(t) \\ & + f_Y(s+\lambda) f_\tau(s) f_B(s) \frac{1}{\bar{f}_V(s+\lambda)} \sum_{n=N}^{\infty} p_{nj}^*(s) \int_0^\infty e^{-(s+\lambda)t} \frac{(\lambda t)^n}{n!} dF_V(t), \end{aligned} \tag{53}$$

$$\begin{aligned} p_{ij}^*(s) = & \sum_{k=1}^{i} q_{j-i+k}^*(s) f_B^{k-1}(s) + p_{1j}^*(s) f_B^i(s) f_\tau(s) \bar{f}_Y(s+\lambda) \\ & + f_Y(s+\lambda) f_B^i(s) \frac{1}{\bar{f}_V(s+\lambda)} \int_0^\infty e^{-st} \bar{V}(t) \left[F_\tau^{(j)}(t) - F_\tau^{(j+1)}(t) \right] dt \\ & + p_{Nj}^*(s) f_Y(s+\lambda) f_B^i(s) \frac{1}{\bar{f}_V(s+\lambda)} \sum_{n=1}^{N-1} \int_0^\infty e^{-(s+\lambda)t} \frac{(\lambda t)^n}{n!} f_\tau^{N-n}(s) \, dF_V(t) \\ & + f_Y(s+\lambda) f_B^i(s) \frac{1}{\bar{f}_V(s+\lambda)} \sum_{n=N}^{\infty} p_{nj}^*(s) \int_0^\infty e^{-(s+\lambda)t} \frac{(\lambda t)^n}{n!} dF_V(t). \end{aligned} \tag{54}$$

For $j \geq N$, the relation between $p_{0j}^*(s)$ and $p_{ij}^*(s)$ is given by

$$p_{ij}^*(s) = \sum_{k=1}^{i} q_{j-i+k}^*(s) f_B^{k-1}(s) + \frac{f_B^{i-1}(s)}{f_\tau(s)} \left[p_{0j}^*(s) - f_\tau(s) q_j^*(s) \right], i \geq 1. \tag{55}$$

Substituting Eq. (55) into (53) leads to Eq. (5), and then substituting Eq. (5) into (55) arrives at Eq. (6).

Acknowledgements The authors thank the Editor-In-Chief and the anonymous referees for their constructive comments and suggestions which have improved the quality of this paper.

Author Contributions The contributions of each author are as follows. (1) Yaxing He: Conceptualization, Writing-original draft, Writing-review & editing, Investigation, Formal analysis. (2) Yinghui Tang: Conceptualization, Methodology, Funding acquisition. (3) Miaomiao Yu: Validation, Writing-review & editing. (4) Wenqing Wu: Software, Formal analysis, Data curation.

Funding This research is supported by the National Natural Science Foundation of China under Grant No.71571127 and the Specialized Project for Subject Construction of Sichuan Normal University under Grant XKZX2021-04.

Availability of Data and Material Not applicable.

Declarations

Competing Interests The authors declare no competing interests.

References

Agarwal RP, Dshalalow JH (2005) New fluctuation analysis of D-policy bulk queues with multiple vacations. Math Comput Modelling 41(2–3):253–269

Artalejo JR (2002) A note on the optimality of the N- and D-policies for the $M/G/1$ queue. Oper Res Lett 30(6):375–376

Ayyappan G, Karpagam S (2019) Analysis of a bulk queue with unreliable server, immediate feedback, N-policy, Bernoulli schedule multiple vacation and stand-by server. Ain Shams Eng J 10(4):873–880

Ayyappan G, Nirmala M (2020) Analysis of bulk queue with unreliable service station, second optional repair, N-policy multiple vacation, loss and immediate feedback in production system. Int J Comput Sci Math 12(4):339–349

Baba Y (2004) Analysis of a $GI/M/1$ queue with multiple working vacations. Oper Res Lett 33(2):201–209

Balachandran KR (1973) Control policies for a single server system. Manage Sci 19(9):1013–1018

Cohen JW (1982) The single server queue. North-Holland New York

Doshi B (1986) Queueing Systems with Vacations - A Survey. Queueing Syst 1:29–66

Heyman DP (1977) The T-policy for the $M/G/1$ queue. Manag Sci 23(7):775–778

Hur S, Kim J, Kang C (2003) An analysis of the $M/G/1$ system with N and T policy. Appl Math Model 27(8):665–675

Kella O (1989) The threshold policy in the $M/G/1$ queue with server vacations. Nav Res Logist 36(1):111–123

Kuang XY, Tang YH, Yu MM, Wu WQ (2022) Performance analysis of an M/G/1 queue with bi-level randomized (p, $N1$, $N2$)-policy. RAIRO-Oper Res 56(1):395–414

Lan SJ, Tang YH (2019) An N-policy discrete-time $Geo/G/1$ queue with modified multiple server vacations and Bernoulli feedback. RAIRO-Oper Res 53(2):367–387

Lee SS, Lee HW, Yoon SH, Chae KC (1994a) Batch arrival queue with N-policy and single vacation. Computer Ops Res 22(2):173–189

Lee HW, Seo WJ (2008) The performance of the $M/G/1$ queue under the dyadic min(N, D)-policy and its cost optimization. Perform Eval 65(10):742–758

Lee HW, Seo WJ, Lee SW, Jeon JW (2010) Analysis of the $MAP/G/1$ queue under the min(N, D)-policy. Stoch Models 26(1):98–123

Lee H, Lee J, Park J, Chae KC (1994b) Analysis of the $M^X/G/1$ queue by N-policy and multiple vacations. J Appl Probab 31(2):476–496

Levy Y, Yechiali U (1975) Utilization of the idle time in an $M/G/1$ queueing system. Manage Sci 22(2):202–211

Li J, Tian N (2010) Performance analysis of a $GI/M/1$ queue with single working vacation. App Math Comput 217(10):4960–4971

Luo CY, Tang YH, Chao BS, Xiang KL (2013) Performance analysis of a discrete-time $Geo/G/1$ queue with randomized vacations and at most J vacations. Appl Math Model 37(9):6489–6504

Luo L, Tang YH, Yu MM, Wu WQ (2023) Optimal control policy of $M/G/1$ queueing system with delayed randomized multiple vacations under the modified min(N, D)-policy control. J Oper Res Soc China 11:857–874

Sethi R, Jain MR, Meena K, Garg D (2020) Cost optimization and ANFIS computing of an unreliable $M/M/1$ queueing system with customers impatience under N-Policy. Int J Appl Comput Math 6(1):97–108

Tang YH, Yun X, Huang SJ (2008) Discrete-time $Geo^X/G/1$ queue with unreliable server and multiple adaptive delayed vacation. J Comput Appl Math 220(3):439–455

Tian NS, Zhang ZG (2008) Vacation queueing models-theory and applications. Springer, New York

Wang KH, Wang TY, Pearn WL (2006) Optimal control of the N policy $M/G/1$ queueing system with server breakdowns and general startup times. Appl Math Model 31(10):2199–2212

Wei YY, Tang YH, Yu MM (2020) Recursive solution of queue length distribution for $Geo/G/1$ queue with delayed min(N, D)-policy. J Syst Sci Inf 8(4):367–386

Wu WQ, Tang YH, Yu MM (2014) Analysis of an $M/G/1$ queue with multiple vacations, N-policy, unreliable service station and repair facility failures. Int J Sup Oper Manag 1(1):1–19

Yadin M, Naor P (1963) Queueing systems with a removable service station. Oper Res Q 14:393–405