STRATEGIC BEHAVIOR AND OPTIMAL STRATEGIES IN AN M/G/1 QUEUE WITH BERNOULLI VACATIONS

SHENG ZHU
1. Department of Mathematics, Beijing Jiaotong University
Beijing, 100044, China
2. School of Mathematics & Information Science, Henan Polytechnic University
Jiaozuo, 454003, China
JINTING WANG*
Department of Mathematics, Beijing Jiaotong University
Beijing, 100044, China

(Communicated by Shoji Kasahara)

Abstract. This paper considers an unobservable M/G/1 queue with Bernoulli vacations in which the server begins a vacation when the system is empty or upon completing a service. In the latter case, the server takes a vacation with probability p or serves the next customer, if any, with probability 1−p. We first give the steady-state equations and some performance measures, and then study the customer strategic behavior and obtain customers’ Nash equilibrium strategies. From the viewpoint of the social planner, we derive the socially optimal joining probability, the socially optimal vacation probability and the socially optimal vacation rate. The socially optimal joining probability is found not greater than the equilibrium probability. In addition, if the vacation scheme does not incur any cost, the socially optimal decision is that the server does not take either a Bernoulli vacation or the normal vacation. On the other hand, if the server incurs the costs due to the underlying loss and the technology upgrade, proper vacations are beneficial to the social welfare maximization. Finally, sensitivity analysis is also performed to explore the effect of different parameters, and some managerial insights are provided for the social planner.

1. Introduction. Queueing systems with Bernoulli vacation have been widely used in various applications such as computer networks, inventory control, equipment maintenance, etc. A real-life situation is that the managers of email service centers always try their best to keep the email system working efficiently and effectively. However, virus attacks have an adverse effect on the email service system and virus scan is an essential and effective maintenance method for the email servers. These virus scans can be regarded as system vacations. That means, if the system is in the working state, the email server handles emails one by one. And if the system is in the vacation state, the email server does not handle any email, but scans itself to check whether the system is infected with viruses or not. After completing

2010 Mathematics Subject Classification. Primary: 60K25; Secondary: 90B22.
Key words and phrases. M/G/1 queue, Nash equilibrium, socially optimal solution, overall welfare, bernoulli vacation.
This work is supported in part by the National Natural Science Foundation of China (Grant nos. 71571014, 71390334.)
* Corresponding author: Jinting Wang.
the virus scan and clearing the viruses, the system enters the working state if the system is nonempty, otherwise continues the next vacation. We can formulate the above problem by a queueing model with vacations. On the other hand, if virus scans are performed only when the server is idle, the system may face a high risk being attacked due to long busy period. A feasible approach to avoid this risk is to execute scans randomly after a service completion. That is, once the server completes a service, it handles the next email with probability $1 - p$ if the system is nonempty, or performs the virus scan with probability $p$, where $p$ is called vacation probability. The state above can be modeled as a queueing model with Bernoulli vacations.

Another interesting example is initial public offerings (IPO), which are mostly used by companies to raise the expansion of capital. In China, companies first submit IPO applications, and then wait for the reviews of China Securities Regulatory Commission (CSRC). Companies are served according to the first-come-first-served (FCFS) service discipline. After passing the reviews of the CSRC (service completion), the companies can sell their shares to the public. However, the CSRC decides whether to continue the IPO audit based on the securities market environment. If the market is weak, the CSRC suspends their reviews until the financial market becomes stable. That means, after a service completion, the CSRC may handle the next IPO application with probability $1 - p$ if the system is nonempty, or suspend their reviews with probability $p$ due to a weak stock market. Here $p$ is also called vacation probability. In the real world, there always are a long queue waiting for reviews. Companies need to decide whether to raise capital through IPO or not. This scenario can also be modeled as a queueing model with Bernoulli vacations.

There are several issues deserving further study. The first issue is that customers (the persons sending emails to the email service center, companies submitting IPO applications, etc.) need to decide whether to join the queue or not. For example, companies needing to raise capital must decide whether to finance through IPO or through other ways such as issuing bonds. In this sense, customers’ strategic behavior is worth of being studied. Second, in the real life, government (CSRC) or other non-profitable organization (email service center operated by education department) always try their best to maximize socially interests and therefore, the socially optimal joining probability needs to be considered. Third, regarding the system’s performance, some measures such as customers’ joining probability, vacation probability, and vacation rate have key effects on social interests. We need to discuss optimal controls from the viewpoint of the social planner. On one hand, individual customer always maximizes her own benefit. On the other hand, the social planner has the need to realize optimal social interests. To reduce the gap, a feasible management method needs to be considered. Finally, a sensitivity analysis should be carried out to illustrate the effect of different parameters.

2. Literature review. Recently, Liu and Wang [18] studied the strategic joining behavior of customers in a single-server Markovian queueing system with Bernoulli vacation. The Nash equilibrium strategies in the fully observable case and the unobservable cases (i.e., the almost unobservable case and the fully unobservable case) have been investigated. Based on the available system information, customers decide whether to join or balk the system. Such a game-theoretic analysis of queueing systems with decentralized behavior of customers has been paid considerable attention during the last several decades. To reflect customers’ desires for service and unwillingness to wait, some suitable reward-cost structures are imposed to the sys-
tems. Arriving customers are allowed to decide whether to join the system or not. Based on different levels of information of the system on hand, they maximize their own utility. These ideas go back to the pioneering work of Naor [20]. Interested readers are referred to the books of Hassin and Haviv [10] and Hassin [11] for more details.

In the literature on the game theoretic analysis of queueing systems, a significant part is devoted to Markovian models (see, for example, Burnetas and Economou [2], Burnetas and Apostolos [3], Guo and Hassin [8, 9], Shi and Lian [22], Wang and Zhang [24, 25], Zhang et al. [27], among others). In contrast to the Markovian queueing models, the analysis of the non-Markovian models seems challenging since some properties do not hold any more. In the standard M/G/1 queue, customers affect each other not only by directly increasing the length of the queue but also by transmitting relevant information on the residual service time. Few works concerned with the M/G/1 cases. Altman and Hassin [1] studied an observable M/G/1 queue in which service times follow a Bernoulli distribution and customers decide whether or not to join the queue after observing the queue length. Under certain assumptions on the arrival rate and the reward-cost structure, it was shown there exists an equilibrium strategy profile of the delay threshold policy type. Haviv and Kerner [12] considered the partial observable M/G/1 queue. Kerner [14] presented a recursive algorithm for computing the Nash equilibrium strategy in the standard M/G/1 queue. Economou et al. [5] investigated equilibrium and social optimization in the M/G/1 queue with multiple vacations under two different information assumptions. Recently, Manou et al. [19] considered an M/G/1 type transportation station system with strategic customers. Customers’ strategic behavior was studied and their symmetric Nash equilibrium strategies under two levels of information were derived. To the best authors’ knowledge, there is no work that investigates the M/G/1 queue with Bernoulli scheduling.

In the present paper we consider the equilibrium analysis of the M/G/1 queue with Bernoulli scheduling. The queueing system with Bernoulli scheduling is first introduced by Keilson and Servi [13]. They mainly demonstrate the existence of a decomposition of the waiting-time distribution for this schedule and hence for the exhaustive schedule and single service schedule as by-products. Follow-up works include Servi [21], Ghaifir and Silio [6], Gray et al. [7], Kumar et al. [15, 16], to mention a few. These works studied various queueing systems with Bernoulli scheduling for the queueing performance. However, the equilibrium customer behavior and social optimal decision subject to Bernoulli vacations are seldom studied. Different from Liu and Wang [18], we assume in this paper that the server time follows a general distribution which makes the model become more realistic. We aim to derive the customers’ equilibrium strategic behavior in M/G/1 queueing system with Bernoulli vacations. From the perspective of the planner, we obtain the socially optimal joining probabilities. We find that the socially optimal joining probabilities are not greater than equilibrium joining probabilities. Based on this observation, we present a feasible method to maximize the social welfare. We also derive the socially optimal vacation probability and the socially optimal vacation rate. It is observed that ‘the server does not take either a Bernoulli vacation or the normal vacation’ is a socially optimal decision if the server incurs no cost. On the other hand, if the server incurs the costs due to the underlying loss and the technology upgrade, some vacations are beneficial to the social welfare maximization.
This reminder of this paper is organized as follows. Section 3 presents the model description. In Section 4, we consider the steady-state equations and generating functions. Section 5 derives the equilibrium strategies for a fully unobservable $\text{M}/\text{G}/1$ queue with Bernoulli vacations. The corresponding mixed Nash equilibrium strategies are obtained. Section 6 studies the optimal control for the social planner, and we obtain the socially optimal joining probability, the socially optimal vacation probability and the socially optimal vacation rate. The management of the social planner is explored in Section 7. Finally, some conclusions are given in Section 8.

3. Model description. We consider an unobservable $\text{M}/\text{G}/1$ queue with Bernoulli vacations. Customers arrive according to a Poisson process with intensity $\lambda$. New customers line up at the server based on the order of their arrivals and are served according to the first-come-first-served (FCFS) service discipline. The service times are independent, identically distributed with a common distribution function $B(\cdot)$ and finite first two moments: $\beta_k, k = 1, 2$. The corresponding probability density function $b(x)$ of the service times is given by

$$b(x) = \mu(x) \exp\{-\int_0^x \mu(t)dt\},$$

where $\mu(x)$ is the service completion rate function

$$\mu(x) = \frac{b(x)}{1 - B(x)}.$$

Let $B^*(s)$ be the Laplace-Stieltjes Transform (LST) of $B(x)$ and $\rho = \lambda \beta_1$ be the system load. The server begins a vacation period if the system becomes empty; otherwise, the server begins a vacation period with probability $p$ and serves the next customer with probability $1 - p$ upon completion of a service, where $p$ is called vacation probability in the paper. After a vacation period, the server immediately serves the head customer waiting in the queue if the system is nonempty or continues to take a new vacation if there is no customer in the system. The vacation times of the server are exponentially distributed with vacation rate $\theta$. Furthermore, we assume that the interarrival of customers, the service times and the vacation times are mutually independent. Note that after each service completion vacations may occur no matter whether the system is empty. So Bernoulli vacations are quite different from ordinary vacation scheme which does not occur if the system is nonempty (see Liou [17], Yue et al. [26]).

4. The steady-state equations and generating functions. In this section, we first derive the steady-state condition that the system is stable. In the steady-state situation, we get the partial generating functions of the joint distribution of the server state and queue length, and then the probability that the server is busy (on the Bernoulli vacation) is also obtained. Assume that arriving customers follow a common strategy $q$, i.e., the arriving customers enter the system with probability $q$. The following theorem presents the stability condition of the considered $\text{M}/\text{G}/1$ queueing system with Bernoulli vacations.

**Theorem 4.1.** Consider an $\text{M}/\text{G}/1$ queueing system with Bernoulli vacations, in which arriving customers follow a common strategy $q$. If $\theta(\lambda q \beta_1 - 1) + \lambda qp < 0$, that is, $q < \frac{\theta}{\theta \lambda \beta_1 + \lambda p}$, the system is stable.
Proof. The proof is similar to Theorem 3.1 in the work of Wang et al. [23]. Let $\Upsilon_n$ be the number of customers in the system at the instant after $n$-th departure, and $\Psi_k$ denotes $E(\Upsilon_{n+1} - \Upsilon_n | \Upsilon_n = k), k = 0, 1, 2, \ldots$. After $n$-th departure, the server takes a vacation with $p$ or serves the next customer, if any, with $1 - p$. If the server serves the next customer, the mean number of customers in the system at the instant after $n + 1$-th departure equals $k - 1 + \lambda q \beta_1$; if the server takes a vacation, the mean number of customers is $k - 1 + \lambda q (\beta_1 + \frac{1}{\theta})$. Then $\Psi_k$ can be computed from

$$\Psi_k = (1 - p)(k - 1 + \lambda q \beta_1 - k) + p(k - 1 + \lambda q (\beta_1 + \frac{1}{\theta}) - k) = \frac{\theta (\lambda q \beta_1 - 1) + \lambda q p}{\theta}. $$

If $\theta (\lambda q \beta_1 - 1) + \lambda q p < 0$, we have

$$|\Psi_k| = \frac{\theta (\lambda q \beta_1 - 1) + \lambda q p}{\theta} < \infty, \quad (3)$$

and

$$\limsup_{k \to \infty} \Psi_k = \frac{\theta (\lambda q \beta_1 - 1) + \lambda q p}{\theta} < 0. \quad (4)$$

From Eqs. (3)-(4), we know that $\{\Upsilon_n, n \geq 1\}$ is positive recurrent. Clearly, $\{\Upsilon_n, n \geq 1\}$ is irreducible and aperiodic. So $\{\Upsilon_n, n \geq 1\}$ is ergodic. According to the results in the work of Cinlar [4], the system is stable. This completes the proof. \hfill $\square$

Let $N(t)$ be the number of customers in the system at time $t$, and $X(t)$ be the elapsed service time of the customer being served at time $t$. The stochastic process $\{(N(t), X(t)), t \geq 0\}$ is a Markovian process with state space $\{(n, x) | n \in \mathbb{N}, 0 \leq x < +\infty\}$.

1. $P_n(t, x)dx$ is the joint probability that at time $t$ there are $n$ customers in the system and a customer is being served with elapsed service time between $x$ and $x + dx$.

2. $Q_n(t)$ is the probability that at time $t$ the server is on Bernoulli vacation state and there are $n$ customers in the system.

By considering transitions of the process between time $t$ and $t + \Delta t$ and letting $\Delta t \to 0$, we have

$$\left[ \frac{\partial}{\partial x} + \frac{\partial}{\partial t} \right] P_n(t, x) = - (\lambda q + \mu(x)) P_n(t, x) + \lambda q P_{n-1}(t, x), \quad n = 2, 3, \ldots; \quad (5)$$

$$\left[ \frac{\partial}{\partial x} + \frac{\partial}{\partial t} \right] P_1(t, x) = - (\lambda q + \mu(x)) P_1(t, x); \quad (6)$$

$$\frac{d}{dt} Q_n(t) = - (\lambda q + \theta) Q_n(t) + \int_0^\infty P_{n+1}(t, x) \mu(x) dx + \lambda q Q_{n-1}(t), \quad n = 1, 2, \ldots; \quad (7)$$

$$\frac{d}{dt} Q_0(t) = - \lambda q Q_0(t) + \int_0^\infty P_1(t, x) \mu(x) dx; \quad (8)$$

$$P_n(t, 0) = \theta Q_n(t) + \int_0^\infty P_{n+1}(t, x) \mu(x)(1 - p) dx, \quad n = 1, 2, \ldots. \quad (9)$$

We assume that the stable condition of the system is satisfied and set $Q_n = \lim_{t \to \infty} Q_n(t), P_n(x) = \lim_{t \to \infty} Q_n(t, x)$. From (5)-(9) we obtain the following equations:
Substituting (13) into (18), we get

\[ \frac{d}{dx} + \lambda q + \mu(x) P_n(x) = \lambda q P_{n-1}(x), \quad n = 2, 3, \ldots; \] (10)

\[ \frac{d}{dx} + \lambda q + \mu(x) P_1(x) = 0; \] (11)

\[
(\lambda q + \theta) Q_n = \int_0^\infty P_{n+1}(x) \mu(x) p dx + \lambda q Q_{n-1}, \quad n = 1, 2, \ldots; \] (12)

\[ \lambda q Q_0 = \int_0^\infty P_1(x) \mu(x) dx; \] (13)

\[ P_n(0) = \theta Q_n + \int_0^\infty P_{n+1}(x) \mu(x)(1-p) dx, \quad n = 1, 2, \ldots. \] (14)

The normalizing equation is given as

\[
\sum_{n=0}^\infty Q_n + \sum_{n=1}^\infty \int_0^\infty P_n(x) dx = 1. \] (15)

To solve the system of equations (10)-(14), we define the following generating functions:

\[ Q(z) = \sum_{n=0}^\infty Q_n z^n, \quad P(z, x) = \sum_{n=1}^\infty P_n(x) z^n, \]

where \(|z| \leq 1\). Multiplying equations (10), (12), (14) by \(z^n\) and summing over \(n\), we obtain the following basic equations after some algebraic manipulations:

\[
\left[\frac{d}{dx} + \lambda q + \mu(x)\right] P(z, x) = \lambda q z P(z, x), \] (16)

\[ (\lambda q + \theta)(Q(z) - Q_0) = \frac{1}{z} \int_0^\infty \mu(x) p(P(z, x) - P_1(x)z) dx + \lambda q z Q(z), \] (17)

\[ P(z, 0) = \theta Q(z) - \theta Q_0 + \frac{1}{z} \int_0^\infty \mu(x)(1-p)(P(z, x) - P_1(x)z) dx. \] (18)

From the system of equations (16)-(18), the partial generating functions \(Q(z), P(z, x)\) can be obtained. We summarize the results in the following theorem.

**Theorem 4.2.** In \(M/G/1\) queue with Bernoulli vacations, the joint distribution of the server state and queue length has partial generating functions:

\[ Q(z) = \frac{[B^*(\lambda q(1-z)) - z][\theta(1 - \lambda q \beta_1) - \lambda q p]}{\theta[B^*(\lambda q(1-z)) - z] - \lambda q(1-z)[z - (1-p)B^*(\lambda q(1-z))]}, \] (19)

\[ P(z, x) = \frac{\lambda q(1-z)z[\theta(1 - \lambda q \beta_1) - \lambda q p]e^{-\lambda q(1-z)x\overline{B}(x)}}{\theta[B^*(\lambda q(1-z)) - z] - \lambda q(1-z)[z - (1-p)B^*(\lambda q(1-z))]}, \] (20)

**Proof.** Solving the differential equations (11) and (16), we obtain

\[ P_1(x) = P_1(0)e^{-\lambda q x\overline{B}(x)}, \] (21)

\[ P(z, x) = P(z, 0)e^{-\lambda q(1-z)x\overline{B}(x)}. \] (22)

Substituting (13) into (18), we get

\[ P(z, 0) = \theta Q(z) + \frac{1-p}{z} \int_0^\infty \mu(x) P(z, x) dx - [\theta + \lambda q(1-p)]Q_0. \] (23)

From (22), (23) can be rewritten as

\[ P(z, 0) = \frac{z}{z - (1-p)B^*(\lambda q(1-z))}[\theta Q(z) - (\theta + \lambda q(1-p))Q_0]. \] (24)
Plugging (13) in (17) and summing (23), we have
\[ P(z,0) = \frac{\lambda q(1-z)z}{B^*(\lambda q(1-z)) - z}Q(z). \] (25)

From (24) and (25), we can get the following equation after some algebraic manipulations:
\[ Q(z) = \frac{[\theta + \lambda q(1-p)][B^*(\lambda q(1-z)) - z]}{\theta[B^*(\lambda q(1-z)) - z] - \lambda q(1-z)[z - (1-p)B^*(\lambda q(1-z))]}Q_0. \] (26)

Then plugging (26) in (25) yields
\[ P(z,0) = \frac{[\theta + \lambda q(1-p)]\lambda q(1-z)z}{\theta[B^*(\lambda q(1-z)) - z] - \lambda q(1-z)[z - (1-p)B^*(\lambda q(1-z))]}Q_0. \] (27)

Letting \( z \to 1 \) in (26) and (27), we can obtain
\[ Q(1) = \frac{[\theta + \lambda q(1-p)](\lambda q\beta_1 - 1)}{\theta(\lambda q\beta_1 - 1) + \lambda qp}Q_0, \] (28)
\[ P(1,0) = \frac{-\lambda q[\theta + \lambda q(1-p)]}{\theta(\lambda q\beta_1 - 1) + \lambda qp}Q_0. \] (29)

And then by (22), we can get
\[ P(1,x) = P(1,0)B(x) = \frac{-\lambda q[\theta + \lambda q(1-p)]}{\theta(\lambda q\beta_1 - 1) + \lambda qp}B(x)Q_0. \] (30)

From the total probability \( Q(1) + \int_0^\infty P(1,x)dx = 1 \), we obtain
\[ Q_0 = \frac{\theta(1 - \lambda q\beta_1) - \lambda qp}{\theta + \lambda q(1-p)}. \] (31)

(19) can be immediately obtained by substituting (31) into (26). From (22), (27) and (31), (20) is also derived. This completes the proof. \( \square \)

‘Busy’ and ‘vacation’ are two different states of the system. The probabilities that the server is in these two different states embody the efficiency of system operation. For example, if the server takes vacations with larger probability, more customers will be accumulated in the system, and the expected waiting time of each customer will become longer. The following lemma gives the corresponding probabilities.

**Lemma 4.3.** In steady-state situation, the following results hold in the \( M/G/1 \) queue with Bernoulli vacations.

1. The probability that the server is on Bernoulli vacation is \( I = 1 - \lambda q\beta_1 \);
2. The probability that the server is busy is \( B = \lambda q\beta_1 \).

**Proof.** Note that \( I = \lim_{z \to 1} Q(z) \) and \( B = \int_0^\infty \lim_{z \to 1} P(z,x)dx \). According to Theorem 4.2, we obtain
\[ I = \lim_{z \to 1} \frac{[B^*(\lambda q(1-z)) - z][\theta(1 - \lambda q\beta_1) - \lambda qp]}{\theta[B^*(\lambda q(1-z)) - z] - \lambda q(1-z)[z - (1-p)B^*(\lambda q(1-z))]}Q_0 \]
and
\[ B = \int_0^\infty \frac{\lambda q(1-z)z[\theta(1 - \lambda q\beta_1) - \lambda qp]e^{-\lambda q(1-z)x}B(x)}{\theta[B^*(\lambda q(1-z)) - z] - \lambda q(1-z)[z - (1-p)B^*(\lambda q(1-z))]}dx, \]

From the above equations, we obtain \( I = 1 - \lambda q\beta_1 \) and \( B = \lambda q\beta_1 \). This completes the proof. \( \square \)
we obtain

Differentiating (34) and (35) with respect to \( z \)

\[
K(z) = \frac{\theta(1 - \lambda q \beta_1) - \lambda qp}{\theta[B^*(\lambda q(1 - z))] - z - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]}. \tag{32}
\]

**Proof.** Let \( P(z) = \int_0^\infty P(z, x)dx \). By (20), we get

\[
P(z) = \frac{z[\theta(1 - \lambda q \beta_1) - \lambda qp][1 - B^*(\lambda q(1 - z))] - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]}{\theta[B^*(\lambda q(1 - z))] - z - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]}. \tag{33}
\]

Obviously, \( K(z) \) is the sum of \( Q(z) \) and \( P(z) \), so we have

\[
K(z) = Q(z) + P(z)
\]

\[
= \frac{\theta(1 - \lambda q \beta_1) - \lambda qp}{\theta[B^*(\lambda q(1 - z))] - z - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]}.
\]

This completes the proof. \( \square \)

**Lemma 4.4.** In steady-state situation, for the \( M/G/1 \) queue with Bernoulli vacations, the distribution of the number of customers in the system, \( K_n = P(N(t) = n) \), has the following generating function:

\[
P(z) = \frac{z[\theta(1 - \lambda q \beta_1) - \lambda qp][1 - B^*(\lambda q(1 - z))] - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]}{\theta[B^*(\lambda q(1 - z))] - z - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]}.
\]

5. **Equilibrium strategy.** In this section, we will take the customers’ equilibrium strategies into account. In the fully unobservable case, a mixed strategy has the form ‘while arriving at time \( t \), enter with a certain probability \( q' \). We focus on the behavior of customers upon their arrivals in \( M/G/1 \) queue with Bernoulli vacations. It is assumed that each customer receives a reward of \( R \) units after service. This reward denotes his added value of being served. Moreover, there exists a waiting cost of \( C \) units per time unit. The cost is accumulated both in queue and in service. Customers are risk neutral, that is, each customer wishes to maximize his expected net benefit. Here we define the following functions:

\[
f(z) = B^*(\lambda q(1 - z))(1 - z), \tag{34}
\]

\[
g(z) = \theta[B^*(\lambda q(1 - z))] - z - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))], \tag{35}
\]

then we get

\[
K(z) = \frac{\theta(1 - \lambda q \beta_1) - \lambda qp}{\theta[B^*(\lambda q(1 - z))] - z - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]}f(z) \tag{36}
\]

Differentiating (34) and (35) with respect to \( z \) and letting \( z = 1 \), respectively, we obtain

\[
f'(1) = -1, \tag{37}
\]

\[
g'(1) = \theta(\lambda q \beta_1 - 1) + \lambda qp. \tag{38}
\]

From (36)-(38), we get

\[
K(1) = \lim_{z \to 1} K(z) = \lim_{z \to 1}\frac{\theta(1 - \lambda q \beta_1) - \lambda qp}{\theta[B^*(\lambda q(1 - z))] - z - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]}f(z) \tag{39}
\]

\[
= \frac{\theta(1 - \lambda q \beta_1) - \lambda qp}{\theta[B^*(\lambda q(1 - z))] - z - \lambda q(1 - z)[z - (1 - p)B^*(\lambda q(1 - z))]} = 1.
\]

Differentiating (34) and (35) with respect to \( z \) twice and letting \( z = 1 \), respectively, we obtain

\[
f''(1) = -2\lambda q \beta_1, \tag{40}
\]

\[
g''(1) = \lambda^2 q^2[\theta \beta_2 - 2(1 - p)\beta_1] + 2\lambda q. \tag{41}
\]
Then
\[ K'(1) = \lim_{z \to 1} K'(z) = [\theta(1 - \lambda q \beta_1) - \lambda q p] \lim_{z \to 1} \frac{f'(z)g(z) - f(z)g'(z)}{g^2(z)} \]
\[ = [\theta(1 - \lambda q \beta_1) - \lambda q p] \lim_{z \to 1} \frac{f''(z)g(z) - f(z)g''(z)}{2g(z)g'(z)} \]
\[ = 2\lambda q \beta_1 [\theta(1 - \lambda q \beta_1) - \lambda q p] + \lambda^2 q^2 [\theta \beta_2 - 2(1 - p) \beta_1] + 2\lambda q. \] (42)

Then the mean number of the customers in the system can be computed from
\[ EN(t) = K'(1) = \lambda q \beta_1 + \frac{\lambda^2 q^2 [\theta \beta_2 - 2(1 - p) \beta_1] + 2\lambda q}{2[\theta(1 - \lambda q \beta_1) - \lambda q p]}. \] (43)

Using Little's law, the expected sojourn time if a customer decides to enter upon arrival is
\[ EW = \frac{EN(t)}{\lambda q} = \beta_1 + \frac{\lambda q [\theta \beta_2 - 2(1 - p) \beta_1] + 2}{2[\theta(1 - \lambda q \beta_1) - \lambda q p]}. \] (44)

Obviously, \( EW > \beta_1 \). Under the condition that the steady-state condition, \( \theta(\lambda q \beta_1 - 1) + \lambda q p > 0 \), is satisfied, we have
\[ \lambda q [\theta \beta_2 - 2(1 - p) \beta_1] + 2 \geq 0. \] (45)

It is easy to verify that the expected mean sojourn time of a customer who decides to enter is strictly increasing in \( q \). Let \( q^e \) be customers’ equilibrium joining probability. The customers’ equilibrium behavior is summarized in the following theorem.

**Theorem 5.1.** In the fully unobservable \( M/G/1 \) queue with Bernoulli vacations, a unique Nash equilibrium strategy ‘enter with probability \( q^e \)’ exists, where \( q^e \) is given by
\[ q^e = \begin{cases} 0, & \text{if } \frac{R}{\theta} < \beta_1 + \frac{1}{\theta}; \\ q_1, & \text{if } \beta_1 + \frac{1}{\theta} \leq \frac{R}{\theta} \leq \beta_1 + \frac{\lambda \theta \beta_2 - 2\lambda (1 - p) \beta_1 + 2}{2(\theta - \lambda q \beta_1 - \lambda p)}; \\ 1, & \text{if } \frac{R}{\theta} > \beta_1 + \frac{\lambda \theta \beta_2 - 2\lambda (1 - p) \beta_1 + 2}{2(\theta - \lambda q \beta_1 - \lambda p)}; \end{cases} \] (46)

where \( q_1 = \frac{2(\theta R - \theta C) - C}{2(\lambda \theta q \beta_1 + p) + C(\theta \beta_2 - 2\theta q^2 - 2\beta_1)}. \)

**Proof.** If an arriving customer decides to enter the system, his expected net benefit per unit time is
\[ S(q) = R - C \left( \beta_1 + \frac{\lambda q (\theta \beta_2 - 2(1 - p) \beta_1 + 2)}{2(\theta(1 - \lambda q \beta_1) - \lambda q p)} \right). \] (47)

and
\[ S(0) = R - C \left( \beta_1 + \frac{1}{\theta} \right), \] (48)
\[ S(1) = R - C \left( \beta_1 + \frac{\lambda \theta \beta_2 - 2\lambda (1 - p) \beta_1 + 2}{2(\theta - \lambda q \beta_1 - \lambda p)} \right). \] (49)

From the above discussion, we know that \( EW \) is strictly increasing in \( q \), so \( S(q) \) is strictly decreasing in \( q \). If \( \beta_1 + \frac{1}{\theta} \leq \frac{R}{\theta} \leq \beta_1 + \frac{\lambda \theta \beta_2 - 2\lambda (1 - p) \beta_1 + 2}{2(\theta - \lambda q \beta_1 - \lambda p)} \), we easily prove that (47) has a unique root in \((0, 1)\) which is denoted by \( q_1 \). Obviously, \( q_1 \) is a Nash equilibrium strategy. We obtain the second part of (46). If \( \frac{R}{\theta} > \beta_1 + \frac{\lambda \theta \beta_2 - 2\lambda (1 - p) \beta_1 + 2}{2(\theta - \lambda q \beta_1 - \lambda p)} \), \( S(q) \) is positive for every \( q \), i.e., customers’ best strategy
is ‘join the queue with probability 1’ in this case. Thus, $q^e = 1$ is a unique Nash equilibrium strategy and we get the third part of (46). If $\frac{\theta}{\lambda} < \beta_1 + \frac{1}{\lambda}$, a newly arriving customer will suffer a negative benefit if he enters the system. In this case, his best strategy is ‘balk the system’ and the unique equilibrium point is $q^e = 0$. This completes the proof. 

**Remark 1.** Theorem 4.3 in Liu and Wang [18] is a special case of Theorem 5.1.

If the service times are exponential with rate $\mu$, the $M/G/1$ queue with Bernoulli vacations degenerates into the $M/M/1$ queue with Bernoulli vacations. Then $\beta_1 = \mu^{-1}$, $\beta_2 = 2\mu^{-1}$. Substituting $\beta_1, \beta_2$ into (46), we obtain Theorem 4.3 in Liu and Wang [18].

6. **Optimal controls.** In this section, we consider the decision problem of the social planner. There are two scenarios needing to be considered. We first consider an ideal scenario, in which no cost is incurred when the vacation probability $p$ and vacation rate $\theta$ change. However, in the real situation, the server will incur the costs if the vacation probability $p$ and vacation rate $\theta$ change. The reason is given as follows. Take the mail service system, which is described in Section 1, as an example. The email service system can be modeled as a queueing model with Bernoulli vacations, where virus scans, the scan completing rate are regarded as system vacations and vacation rate, respectively. After completing an email service, the server handles the next email with probability $1 - p$ if the system is nonempty, or performs the virus scan with probability $p$. $p$ is called vacation probability. In the real situation, if the email server scans viruses with less probability $p$, the server will face more risks of virus attacks, which result in a greater underlying loss due to breakdowns of the system. So the underlying loss is decreasing in vacation probability $p$, and is increasing in $1 - p$. In addition, improving the scan completing rate (i.e., vacation rate) results in the costs due to technical upgrades. Overall, the cost of the server is increasing in $1 - p$ and $\theta$. Therefore we also need to consider another more real scenario, i.e. the case that the server has costs. In this case, the server will incur the costs if the vacation probability $p$ and vacation rate $\theta$ change.

The social planner needs to determine the optimal policies under which the social welfare is maximized. We assume that $SW(q, p, \theta)$ denotes the social welfare which is the sum of the expected total utility of customers per unit time and the expected total utility of the server per unit time. Let $q^*$, $p^*$, $\theta^*$ be the socially optimal joining probability, the socially optimal vacation probability and the socially optimal vacation rate, respectively. Recalling Theorem 4.1, we know that the system is stable if the following inequality holds:

$$\theta(\lambda q \beta_1 - 1) + \lambda qp < 0.$$  \hspace{1cm} (50)

(50) is identical with $(1 - \lambda q \beta_1) > \frac{\lambda qp}{\theta}$. If (50) holds, $(1 - \lambda q \beta_1) > 0$ since $\frac{\lambda qp}{\theta} > 0$. We assume $(1 - \lambda q \beta_1) > 0$ in the rest part of this paper. In the steady-state situation, from (50), $q$, $p$ and $\theta$ should satisfy

$$q \in [0, 1] \cap \left[0, \frac{\theta}{\theta \lambda \beta_1 + \lambda p}\right),$$  \hspace{1cm} (51)

$$p \in [0, 1] \cap \left[0, \frac{\theta(1 - \lambda q \beta_1)}{\lambda q}\right),$$  \hspace{1cm} (52)

$$\theta \in \left(\frac{pq \lambda}{1 - q \lambda \beta_1}, \infty\right).$$  \hspace{1cm} (53)
Hence we should limit our search for \( q^*, p^* \) and \( \theta^* \) in the corresponding intervals. To ensure that a customer who faces an idle system must join, we assume that the expected net benefit of the joining customer is positive, that is,

\[
R - C\left(\frac{1}{\theta} + \beta_1\right) > 0.
\]  

(54)

If (54) does not hold, no customer joins the queue after the first departure leaving behind an empty system, and then the server will never be busy. We ignore this case in our paper.

6.1. The case that the server has no cost. In this subsection, we consider the case that the server has no cost. In this case, the social welfare \( SW(q, p, \theta) \) equals the expected total utility of customers per unit time since the server has no cost, that is

\[
SW(q, p, \theta) = \lambda q [R - CEW] = \lambda q \left[R - C\beta_1 - \frac{C\lambda q (\theta \beta_2 - 2(1 - p)\beta_1) + 2}{2(\theta(1 - \lambda q \beta_1) - \lambda q p)}\right].
\]  

(55)

To ensure the stability of the system, from (51) we should limit our search for socially optimal joining probability \( q^* \) in the interval \( [0, 1] \cap [0, \frac{\theta}{\lambda q \beta_1 + \lambda p}] \).

Proposition 1. For the \( M/G/1 \) queueing system with Bernoulli vacations, under the condition that the steady-state condition is satisfied, the social welfare function \( SW(q, p, \theta) \) is strictly concave in \( q \).

Proof. We first compute the first order derivative of the social welfare function \( SW(q, p, \theta) \). It can be written as

\[
\frac{\partial SW(q, p, \theta)}{\partial q} = \lambda \left[R + \frac{-2C(1 + \theta \beta_1)(\theta + q\lambda \beta_1(\psi(q) - \theta)) + Cq\theta \lambda(-\theta)\beta_2}{2(\psi(q))^2}\right],
\]  

(56)

where \( \psi(q) = -\theta + p q \lambda + q \theta \lambda \beta_1 \). The second order derivative of the social welfare function with respect to \( q \) is

\[
\frac{\partial^2 SW(q, p, \theta)}{\partial q^2} = \frac{C \theta \lambda^2 (2p + 2p \theta \beta_1 + \theta^2 \beta_2)}{(-\theta + pq \lambda + q \theta \lambda \beta_1)^3}.
\]  

(57)

Clearly, the numerator of the right term in (57) is positive. From the steady-state condition \( \theta(\lambda q \beta_1 - 1) + \lambda q p < 0 \), we know that the denominator is negative. Hence we have

\[
\frac{\partial^2 SW(q, p, \theta)}{\partial q^2} < 0.
\]  

(58)

The above inequality implies that the social welfare function \( SW(q, p, \theta) \) is strictly concave in \( q \). This completes the proof.

Based on the concavity of the social welfare function, we can easily prove that the socially optimal joining probability \( q^* \) is existent. Theorem 6.1 gives the specific form of the socially optimal joining probability.

Theorem 6.1. For \( M/G/1 \) queueing system with Bernoulli vacations, under the condition that the steady-state condition is satisfied, the socially optimal joining probability \( q^* \) exists and can be given by

\[
q^* = \min\{\hat{q}, 1\},
\]  

(59)

where \( \hat{q} \) is the solution of the following equation:

\[
\frac{2(1 + \theta \beta_1)(\theta + q \lambda \beta_1(\psi(q) - \theta)) - q \theta \lambda(\psi(q) - \theta)\beta_2}{2(\psi(q))^2} = \frac{R}{C},
\]  

(60)
and \( \psi(q) = -\theta + pq\lambda + q\theta\beta_1 \).

**Proof.** From (58), we easily find that the first order derivative of the social welfare function is strictly decreasing in \( q \). From (54) and (56), we have

\[
\frac{\partial SW(q,p,\theta)}{\partial q} \bigg|_{q=0} = \lambda [R - C(\frac{1}{\beta_1} + \beta_1)] > 0. \tag{61}
\]

(61) implies there exists a sufficient small positive number \( \epsilon \) such that \( SW(0, p, \theta) < SW(\epsilon, p, \theta) \). Hence \( q = 0 \) is not socially optimal joining probability. According to the concavity of the social welfare function, we know the socially optimal joining probability \( q^* \) must be one of \( \hat{q} \) and 1. If \( 0 < \hat{q} < 1 \), the socially optimal joining probability \( q^* = \hat{q} \); otherwise, \( q^* = 1 \). This completes the proof. \( \blacksquare \)

Now we consider the socially optimal joining probability. If \( q^* = 1 \), clearly, \( q^* \leq q^e \). Let us consider the case of \( 0 < q^* < 1 \). From (47) and (55), we know that \( SW(q,p,\theta) = \lambda q S(q) \) for any probability \( q \in [0, 1] \cap [0, \frac{\lambda(1-q\beta_1)}{\lambda p}] \). That means the social welfare function and the individual net benefit function have the same sign. \( S(q) < 0 \) for \( q > q^e \) since \( S(q) \) is decreasing in \( q \) and \( S(q^e) = 0 \). The fact implies that \( SW(q,p,\theta) < 0 \) for \( q > q^e \). Clearly, all probabilities greater than \( q^e \) are not socially optimal. Then we know that \( q^* \leq q^e \) as \( 0 < q^* < 1 \). We summarize the result in Proposition 2.

**Proposition 2.** For \( M/G/1 \) queueing system with Bernoulli vacations, under the condition that the steady-state condition is satisfied, the socially optimal joining probability \( q^* \) is not greater than the equilibrium joining probability \( q^e \), that is, \( q^* \leq q^e \).

The above result expresses that the socially optimal joining probability \( q^* \) is less than or equals the equilibrium joining probability \( q^e \). That means the social planner always hopes that each customer joins the queue with a relatively lower joining probability.

Now we consider the socially optimal vacation probability \( p^* \) and the socially optimal vacation rate \( \theta^* \) when the server has no cost. As the earlier state (see (52) and (53)), we search the socially optimal vacation probability in the interval \( [0, 1] \cap [0, \frac{\lambda(1-q\beta_1)}{\lambda p}] \) and the socially optimal vacation rate in the interval \( (\frac{\lambda p}{1-q\beta_1}, \infty) \). Under the condition that the system is stable, we easily find the social welfare \( SW(q, p, \theta) \) is decreasing in \( p \) and is increasing in \( \theta \). So the socially optimal vacation probability \( p^* = 0 \) and the socially optimal vacation rate \( \theta^* = \infty \). We summarize the results in Theorem 6.2.

**Theorem 6.2.** For \( M/G/1 \) queueing system with Bernoulli vacations, under the condition that the steady-state condition is satisfied, the socially optimal vacation probability \( p^* = 0 \) and the socially optimal vacation rate \( \theta^* = \infty \).

**Proof.** Observing (55), we immediately find the social welfare \( SW(q,p,\theta) \) is decreasing in \( p \). So the socially optimal vacation probability \( p^* = 0 \). Now we prove the socially optimal vacation rate \( \theta^* = \infty \). The first order derivative of the social welfare function \( SW(q,p,\theta) \) with respect to \( \theta \) is

\[
\frac{\partial SW(q,p,\theta)}{\partial \theta} = -q\lambda(-2 + 2(1 + C - Cp)q\beta_1\lambda - C\lambda^2q^2(\beta_2p - 2(-1 + p)\beta_1^2)) \frac{2(-\theta + pq\lambda + q\theta\lambda\beta_1)}{2(-\theta + pq\lambda + q\theta\lambda\beta_1)^2}. \tag{62}
\]
Obviously, the denominator of (62) is positive and the numerator is independent of \( \theta \), so the social welfare function is monotonous. From (55), we have
\[
\lim_{\theta \to \frac{pq\lambda}{1 - \lambda q\beta_1}} SW(q, p, \theta) = -\infty, \tag{63}
\]
and
\[
\lim_{\theta \to \infty} SW(q, p, \theta) = \lambda q \left( R - C\beta_1 - \frac{Cq\beta_2}{2(1 - \lambda q\beta_1)} \right). \tag{64}
\]
From (63) and (64), we get
\[
\lim_{\theta \to \frac{pq\lambda}{1 - \lambda q\beta_1}} SW(q, p, \theta) < \lim_{\theta \to \infty} SW(q, p, \theta). \tag{65}
\]
Hence, under the condition that the steady-state condition is satisfied, the social welfare \( SW(q, p, \theta) \) is increasing in \( \theta \), so the socially optimal vacation rate \( \theta^* = \infty \). This completes the proof.

Theorem 6.2 gives the socially optimal vacation probability \( p^* \) and the socially optimal vacation rate \( \theta^* \) when the server has no cost. According to Theorem 6.2, the socially optimal vacation probability \( p^* \) equals zero. That means, the socially optimal decision is that the sever does not take a Bernoulli vacation when the system is nonempty. In addition, Theorem 6.2 also shows the socially optimal vacation rate \( \theta^* = \infty \), which means the sever takes a vacation with infinite vacation rate when the system is idle. Then it is a socially optimal decision that the server does not take the normal vacation when the server has no cost. Based on the analysis above, in the case that the server does not incur the cost, the socially optimal decision is that the server does not take a Bernoulli vacation or the normal vacation.

The sensitivity analysis is performed to explore the effect of different parameters when the server has no cost. Fig.1 shows that the social welfare is concave in joining probability \( q \). According to the property of the concave function, socially optimal joining probability \( q^* \) can be computed. Figs.2-4 depict that both equilibrium joining probability \( q_e \) and socially optimal joining probability \( q^* \) are non-increasing in \( \lambda, \beta_1, p \). With the increase of the potential arrival rate, the system becomes more congested if the proportion of customers entering the system remains unchanged. Based on the pessimistic expectation, customers have less willingness to enter the system, so customers’ equilibrium joining probability is non-increasing in \( \lambda \). In addition, an over congested system may produce the negative effect on social welfare. To maximize the social welfare, the social planner needs to decrease the negative externality due to congestion. Hence, from the viewpoint of the social planner, it is the socially optimal decision to permit less proportion of customers to enter the system with the increase of the potential arrival rate. That means the socially optimal joining probability is non-increasing in the potential arrival rate. As the first moment of server time grows, the mean server time becomes longer, and the system becomes more congested. Accordingly, the equilibrium joining probability \( q_e \) and the socially optimal joining probability \( q^* \) also becomes smaller. As the earlier definition, \( p \) denotes the probability of taking vacation at the instant after the server completion. A bigger \( p \) results in more frequent vacations which make the system more congested. Customers’ willingness joining the queue becomes weaker. That is to say, customers join the queue with smaller equilibrium joining probability \( q_e \) and even do not join the queue. Hence the equilibrium joining probability \( q_e \) is non-increasing in \( p \). In addition, more congested system produces more negative externality. So the socially optimal joining probability \( q^* \) is also non-increasing in
For a given the first moment of server time $\beta_1$, the second moment of server time embodies the volatility of server time. Fig. 5 shows the increase of the volatility leads to the decrease of both the equilibrium joining probability and the socially optimal joining probability. Fig. 6 shows that both equilibrium joining probability $q^e$ and socially optimal joining probability $q^*$ are increasing in the vacation rate $\theta$. As we know, the mean vacation time equals $1/\theta$ and decreases with $\theta$. After a vacation completion, the system reenters the working state if the system is nonempty. If $\theta$ is bigger, the mean vacation time is shorter and then the system has a chance to reenter the working state with shorter time. So the system becomes more effective and more sparse as $\theta$ grows, and customers are more willing to join the queue. That is why equilibrium joining probability $q^e$ is nondecreasing in the vacation rate $\theta$. In addition, more sparse system produces less negative externality, so the socially optimal joining probability $q^*$ is increasing in the vacation rate $\theta$. Figs. 7-8 imply that with the increase of $C$ or the decrease of $R$, the joining customers obtain less and less net benefit. They have less willing to join the queue, so the equilibrium joining probability $q^e$ is non-increasing in $C$. On the other hand, the social welfare is decreasing as $C$ grows or $R$ decreases, so the socially optimal joining probability $q^*$ is non-increasing in $C$ and nondecreasing in $R$. Finally, comparing the size relationship between $q^*$ and $q^e$ in Figs. 2-8, we easily find the socially optimal joining probability $q^*$ is not greater than equilibrium joining probability $q^e$. The observation verifies the result of Proposition 2.

6.2. The case that the server has costs. In this subsection, we consider the case that the server has costs. As the analysis in the first paragraph of Section 6, the scenario that the server has costs is more practical. In the real situation, the cost of the server often is increasing in $1 - p$ and $\theta$. Assume that $C_b$, $C_s$ are the constant marginal loss of ‘$1 - p$’ per unit time and the constant marginal cost of the vacation rate $\theta$ per unit time, respectively. Social welfare $SW(q, p, \theta)$ equals the sum of the expected total utility of customers per unit time and the expected total utility of the server per unit time.

**Figure 1.** Social welfare $SW$ vs. joining probability $q$ for $R = 4, C = 1, \lambda = 0.5, \beta_1 = 1.2, \beta_2 = 1, p = 0.1$. 


utility of the server per unit time, that is
\[ SW(q, p, \theta) = \lambda q [R - CEW] - (1 - p) C_b - \theta C_s \] (66)

From (44), the above equation can be rewritten as follows:
\[ SW(q, p, \theta) = \lambda q \left( R - C\beta_1 - \frac{C\lambda q \beta_2 - 2(1 - p)\beta_1}{2(\theta(1 - \lambda q \beta_1) - \lambda q p)} \right) - (1 - p) C_b - \theta C_s. \] (67)
Comparing (55) with (67), we find that the socially optimal joining probability in the no-cost case is identical with the socially optimal joining probability in the having-costs case since the last two terms of (67) do not depend on $q$. Hence the socially optimal joining probability $q^*$ can be obtained from Theorem 6.1 no matter whether the server has costs or not. Obviously, Proposition 2 still holds, that is, the socially optimal joining probability $q^*$ is not greater than the equilibrium joining probability $q^e$. Now we consider the socially optimal vacation probability. From (52), we search the socially optimal vacation rate $p^*$ in the interval $[0,1] \cap \{0, \frac{\theta(1-\lambda \beta_1)}{\lambda q}\}$.

**Figure 4.** Equilibrium joining probability $q^e$ and socially optimal joining probability $q^*$ vs. $p$ for $R = 2, C = 1, \lambda = 0.5, \beta_1 = 0.56, \beta_2 = 0.25, \theta = 0.83$.

**Figure 5.** Equilibrium joining probability $q^e$ and socially optimal joining probability $q^*$ vs. the second moment of server time $\beta_2$ for $R = 2, C = 1, \lambda = 0.5, \beta_1 = 0.56, \theta = 0.83, p = 0.1$. 

Theorem 6.3. Consider an M/G/1 queueing system with Bernoulli vacations. Under the condition that the steady-state condition is satisfied, SW(q, p, θ) is strictly concave in p.

Proof. The first order derivative of the social welfare function SW(q, p, θ) with respect to p can be written as

\[
\frac{\partial SW(q, p, \theta)}{\partial p} = C_b + \frac{q^2 \lambda^2 (-2 + C(-2\beta_1 \theta + 2q\beta_1 \lambda - \beta_2 q\theta\lambda + 2q\beta_1^2 \theta\lambda))}{2(-\theta + pq\lambda + q\theta\lambda\beta_1)^2},
\]  

(68)
and the second order derivative is
\[
\frac{\partial^2 SW(q,p,\theta)}{\partial p^2} = -q^3 \lambda^3 \left( -2 + C(-2\beta_1 \theta + 2q\beta_1\lambda - \beta_2 q\theta\lambda + 2q\beta_2^2\theta\lambda) \right) \left( -\theta + pq\lambda + q\theta\lambda\beta_1 \right)^3.
\] (69)

Let \( f = -2 + C(-2\beta_1 \theta + 2q\beta_1\lambda - \beta_2 q\theta\lambda + 2q\beta_2^2\theta\lambda) \). Now we prove that \( f < 0 \). A proof by contradiction is adopted. If \( f \geq 0 \), from (68) we get
\[
\frac{\partial SW(q,p,\theta)}{\partial p} = C_b + \frac{q^2 \lambda^2 f}{2(-\theta + pq\lambda + q\theta\lambda\beta_1)^2} \geq 0.
\] (70)
(70) implies \( SW(q,p,\theta) \) is nondecreasing in \( p \). However, we easily find
\[
SW(q,0,\theta) = \lambda q \left( R - C\beta_1 - \frac{C\lambda q(\theta\beta_2 - 2\beta_1)}{2\theta(1 - \lambda q\beta_1)} \right) - C_b - \theta C_s < \infty.
\] (71)

From (45), we get
\[
\lim_{p \to \frac{\theta q - \lambda q\beta_1}{\lambda q}} SW(q,p,\theta) = -\infty.
\] (72)
(71) and (72) imply \( SW(q,p,\theta) \) is not nondecreasing in \( p \). It contradicts (70). So \( f < 0 \). From (69) and Theorem 4.1, we have
\[
\frac{\partial^2 SW(q,p,\theta)}{\partial p^2} = -q^3 \lambda^3 f \left( -\theta + pq\lambda + q\theta\lambda\beta_1 \right)^3 < 0.
\] (73)

Therefore, under the condition that the steady-state condition always holds, \( SW(q,p,\theta) \) is strictly concave in \( p \). This completes the proof.

According to Theorem 6.3, under the condition that the steady-state condition is satisfied, the social welfare function \( SW(q,p,\theta) \) is strictly concave in \( p \). From
earlier state (see (53)), we search the socially optimal vacation rate

How to decide the socially optimal vacation rate for the social planner? As the

θ

vacation rate

If ˆp < 0, the socially optimal vacation probability

Since

f

vacation probability in the interval [0, 1]. Obviously, the socially optimal vacation probability p∗ = 1. We summarize the results in the following theorem.

Theorem 6.4. Consider an M/G/1 queueing system with Bernoulli vacations. Under the condition that the steady-state condition is satisfied, the socially optimal vacation probability p∗ exists and can be given by

\[ p^* = \begin{cases} 0, & \hat{p} < 0, \\ \frac{\theta(1 - q\lambda\beta_1) - \sqrt{\frac{-q^2\lambda^3(2 + C(2\beta_1\theta - 2q\beta_1\lambda + \beta_2q\theta\lambda - 2q\beta_1^2\theta\lambda))}{2C_b}}}{\lambda q}, & 0 < \hat{p} < 1, \\ 1, & \hat{p} \geq 1. \end{cases} \]  

(75)

Theorem 6.4 gives the socially optimal vacation probability p∗. However, the vacation rate θ also plays a key role in the queueing system with Bernoulli vacations. How to decide the socially optimal vacation rate for the social planner? As the earlier state (see (53)), we search the socially optimal vacation rate θ∗ in the interval \( \left( \frac{pq\lambda}{1-q\lambda\beta_1}, \infty \right) \).

Theorem 6.5. Consider an M/G/1 queueing system with Bernoulli vacations. Under the condition that the steady-state condition is satisfied, the social welfare function SW(q, p, θ) is strictly concave in θ.

Proof. The first two order derivatives of the social welfare function SW(q, p, θ) with respect to θ are

\[ \frac{\partial SW(q, p, \theta)}{\partial \theta} = \frac{-q\lambda(-2 + C - C p)q\beta_1\lambda - C\lambda^2q^2(\beta_2p - 2(-1 + p)\beta_1^2)}{2(-\theta + pq\lambda + q\theta\lambda\beta_1)^2} - C_s, \]  

(76)

and

\[ \frac{\partial^2 SW(q, p, \theta)}{\partial \theta^2} = \frac{g(\theta)}{(-\theta + pq\lambda + q\theta\lambda\beta_1)^3}, \]  

(77)

where \( g(\theta) = q\lambda(-1 + q\beta_1\lambda)(-2 + 2(1 + C - C p)q\beta_1\lambda - C\lambda^2q^2(\beta_2p - 2(-1 + p)\beta_1^2)). \) From Theorem 4.1, the denominator of (77) is negative. If the numerator of (77) is positive, \( \frac{\partial^2 SW(q, p, \theta)}{\partial \theta^2} < 0, \) and then SW(q, p, θ) is strictly concave with respect to θ. Now we only need to prove that \( g(\theta) > 0. \) A proof by contradiction is adopted. Assume the inequality \( g(\theta) > 0 \) does not hold. Then \( g(\theta) \leq 0. \) From (77), we get

\[ \frac{\partial^2 SW(q, p, \theta)}{\partial \theta^2} = \frac{g(\theta)}{(-\theta + pq\lambda + q\theta\lambda\beta_1)^3} \geq 0. \]  

(78)
(78) implies that $SW(q,p,\theta)$ is convex with respect to $\theta$. From (76), the solution to $\frac{\partial SW(q,p,\theta)}{\partial \theta} = 0$ can be obtained as follows:

$$pq\lambda + \frac{\sqrt{q\lambda(2-2(1+C-Cp)q\beta_1\lambda + C\lambda^2q^2(\beta_2p-2(-1+p)\beta_2^2))}}{2C_s} \frac{1}{1 - q\lambda\beta_1}. \quad (79)$$

Since the condition that the steady-state condition holds, we have

$$1 - q\lambda\beta_1 > \frac{pq\lambda}{\theta} > 0. \quad (80)$$

Then we get

$$\frac{pq\lambda}{1 - q\lambda\beta_1} < \frac{pq\lambda + \sqrt{q\lambda(2-2(1+C-Cp)q\beta_1\lambda + C\lambda^2q^2(\beta_2p-2(-1+p)\beta_2^2))}}{2C_s} < \infty. \quad (81)$$

From (45), We can easily find

$$\lim_{\theta \to \frac{pq\lambda}{1 - q\lambda\beta_1}} SW(q,p,\theta) = \lim_{\theta \to \infty} SW(q,p,\theta) = -\infty. \quad (82)$$

For $\forall \theta \in (\frac{pq\lambda}{1 - q\lambda\beta_1}, \infty)$, we have

$$|SW(q,p,\theta)| < \infty. \quad (83)$$

(81)-(83) imply that $SW(q,p,\theta)$ is not convex with respect to $\theta$ in the interval $(\frac{pq\lambda}{1 - q\lambda\beta_1}, \infty)$. It contradicts (78). Hence $g(\theta) > 0$. This completes the proof. \hfill \Box

According to Theorem 6.5, the social welfare $SW(q,p,\theta)$ exists a unique maximum point in the interval $(\frac{pq\lambda}{1 - q\lambda\beta_1}, \infty)$ since it is strictly concave with respect to $\theta$. The maximum point can be obtained by solving the equation $\frac{\partial SW(q,p,\theta)}{\partial \theta} = 0$, and the solution can be written as

$$\theta^* = \frac{pq\lambda + \sqrt{q\lambda(2-2(1+C-Cp)q\beta_1\lambda + C\lambda^2q^2(\beta_2p-2(-1+p)\beta_2^2))}}{2C_s} \frac{1}{1 - q\lambda\beta_1}. \quad (84)$$

From (81), we know that the solution belongs to the interval $(\frac{pq\lambda}{1 - q\lambda\beta_1}, \infty)$. Based on the concavity of the social welfare, we get that (84) is the socially optimal vacation rate.

**Theorem 6.6.** Consider an $M/G/1$ queuing system with Bernoulli vacations. Under the condition that the steady-state condition is satisfied, the socially optimal vacation rate $\theta^*$ exists and can be given by

$$\theta^* = \frac{pq\lambda + \sqrt{q\lambda(2-2(1+C-Cp)q\beta_1\lambda + C\lambda^2q^2(\beta_2p-2(-1+p)\beta_2^2))}}{2C_s} \frac{1}{1 - q\lambda\beta_1}. \quad (85)$$

Theorem 6.4, Theorem 6.6 give the socially optimal vacation probability $p^*$ and the socially optimal vacation rate $\theta^*$, respectively. These are instructive for the social planner in the real life. In addition, since $\theta^* \neq \infty$, proper vacations are beneficial to the social welfare when the server has costs.

The sensitivity analysis is also performed to explore the effect of different parameters when the server has costs. Figs.9-10 show that the social welfare is strictly concave in both $p$ and $\theta$. Fig.11 implies that the socially optimal vacation probability $p^*$ is nondecreasing in $C_b$. As we know, the underlying loss increases with $C_b$. If $C_b$ is very small, the underlying loss is negligible. In this situation, the social
planner does not need to worry about the underlying loss, and can maximize the social welfare by reducing the waiting times of customers. So the socially optimal vacation probability \( p^* = 0 \). Moreover, to maximize the social welfare, the social planner should enhance the vacation probability as \( C_b \) grows. Fig.12-13 show that the socially optimal vacation probability \( p^* \) is non-increasing in the joining probability \( q \) and non-decreasing in the vacation rate \( \theta \). So the social planner should set the larger vacation probability if the joining probability \( q \) becomes smaller or the vacation rate \( \theta \) becomes bigger. Figs.14-15 show that the socially optimal vacation rate \( \theta^* \) is decreasing in \( C_s \) and increasing in the joining probability \( q \). It tells us
that the social planner should set a smaller vacation rate as $C_s$ grows or $q$ decreases.

![Figure 11. Socially optimal vacation probability $p^*$ vs. $C_b$ for $R = 10, C = 1, \theta = 1, \lambda = 0.8, \beta_1 = 0.6, \beta_2 = 1$.](image1)

Figure 11. Socially optimal vacation probability $p^*$ vs. $C_b$ for $R = 10, C = 1, \theta = 1, \lambda = 0.8, \beta_1 = 0.6, \beta_2 = 1$.

![Figure 12. Socially optimal vacation probability $p^*$ vs. joining probability $q$ for $R = 10, C = 1, \theta = 1, \lambda = 0.8, \beta_1 = 0.6, \beta_2 = 1$.](image2)

Figure 12. Socially optimal vacation probability $p^*$ vs. joining probability $q$ for $R = 10, C = 1, \theta = 1, \lambda = 0.8, \beta_1 = 0.6, \beta_2 = 1$.

7. The management of the social planner. Based on the analysis in Section 6, the socially optimal joining probability $q^*$ can be obtained from Theorem 6.1 no matter whether the server has costs or not, and it is not greater than the equilibrium joining probability $q_e^*$. However, all customers are selfish. They only consider individual benefits, and form an equilibrium joining probability which is not socially optimal. To maximize the social welfare, the social planner needs to establish some
measures to make that the equilibrium joining probability of customers equals the social optimal joining probability. An ordinary method is that the server imposes a permission fee $P_f$ on the customers who join the queue. How to determine the permission fee? Clearly, $q^*$ needs to satisfy the following equations:

\[
\begin{cases}
S(q^*) - P_f = 0, \\
q^* = q^*.
\end{cases}
\tag{86}
\]

where $S(q)$ is the expected net benefit of an arriving customer deciding to join the queue, and is given in (47). From the simple computation, we can obtain $P_f = S(q^*)$. 

---

**Figure 13.** Socially optimal vacation probability $p^*$ vs. vacation rate $\theta$ for $R = 10, C = 1, \theta = 1, \lambda = 0.8, \beta_1 = 0.6, \beta_2 = 1, C_b = 1$.

**Figure 14.** Socially optimal vacation rate $\theta^*$ vs. $C_s$ for $R = 10, C = 1, \lambda = 0.8, \beta_1 = 0.6, \beta_2 = 1, p = 0.32$. 
Figure 15. Socially optimal vacation rate $\theta^*$ vs. joining probability $q$ for $R = 10, C = 1, \lambda = 0.8, \beta_1 = 0.6, \beta_2 = 1, p = 0.32$.

Note that levying a permission fee $P_f$ does not change the social welfare. The social welfare is the sum of the expected total utility of customers per unit time and the expected total utility of the server per unit time. After the server levies a permission fee, the permission fee is transferred from a customer to the server, so the sum of the expected total utility of customers per unit time and the expected total utility of the server per unit time does not change. That means the permission fee has no effect on the social welfare. But levying the admission fee can make customers’ strategy be a socially optimal decision.

8. Conclusions. In this paper we study an unobservable $M/G/1$ queue with Bernoulli vacations. We first give the steady-state condition which ensures the system is stable, and then derive the steady-state equations and the partial generating functions of the joint distribution of the server state and the queue length. The probability that the server is busy (on Bernoulli vacation) can be obtained. We also derive the mean queue length and the mean waiting time. From the viewpoint of customers, we obtain customers’ Nash equilibrium strategies based on game theory. From the viewpoint of the social planner, we derive the socially optimal joining probability, the socially optimal vacation probability and the socially optimal vacation rate, respectively. These are instructive for the social planner in the real life. Comparing the socially optimal joining probabilities and the equilibrium joining probabilities, we find that the socially optimal joining probabilities are not greater than equilibrium joining probabilities. To maximize the social welfare, the social planner may impose a permission fee to make that customers’ equilibrium joining probability equals the social optimal joining probability. In addition, proper vacations are beneficial to the social welfare when the server has costs. The results obtained in this paper can provide managerial insights in web-clou ding service, file transfer service, and mail service, among others.

Acknowledgments. We would like to thank the anonymous referees for their valuable comments and suggestions that improved the presentation and the quality of this paper.
REFERENCES

[1] E. Altman and R. Hassin, Non-threshold equilibrium for customers joining an M/G/1 queue, *Proceedings of 10th International Symposium on Dynamic Game and Applications*, (2002), 56–64.

[2] A. Burnetas and A. Economou, Equilibrium customer strategies in a single server Markovian queue with setup times, *Queueing Systems*, **56** (2007), 213–228.

[3] A. Burnetas and N. Apostolos, Customer equilibrium and optimal strategies in Markovian queues in series, *Annals of Operations Research*, **208** (2013), 515–529.

[4] E. Cinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood cliffs, 1975.

[5] A. Economou, A. Gómez-Corral and S. Kanta, Optimal balking strategies in single-server queues with general service and vacation times, *Performance Evaluation*, **68** (2011), 967–982.

[6] H. M. Ghafir and C. B. Silio, Performance analysis of a multiple-access ring network, *IEEE Transactions on Communications*, **41** (1993), 1494–1506.

[7] W. J. Gray, P. P. Wang and M. K. Scott, A vacation queueing model with service breakdowns, *Applied Mathematical Modelling*, **24** (2000), 391–400.

[8] P. Guo and R. Hassin, Strategic behavior and social optimization in Markovian vacation queues, *Operations Research*, **59** (2011), 986–997.

[9] P. Guo and R. Hassin, Strategic behavior and social optimization in Markovian vacation queues: The case of heterogeneous customers, *European Journal of Operational Research*, **222** (2012), 278–286.

[10] R. Hassin and M. Haviv, *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, Kluwer Academic Publishers, Boston, 2003.

[11] R. Hassin, *Rational Queueing*, CRC Press, Raton, 2016.

[12] M. Haviv and Y. Kerner, On balking from an empty queue, *Queueing Systems*, **55** (2007), 239–249.

[13] J. Keilson and L. D. Servi, Oscillating random walk models for GI/G/1 vacation systems with Bernoulli schedules, *Journal of Applied Probability*, **23** (1986), 790–802.

[14] Y. Kerner, Equilibrium joining probabilities for an M/G/1 queue, *Games and Economic Behavior*, **71** (2011), 521–526.

[15] B. K. Kumar and S. P. Madheswari, Analysis of an M/M/N queue with Bernoulli service schedule, *International Journal of Operational Research*, **5** (2009), 48–72.

[16] B. K. Kumar, R. Rukmani and S. R. A. Lakshmi, Performance analysis of an M/G/1 queueing system under Bernoulli vacation schedules with server setup and close down periods, *Computers & Industrial Engineering*, **66** (2013), 1–9.

[17] C. D. Liou, Optimization analysis of the machine repair problem with multiple vacations and working breakdowns, *Journal of Industrial & Management Optimization*, **11** (2014), 83–104.

[18] J. Liu and J. Wang, Strategic joining rules in a single server Markovian queue with Bernoulli vacation, *Operational Research*, **17** (2017), 413–434.

[19] A. Manou, A. Economou and F. Karaesmen, Strategic customers in a transportation station: When is it optimal to wait?, *Operations Research*, **62** (2014), 910–925.

[20] P. Naor, The regulation of queue size by levying tolls, *Econometrica*, **37** (1969), 15–24.

[21] L. Servi, Average delay approximation of M/G/1 cyclic service queues with Bernoulli schedules, *IEEE Journal on Selected Areas in Communications*, **4** (1986), 813–822.

[22] Y. Shi and Z. Lian, Optimization and strategic behavior in a passenger-taxi service system, *European Journal of Operational Research*, **249** (2016), 1024–1032.

[23] J. Wang, J. Cao and Q. Li, Reliability analysis of the retrial queue with server breakdowns and repairs, *Queueing Systems*, **38** (2001), 363–380.

[24] J. Wang and F. Zhang, Equilibrium analysis of the observable queues with balking and delayed repairs, *Applied Mathematics and Computation*, **218** (2011), 2716–2729.

[25] J. Wang and F. Zhang, Strategic joining in M/M/1 retrial queues, *European Journal of Operational Research*, **230** (2013), 76–87.
[26] D. Yue, J. Yu and W. Yue, A Markovian queue with two heterogeneous servers and multiple vacations, Journal of Industrial & Management Optimization, 5 (2009), 453–465.

[27] F. Zhang, J. Wang and B. Liu, On the optimal and equilibrium retrial rates in an unreliable retrial queue with vacations, Journal of Industrial & Management Optimization, 8 (2013), 861–875.

Received August 2016; revised August 2017.

E-mail address: shengzhu@bjtu.edu.cn
E-mail address: jtwang@bjtu.edu.cn