Probability in the Engineering and Informational Sciences, **36**, 2022, 616–643.
doi:10.1017/S0269964820000704

ON MARKOVIAN QUEUES WITH SINGLE WORKING VACATION AND BERNOULLI INTERRUPTIONS

Ruiling Tian
College of Sciences, Yanshan University, Qinhuangdao, 066004 Hebei, China
E-mail: tianrl@ysu.edu.cn

Zhe George Zhang
Beedie School of Business, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
College of Business and Economics, Western Washington University, Bellingham, WA 98229, USA
E-mail: gzhang@sfu.ca

Siping Su
College of Business and Economics, Western Washington University, Bellingham, WA 98229, USA

This paper considers the customers' equilibrium and socially optimal joining–balking behavior in a single-server Markovian queue with a single working vacation and Bernoulli interruptions. The model is motivated by practical service systems where the service rate can be adjusted according to whether or not the system is empty. Specifically, we focus on a single-server queue in which the server's service rate is reduced from a regular to a lower one when the system becomes empty. This lower rate period is called a working vacation for the server which may represent that part of the service facility is under a maintenance process or works on other non-queueing job, or simply for saving the energy (for a machine server case). In this paper, we assume that the working vacation period is terminated after a random period or with probability p after serving a customer in a non-empty system. Such a system is called a queue with single working vacation and Bernoulli interruptions. Customers are strategic and can make choice of joining or balking based on different levels of system information. We consider four scenarios: fully observable, almost observable, almost unobservable, and fully unobservable queue cases. Under a reward-cost structure, we analyze the customer's equilibrium and social-optimal strategies. In addition, the effects of system parameters on optimal strategies are illustrated by numerical examples.

Keywords: Bernoulli vacation interruptions, equilibrium balking strategy, queuing system, social welfare, working vacation

1. INTRODUCTION

We consider a queueing system where the server may switch between two service rates. When the system becomes empty, the service rate is reduced from regular rate to low rate. This lower rate period is random and called a working vacation for the server. The word "working" means that the service rate is not zero. The working vacation is terminated either after a random period or with a certain probability at a service completion instant when some waiting customers exist. Such a model is motivated by some practical service systems. For example, the server may be a service facility that requires maintenance when a busy period ends. While the maintenance is in process, the service rate will be low (if the service can still be offered). Thus, the maintenance period becomes a "working vacation." In this paper, we consider a special type of interruptions during a working vacation. That is, during the working vacation period, if there are customers waiting at a service completion instant, the server can resume the normal service rate (interrupt the working vacation) with probability p $(0 < p < 1)$ or keep the slow service rate (continue the vacation) with probability $1 - p$. Such a vacation mechanism is called a single working vacation with Bernoulli interruptions. Introducing the Bernoulli vacation interruptions into the model makes the model more flexible to represent the real system with trade-off between minimizing customer waiting time and utilizing the idle time for maintaining the service facility. For example, if minimizing customer waiting time has higher priority than performing the maintenance on the server, then the vacation interruption probability p can be set higher. An extreme value of $p = 1$ represents that the maintenance is terminated whenever there is a waiting customer in the system at a service completion during the maintenance process, reflecting the highest priority of serving customers. The other extreme value of $p = 0$ represents that the maintenance must be completed once it starts. In fact, the latter is the classical single working vacation model. Another example is network services. In order to keep the server running smoothly, virus scanning regularly is desirable for the server. This type of scanning can be programmed to execute on a regular basis (i.e., whenever a service completion leaves an idle system). Although the virus scanning consumes some system resources and reduces the processing speed, the server can still provide service at a lower processing speed during the virus scanning period. Again, the Bernoulli interruptions can be set to balance chances of completing the virus scanning process $(1 - p)$ and chances of quickly resuming the regular processing speed p.

This single-server model can be also used to study the multi-server system approximately where a high service rate represents more servers are serving customers and a low service rate represents fewer servers are on duty. Then, when the system is empty, some idle servers may take a break or work on other non-queue jobs. The off-duty servers may return after a break (or finishing the non-queue job) or called back if some waiting customers exist. Another feature of our model is that customers can make choice of joining or balking based on the delay information (i.e., queue length). Four information scenarios are considered: fully observable queue, almost observable queue, almost unobservable queue, and unobservable queue. In a fully observable case, arriving customers can observe both the state of the server (either normal service or working vacation) and the queue length (i.e., the number of customers in the system). In the almost observable case, arriving customers can observe the queue length but do not know the state of the server. In an almost unobservable case, arriving customers cannot observe the queue length but know the state of the server. Finally, in an unobservable case, a customer observes neither the queue length nor the server state. An arriving customer decides whether to join the system or balk according to his or her service utility. This service utility can be computed based on the reward-cost structure and delay information. We study customer's equilibrium and socially optimal strategies for the service system of our interest.

Since our model belongs to a class of queues with server vacations, we briefly review the related literature. In the past decades, queueing models with vacations were developed as useful performance analysis tools for computer systems, communication networks, and flexible manufacturing systems. Recently, researchers paid more attention to customers' strategic behaviors in vacation models under a reward-cost structure. Burnetas and Economou [2] first studied a Markovian single-server queueing system with setup times and strategic customers. They derived equilibrium joining strategies for customers under various levels of delay information and analyzed the stationary behavior of the system under these strategies. Economou and Kanta [4] considered a single-server queue with breakdowns and repairs. They derived equilibrium threshold strategies in fully observable and almost observable queues. Sun *et al.* [15,16] studied customers' equilibrium and socially optimal balking strategies in observable and unobservable queues with several types of setup/closedown policies, respectively. More research in this area can be referred to Guo and Hassin [6,7], Tian *et al.* [22], Yu *et al.* [25,26], Liu and Wang [11], Doo [8], and references therein.

As a generalization from zero to non-zero service rate during a vacation, queueing systems with working vacations have been studied extensively. Studies on various working vacation queues can be found in a survey given by Tian *et al.* [21]. For work on customers' choice behavior in queueing systems with working vacations, Zhang *et al.* [27] and Sun and Li [14] studied equilibrium balking strategies in M/M/1 queues with multiple working vacations under different information scenarios. Subsequently, Sun *et al.* [17,18] considered customers' optimal balking behavior in some single-server Markovian queues with two-stage working vacations and double adaptive working vacations, respectively. Wang and Zhang [24] considered equilibrium strategies in Markovian queues with a single working vacation. Then, Tian *et al.* [23] studied customer equilibrium and social-optimal strategies in M/M/1 queues with multiple working vacations and vacation interruptions under three different levels of system information. Doo [9] studied customer's equilibrium joining/balking behaviors in M/M/1 queues with a single working vacation and vacation interruptions.

Vacation models with Bernoulli interruptions were studied by Tao *et al.* [19,20], Gao and Liu [5], and Li *et al.* [10]. All these studies did not consider strategic customers who can decide to join the system or not. In other words, all customers enter the system for service. However, to the best of the authors' knowledge, the equilibrium strategies of joining or balking in Markovian queueing systems with a single working vacation and Bernoulli interruptions have not been studied. Therefore, we investigate an M/M/1 queueing model with a single working vacation and Bernoulli interruptions where customers can make choice of joining or balking under different information scenarios. The only work that is closely related to ours is Doo [9] who studied the equilibrium balking strategies in a single vacation model with working vacation and vacation interruptions. However, our study is significantly different from his in two aspects. First, Doo's model is a special case of ours with $p = 1$ (i.e., the vacation must be interrupted if waiting customers exist at a service completion instant during vacation). From practical point of view, our model is more flexible as the trade-off between minimizing customer wait and completing some non-queue job (working vacation) can be addressed by changing p. Second, our analysis approach is different and more complete. For example, in the unobservable queue case, we not only obtain the stationary distribution but also show the decomposition property in the vacation model with strategic customers. Therefore, our results are more general than Doo's from both practical and theoretical perspectives.

This paper is organized as follows. Model description is given in Section 2. In Sections 3–6, we analyze the customer equilibrium and socially optimal strategies in the fully observable, almost observable, almost unobservable, and unobservable cases, respectively. In Section 7, some numerical examples are presented to illustrate the effects of several parameters on customers' behaviors and system performance. Finally, Section 8 concludes the paper with a brief summary.

2. MODEL DESCRIPTION

Consider an M/M/1 queueing system where customers arrive at the system according to a Poisson process with rate λ and their service times are assumed to be exponentially distributed with rate μ_1. At a service completion instant, if there is no customer in the system, the server begins a working vacation and the vacation time is assumed to be exponentially distributed with rate θ. During a vacation period, arriving customers can be still served and service times are exponentially distributed at a lower rate of μ_0 ($\mu_0 < \mu_1$). At a service completion during the vacation, if there are customers waiting in the system, the vacation is either interrupted (i.e., terminated) with probability p ($0 < p < 1$) or continues with probability $1 - p$. Meanwhile, when a vacation ends, if there are customers in the system, the server resumes service rate μ_1 from μ_0 and a regular service period starts. Otherwise, the server enters an idle period and a new regular busy period with service rate μ_1 starts when a customer arrives. We assume that the inter-arrival times, service times, and working vacation times are mutually independent. In addition, the service discipline is first in first out (FIFO).

Let $N(t)$ denote the number of customers in the system at time t and $J(t)$ be the state of server being with $J(t) = 0$ representing "on a working vacation" and $J(t) = 1$ representing "in a regular service period" at time t. It is assumed that customers are homogeneous and decide whether to join or balk upon arrival based on their own service utility. The customer's utility function is equal to a reward R for receiving service minus a waiting cost which is computed as expected waiting time, denoted by $E[W]$ times waiting cost per time unit, denoted by C. Finally, we assume that there are no retrials for balking customers nor reneging for waiting customers. Thus, the system state can be completely represented by $(N(t), J(t))$ which becomes a two-dimensional Markov chain due to the fact that all random variables are exponential and mutually independent.

3. FULLY OBSERVABLE QUEUES

We begin with the fully observable case in which arriving customers not only know the number of customers in the system, $N(t)$, but also the state of the server, $J(t)$, at arrival time t. In such a system, there exists a balking threshold $n(i)$ for a customer arriving at state with $(N(t) = n, J(t) = i)$, where $n = 0, 1, \ldots,\ i = 0, 1$. If $n \le n(i)$, the customer joins the system; otherwise balks.

Let T_{ni} be the customer's system time (waiting plus service), given that he arrives at state (n, i). Then, we have the following equations:

$$T_{00} = \frac{1}{\mu_0 + \theta} + \frac{\theta}{\mu_0 + \theta}\frac{1}{\mu_1}, \tag{1}$$

$$T_{n0} = \frac{\mu_0}{\mu_0 + \theta}\left(\frac{1}{\mu_0} + pT_{n-1,1} + (1-p)T_{n-1,0}\right) + \frac{\theta}{\mu_0 + \theta}T_{n1}, \quad n = 1, 2, \ldots, \tag{2}$$

$$T_{n1} = \frac{n+1}{\mu_1}, \quad n = 0, 1, 2, \ldots. \tag{3}$$

Taking into account (2) and (3), we obtain

$$\begin{aligned} T_{n0} - T_{n-1,0} - \frac{1}{\mu} &= \frac{(1-p)\mu_0}{\mu_0 + \theta}\left(T_{n-1,0} - T_{n-2,0} - \frac{1}{\mu}\right) \\ &= \left(\frac{(1-p)\mu_0}{\mu_0 + \theta}\right)^{n-1}\left(T_{10} - T_{00} - \frac{1}{\mu}\right). \end{aligned} \tag{4}$$

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

Using (1) and (2), we can get

$$T_{n0} - T_{n-1,0} = \frac{1}{\mu_1} + \frac{\mu_1 - \mu_0}{\mu_1(\mu_0 + \theta)} \left(\frac{(1-p)\mu_0}{\mu_0 + \theta} \right)^n. \tag{5}$$

After the iteration of (5), we have

$$T_{n0} = \frac{n}{\mu_1} + \frac{(1-p)\mu_0(\mu_1 - \mu_0)}{\mu_1(\mu_0 + \theta)(p\mu_0 + \theta)} \left(1 - \left(\frac{(1-p)\mu_0}{\mu_0 + \theta} \right)^n \right) + \frac{\mu_1 + \theta}{\mu_1(\mu_0 + \theta)}. \tag{6}$$

Based on the reward-cost structure, the expected utility of a joining customer is

$$U(n,i) = R - CT_{ni}.$$

From (3) and (6), we obtain $U(n,i)$ as follows:

$$\begin{aligned} U(n,0) &= R - CT_{n0} \\ &= R - C\left(\frac{n}{\mu_1} + \frac{(1-p)\mu_0(\mu_1 - \mu_0)}{\mu_1(\mu_0 + \theta)(p\mu_0 + \theta)} \left(1 - \left(\frac{(1-p)\mu_0}{\mu_0 + \theta} \right)^n \right) + \frac{\mu_1 + \theta}{\mu_1(\mu_0 + \theta)} \right), \end{aligned} \tag{7}$$

$$U(n,1) = R - CT_{n1} = R - \frac{C(n+1)}{\mu_1}. \tag{8}$$

Obviously, T_{n0} is increasing with respect to n, so $U(n,0)$ decreases in n. A customer strictly prefers to join the queue if $U(n,i) > 0$ and is indifferent between joining and balking if $U(n,i) = 0$. By solving $U(n,i) = 0$ for n, we obtain the customer's equilibrium thresholds $(n_e(0), n_e(1))$.

In order to obtain the closed form of equilibrium threshold, we make use of the non-elementary Lambert W Function which is known as the product logarithm or productlog function (See [1,3]).

Definition 1: *For all $z \geq -1/e$, the Lambert W function is defined as either one or two real-valued functions giving the solution to*

$$W(z)\, e^{W(z)} = z.$$

We refer to the specific branches of this function as $W_0(z)$ and $W_{-1}(z)$, with $W_0(z) > W_{-1}(z)$, $z \geq -1/e$. The graphs of both branches are shown in Figure 1.

Hence, a pure threshold strategy for customers is specified by a pair $(n_e(0), n_e(1))$ as stated in the following theorem.

Theorem 1: *In a fully observable case, there exist a pair of thresholds $(n_e(0), n_e(1))$ and equilibrium strategy has the form "a customer arriving at state $(N(t), J(t))$ enters if $N(t) \leq$*

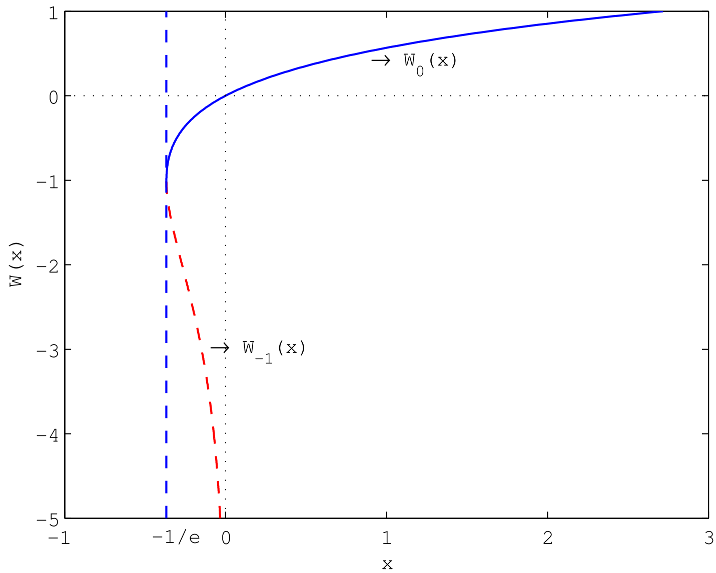

FIGURE 1. The real-valued branches of the Lambert W function.

$n_{\mathrm{e}}(J(t))$ *or balks otherwise." In addition,*

$$n_{\mathrm{e}}(0) = \left\lfloor d - \frac{W_0(bc^d \ln(c))}{\ln(c)} \right\rfloor, \tag{9}$$

$$n_{\mathrm{e}}(1) = \left\lfloor \frac{R\mu_1}{C} \right\rfloor - 1, \tag{10}$$

where

$$b = -\frac{(1-p)\mu_0(\mu_1-\mu_0)}{(\mu_0+\theta)(p\mu_0+\theta)}, \quad c = \frac{(1-p)\mu_0}{\mu_0+\theta},$$
$$d = \frac{R\mu_1}{C} - \frac{(1-p)\mu_0(\mu_1-\mu_0)+(\mu_1+\theta)(p\mu_0+\theta)}{(\mu_0+\theta)(p\mu_0+\theta)}.$$

PROOF: To obtain the customer's equilibrium thresholds $(n_{\mathrm{e}}(0), n_{\mathrm{e}}(1))$, we need to solve $U(n,i)=0$ for n and extend $U(n,i)$ from the non-negative integers to the reals.

Through mathematical operations, we see that the equation $U(x,0)=0$ is equivalent to

$$x + \frac{(1-p)\mu_0(\mu_1-\mu_0)}{(\mu_0+\theta)(p\mu_0+\theta)}\left(1-\left(\frac{(1-p)\mu_0}{\mu_0+\theta}\right)^x\right) + \frac{\mu_1+\theta}{\mu_0+\theta} = \frac{R\mu_1}{C}.$$

Using the help of Mathematica, we obtain

$$(d-x)\ln(c)\, e^{(d-x)\ln(c)} = \ln(c) bc^d. \tag{11}$$

By Definition 1, the solution of the equation $Xe^X = Y$ is $X = W(Y)$. Since equation (11) satisfies the form $Xe^X = Y$, we have

$$x = d - \frac{1}{\ln(c)} W_i(bc^d \ln(c)), \quad i = -1, 0. \tag{12}$$

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

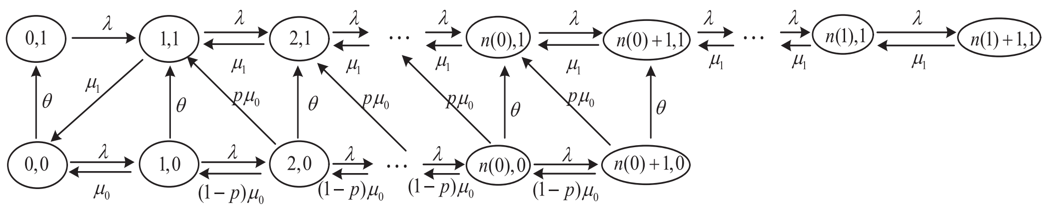

FIGURE 2. Transition rate diagram for a fully observable queue.

Due to $b < 0$ and $0 < c = ((1-p)\mu_0)/(\mu_0+\theta) < 1$, we can obtain $bc^d \ln(c) > 0$. It follows that the equilibrium threshold $n_e(0)$ is given by (10). By solving $U(n,1)=0$ for n, we obtain the customer's equilibrium threshold $n_e(1)$. This completes the proof. ■

To avoid the trivial case, the service reward must exceed the expected cost of joining a system empty. Thus, we assume

$$R > C \max\left\{\frac{1}{\mu_1}, \frac{\mu_1+\theta}{\mu_1(\mu_0+\theta)}\right\}.$$

Since $\mu_0 < \mu_1$, we have $(\mu_1+\theta)/(\mu_0+\theta) > 1$. Therefore, the condition is reduced to

$$R > \frac{C(\mu_1+\theta)}{\mu_1(\mu_0+\theta)}.$$

Next, we discuss social-benefit in fully observable queues. To obtain the performance measures, we consider the system process as a steady-state Markov chain with the state space

$$\Omega_{fo} = \{(n,i) \mid n = 0,1,2,\ldots,n(i)+1,\ i=0,1\}.$$

The transition diagram is shown in Figure 2.

Due to the customer choice behavior, the Markov chain eventually reaches steady-state. Let $\pi_{ni} = \lim_{t\to\infty} P\{N(t)=n, J(t)=i\}$ with $(n,i)\in\Omega_{fo}$; then $\{\pi_{ni} : (n,i)\in\Omega_{fo}\}$ is the stationary distribution of the process $\{(N(t),J(t)),\ t\geq 0\}$. These stationary probabilities satisfy the following flow balance equations:

$$(\lambda+\theta)\pi_{00} = \mu_0\pi_{10} + \mu_1\pi_{11}, \tag{13}$$

$$(\lambda+\theta+\mu_0)\pi_{n0} = \lambda\pi_{n-1,0} + (1-p)\mu_0\pi_{n+1,0}, \quad n = 1,2,\ldots,n(0), \tag{14}$$

$$(\theta+\mu_0)\pi_{n(0)+1,0} = \lambda\pi_{n(0),0}, \tag{15}$$

$$\lambda\pi_{01} = \theta\pi_{00}, \tag{16}$$

$$(\lambda+\mu_1)\pi_{n1} = \lambda\pi_{n-1,1} + \mu_1\pi_{n+1,1} + \theta\pi_{n0} + p\mu_0\pi_{n+1,0}, \quad n = 1,2,\ldots,n(0), \tag{17}$$

$$(\lambda+\mu_1)\pi_{n(0)+1,1} = \lambda\pi_{n(0)1} + \mu_1\pi_{n(0)+2,1} + \theta\pi_{n(0)+1,0}, \tag{18}$$

$$(\lambda+\mu_1)\pi_{n1} = \lambda\pi_{n-1,1} + \mu_1\pi_{n+1,1}, \quad n = n(0)+2,\ldots,n(1), \tag{19}$$

$$\mu_1\pi_{n(1)+1,1} = \lambda\pi_{n(1),1}. \tag{20}$$

The stationary distribution of the system can be obtained by solving these equations (the detailed derivations are presented in the Appendix). Since the probability of customer

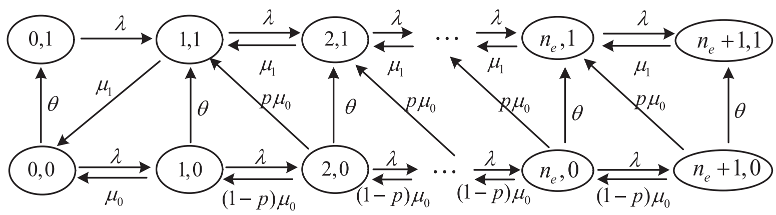

FIGURE **3.** Transition rate diagram for almost observable queues.

balking is $\pi_{n(0)+1,0} + \pi_{n(1)+1,1}$, the social-benefit (also called social-welfare) per time unit when all customers follow a threshold policy $(n(0), n(1))$ is given by

$$S_{fo}(n(0), n(1)) = \lambda R(1 - \pi_{n(0)+1,0} - \pi_{n(1)+1,1}) - C\left(\sum_{n=0}^{n(0)+1} n\pi_{n0} + \sum_{n=0}^{n(1)+1} n\pi_{n1}\right). \quad \mathbf{(21)}$$

Let $(n^*(0), n^*(1))$ be the socially optimal threshold strategy. That is

$$(n^*(0), n^*(1)) = \arg \max_{(n(0),n(1))} S_{fo}(n(0), n(1)).$$

Although we cannot analytically prove the discrepancy between customer equilibrium strategy and socially optimal strategy, we can numerically demonstrate such a discrepancy. For example, in a case with $R = 15$, $C = 1$, $\lambda = 1.5$, $\mu_1 = 2$, $\mu_0 = 0.5$, $\theta = 0.5$, $p = 0.4$, we obtain $(n^*(0), n^*(1)) = (10, 11)$ by numerical search. Meanwhile, the equilibrium strategy is $(n_e(0), n_e(1)) = (25, 29)$. Note that $n_e(0) > n^*(0)$ and $n_e(1) > n^*(1)$, which are consistent with the results for a single queue system [12].

4. ALMOST OBSERVABLE QUEUES

Next, we consider the almost observable case, where a customer can observe the queue length $N(t)$ but not the state of the server $J(t)$ at the arrival instant.

Hence, the corresponding Markov chain is from Section 3 with $n_e(0) = n_e(1) = n_e$ and the state space

$$\Omega_{ao} = \{(n, i) \mid n = 0, 1, 2, \ldots, n_e + 1,\ i = 0, 1\}.$$

The transition diagram is shown in Figure 3.

The stationary probabilities satisfy the following flow balance equations:

$$(\lambda + \theta)\pi_{00} = \mu_0\pi_{10} + \mu_1\pi_{11}, \quad \mathbf{(22)}$$

$$(\lambda + \theta + \mu_0)\pi_{n0} = \lambda\pi_{n-1,0} + (1-p)\mu_0\pi_{n+1,0}, \quad n = 1, 2, \ldots, n_e, \quad \mathbf{(23)}$$

$$(\theta + \mu_0)\pi_{n_e+1,0} = \lambda\pi_{n_e,0}, \quad \mathbf{(24)}$$

$$\lambda\pi_{01} = \theta\pi_{00}, \quad \mathbf{(25)}$$

$$(\lambda + \mu_1)\pi_{n1} = \lambda\pi_{n-1,1} + \mu_1\pi_{n+1,1} + \theta\pi_{n0} + p\mu_0\pi_{n+1,0}, \quad n = 1, 2, \ldots, n_e, \quad \mathbf{(26)}$$

$$\mu_1\pi_{n_e+1,1} = \lambda\pi_{n_e,1} + \theta\pi_{n_e+1,0}. \quad \mathbf{(27)}$$

The stationary distribution of the system can be obtained by solving these equations (the detailed derivations are presented in the Appendix). Now, we begin to look for the

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

expected net benefit of the customer that observes n customers and then decides to enter. We have the following result.

LEMMA 1: *In the almost observable queue where customers use the same balking threshold n_e, the net benefit of an arriving customer that observes n customers in the system and decides to enter is given by*

$$U_{n_e}(n) = R - \frac{C(n+1)}{\mu_1} - \frac{C(\mu_1-\mu_0)}{\mu_1(p\mu_0+\theta)}\left(1-\left(\frac{(1-p)\mu_0}{\mu_0+\theta}\right)^{n+1}\right) \times \frac{x_1^n+\delta x_2^n}{\frac{\mu_1-\mu_0}{\mu_1 x_1-\lambda}x_1^{n+1}+\frac{\mu_1-\mu_0}{\mu_1 x_2-\lambda}\delta x_2^{n+1}+(\frac{\lambda-\mu_0 x_1}{\lambda-\mu_1 x_1}+\frac{\theta}{\lambda}+(\frac{\lambda-\mu_0 x_2}{\lambda-\mu_1 x_2}+\frac{\theta}{\lambda})\delta)\rho^n}, \quad n=0,1,\ldots,n_e+1, \tag{28}$$

where $\rho=\lambda/\mu_1$ and $\delta=-((1-x_2)/(1-x_1))(x_1^{n_e+2}/x_2^{n_e+2})$.

PROOF: The net benefit of an arriving customer who observes n customers and decides to enter is given by $U_{n_e}(n)=R-CT(n)$, where $T(n)=E[S\,|\,N^-=n]$ is his mean sojourn time when he finds n customers in the system right before his arrival instant. Let $\pi_{I|N}(i\,|\,n)$ $(i=0,1)$ be the probability that an arriving customer finds there are n customers given that the server is at state i. Conditioning on the state of the server, we obtain

$$T(n) = T_{n0}\pi_{I|N}(0\,|\,n) + T_{n1}\pi_{I|N}(1\,|\,n) \tag{29}$$

and

$$\pi_{I|N}(i\,|\,n) = \frac{\lambda\pi_{n0}}{\lambda\pi_{n0}+\lambda\pi_{n1}}, \quad n=0,1,\ldots,n_e+1.$$

Using the stationary probabilities, we obtain the probabilities $\pi_{I|N}(i\,|\,n)$ for $n = 0,1,\ldots,n_e+1$. So, we obtain $T(n)$, $n=0,1,2,\ldots,n_e+1$. It is worth noting that customer does not enter the empty system if $U_{n_e}(0)<0$. Otherwise, he enters the queue.

When a customer observes j $(j=1,2,\ldots,n_e+1)$ customers and decides to enter, the expected sojourn time $T(j)$ is greater than the expected sojourn time $T(j-1)$ when $j-1$ customers are observed. Therefore, we have $U_{n_e}(j)<U_{n_e}(j-1)$ and $U_{n_e}(n)$ is decreasing with n.

Substituting $n=n_e$ and $n=n_e+1$ into (28), we get

$$U_{n_e}(n_e) = R - \frac{C(n_e+1)}{\mu_1} - \frac{C(\mu_1-\mu_0)}{\mu_1(p\mu_0+\theta)}\left(1-\left(\frac{(1-p)\mu_0}{\mu_0+\theta}\right)^{n_e+1}\right) \times \frac{\mu_0+\theta}{\frac{\lambda(\mu_1-\mu_0)(\mu_1(\mu_0+\theta)-\lambda(1-p)\mu_0)}{(1-p)\mu_0(\mu_1x_1-\lambda)(\mu_1x_2-\lambda)}+\frac{\lambda}{x_2-x_1}(\frac{(\theta+p\mu_0x_1)x_2}{\lambda-\mu_0x_1}(\frac{\lambda-\mu_0x_1}{\lambda-\mu_1x_1}+\frac{\theta}{\lambda})(\frac{\rho}{x_1})^{n_e} - \frac{(\theta+p\mu_0x_2)x_1}{\lambda-\mu_0x_2}(\frac{\lambda-\mu_0x_2}{\lambda-\mu_1x_2}+\frac{\theta}{\lambda})(\frac{\rho}{x_2})^{n_e})}, \tag{30}$$

$$U_{n_e}(n_e+1) = R - \frac{C(n_e+2)}{\mu_1} - \frac{C(\mu_1-\mu_0)}{\mu_1(p\mu_0+\theta)}\left(1-\left(\frac{(1-p)\mu_0}{\mu_0+\theta}\right)^{n_e+2}\right) \times \frac{(1-p)\mu_0}{\frac{\lambda(\mu_1-\mu_0)(\mu_1-\lambda)}{(\mu_1x_1-\lambda)(\mu_1x_2-\lambda)}+\frac{\lambda}{x_2-x_1}(\frac{\theta+p\mu_0x_1}{\lambda-\mu_0x_1}(\frac{\lambda-\mu_0x_1}{\lambda-\mu_1x_1}+\frac{\theta}{\lambda})(\frac{\rho}{x_1})^{n_e+1} - \frac{\theta+p\mu_0x_2}{\lambda-\mu_0x_2}(\frac{\lambda-\mu_0x_2}{\lambda-\mu_1x_2}+\frac{\theta}{\lambda})(\frac{\rho}{x_2})^{n_e+1})}. \tag{31}$$

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

Let

$$f_1(n) = R - \frac{C(n+1)}{\mu_1} - \frac{C(\mu_1-\mu_0)}{\mu_1(p\mu_0+\theta)}\left(1-\left(\frac{(1-p)\mu_0}{\mu_0+\theta}\right)^{n+1}\right) \times \frac{\mu_0+\theta}{\frac{\lambda(\mu_1-\mu_0)(\mu_1(\mu_0+\theta)-\lambda(1-p)\mu_0)}{(1-p)\mu_0(\mu_1x_1-\lambda)(\mu_1x_2-\lambda)} + \frac{\lambda}{x_2-x_1}\left(\frac{(\theta+p\mu_0x_1)x_2}{\lambda-\mu_0x_1}\left(\frac{\lambda-\mu_0x_1}{\lambda-\mu_1x_1}+\frac{\theta}{\lambda}\right)\left(\frac{\rho}{x_1}\right)^n - \frac{(\theta+p\mu_0x_2)x_1}{\lambda-\mu_0x_2}\left(\frac{\lambda-\mu_0x_2}{\lambda-\mu_1x_2}+\frac{\theta}{\lambda}\right)\left(\frac{\rho}{x_2}\right)^n\right)},$$
$$n = 0,1,2,\ldots \tag{32}$$

$$f_2(n) = R - \frac{C(n+1)}{\mu_1} - \frac{C(\mu_1-\mu_0)}{\mu_1(p\mu_0+\theta)}\left(1-\left(\frac{(1-p)\mu_0}{\mu_0+\theta}\right)^{n+1}\right) \times \frac{(1-p)\mu_0}{\frac{\lambda(\mu_1-\mu_0)(\mu_1-\lambda)}{(\mu_1x_1-\lambda)(\mu_1x_2-\lambda)} + \frac{\lambda}{x_2-x_1}\left(\frac{\theta+p\mu_0x_1}{\lambda-\mu_0x_1}\left(\frac{\lambda-\mu_0x_1}{\lambda-\mu_1x_1}+\frac{\theta}{\lambda}\right)\left(\frac{\rho}{x_1}\right)^n - \frac{\theta+p\mu_0x_2}{\lambda-\mu_0x_2}\left(\frac{\lambda-\mu_0x_2}{\lambda-\mu_1x_2}+\frac{\theta}{\lambda}\right)\left(\frac{\rho}{x_2}\right)^n\right)},$$
$$n = 0,1,2,\ldots. \tag{33}$$

Clearly, $f_1(n) = U_n(n)$, $n = 0,1,\ldots,n_e$, $f_2(n) = U_{n-1}(n)$. Moreover, $f_1(n) \geq f_2(n)$, $n \geq 0$.

We have that if $U_0(0) > 0$ and $\lim_{n\to\infty} f_1(n) = -\infty$, then $f_1(0) > 0$. So, we can find a finite number n_u in the sequence $(f_1(n))$ that satisfies inequality

$$f_1(0), f_1(1), \ldots, f_1(n_u) > 0, \quad f_1(n_u+1) \leq 0. \tag{34}$$

Obviously, $f_1(n) > f_2(n)$, $n = 0,1,\ldots$, so $f_2(n_u+1) < f_1(n_u+1) \leq 0$. In the range from 0 to n_u+1, we can find that a number n_l satisfies the inequality

$$f_2(n_l) > 0, \quad f_2(n_l+1), f_2(n_l+2), \ldots, f_2(n_u+1) \leq 0, \tag{35}$$

where n_l is the first positive term of the sequence $(f_2(n))$. If the sequence $(f_2(n))$ is non-positive between 0 and n_u+1, then we have

$$f_2(0), f_2(1), \ldots, f_2(n_u), f_2(n_u+1) \leq 0. \tag{36}$$

This completes the proof of the lemma. ■

Next, we establish the equilibrium threshold strategies stated in the following theorem.

THEOREM 2: *In the almost observable queues with single working vacation and Bernoulli interruptions, $n_e = n_l, n_l+1, \ldots, n_u$ are equilibrium strategies.*

PROOF: We consider a tagged customer at his arrival instant. It follows from (28), (32), and (34) that the customer prefers to enter when he finds n ($n \leq n_e$) customers in the system.

If an arriving customer finds that the number of customers in the system is n_e+1, the customer prefers to balk. It follows from (33), (35), and (36) that the expected net benefit $f_2(n_e+1) \leq 0$. Therefore, we conclude that there are equilibrium strategies $n_e = n_l, n_l+1, \ldots, n_u$. ■

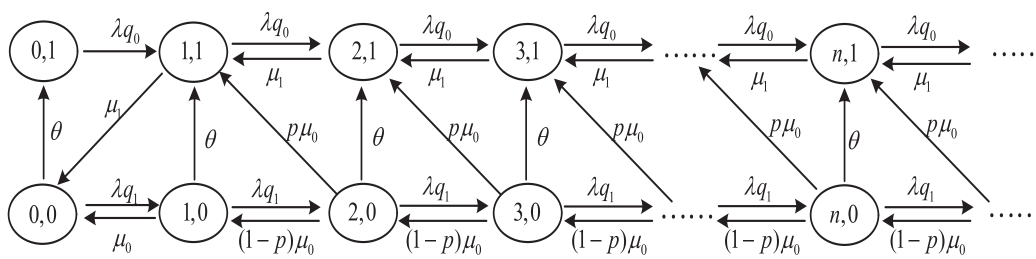

FIGURE 4. Transition rate diagram for almost unobservable queues.

Since the probability of customer balking is $\pi_{n_e+1,0} + \pi_{n_e+1,1}$, the equilibrium social-benefit per time unit when all customers follow the threshold policy n_e is given by

$$S_{ao}(n_e) = \lambda R(1 - \pi_{n_e+1,0} - \pi_{n_e+1,1}) - C \sum_{n=0}^{n_e+1} n(\pi_{n0} + \pi_{n1}). \tag{37}$$

Let n^* be the socially optimal threshold strategy. That is,

$$n^* = \arg\max_n S_{ao}(n).$$

For example, when $R = 10$, $C = 1$, $\lambda = 0.8$, $\mu_1 = 2$, $\mu_0 = 0.6$, $\theta = 0.3$, and $p = 0.3$, we obtain $n^* = 11$. While customers' equilibrium strategy is $n_e = 17, 18$. We observe that $n^* < n_e$, which indicates that individual optimization results in a queue that is longer than the socially desired one.

5. ALMOST UNOBSERVABLE QUEUES

In this section, we turn to the almost unobservable case, where a customer can observe the state of the server $J(t)$ but not the queue length $N(t)$ at the arrival instant. The customer equilibrium strategy in this case can be described by a joining probability q_i $(0 \leq q \leq 1)$, which is the proportion of joining customers when the server is in the state $J(t) = i$, $i = 0, 1$, and the effective arrival rate to the system is λq_i. While the two extreme values of $q_i = 0, 1$ represent two pure strategies of joining and balking, respectively, $0 < q_i < 1$ represents a mixed strategy. The state space for this case is $\Omega_{au} = \{(n, i) \,|\, n \geq 0, \;\; i = 0, 1\}$ and the transition diagram is illustrated in Figure 4.

We denote the stationary distribution as

$$\pi_{ni} = \lim_{t\to\infty} P\{N(t) = n, J(t) = i\}, \quad (n, i) \in \Omega_{po}.$$

$$\boldsymbol{\pi}_n = (\pi_{n0}, \pi_{n1}), \quad n \geq 0.$$

Using the lexicographical sequence for the states, the transition rate matrix $\boldsymbol{Q}$ can be written as a tri-diagonal block form:

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{A}_0 & \boldsymbol{C}_0 & & & \\ \boldsymbol{B}_1 & \boldsymbol{A} & \boldsymbol{C} & & \\ & \boldsymbol{B} & \boldsymbol{A} & \boldsymbol{C} & \\ & & \boldsymbol{B} & \boldsymbol{A} & \boldsymbol{C} \\ & & & \ddots & \ddots & \ddots \end{pmatrix},$$

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

where

$$\boldsymbol{A}_0 = \begin{pmatrix} -(\lambda q_0 + \theta) & \theta \\ 0 & -\lambda q_1 \end{pmatrix}, \quad \boldsymbol{B}_1 = \begin{pmatrix} \mu_0 & 0 \\ \mu_1 & 0 \end{pmatrix},$$
$$\boldsymbol{A} = \begin{pmatrix} -(\lambda q_0 + \theta + \mu_0) & \theta \\ 0 & -(\lambda q_1 + \mu_1) \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} (1-p)\mu_0 & p\mu_0 \\ 0 & \mu_1 \end{pmatrix},$$
$$\boldsymbol{C} = \begin{pmatrix} \lambda q_0 & 0 \\ 0 & \lambda q_1 \end{pmatrix}.$$

The structure of $\boldsymbol{Q}$ indicates that $\{(N(t), J(t)), t \geq 0\}$ is a quasi-birth-and-death (QBD) process (see [13]). To analyze this QBD process, it is necessary to solve for the minimal non-negative solution of the matrix quadratic equation

$$\boldsymbol{R}^2\boldsymbol{B} + \boldsymbol{R}\boldsymbol{A} + \boldsymbol{C} = 0, \tag{38}$$

and this solution, denoted by $\boldsymbol{R}$ and called the rate matrix, is obtained in the following lemmas.

LEMMA 2: *If $\rho_1 = \lambda q_1 \mu^{-1} < 1$, the matrix equation (38) has the minimal non-negative solution*

$$\boldsymbol{R} = \begin{pmatrix} r_0 & \dfrac{(\theta + p\mu_0 r_0) r_0}{\mu_1(1 - r_0)} \\ 0 & \rho_1 \end{pmatrix},$$

where

$$r_0 = \frac{\lambda q_0 + \mu_0 + \theta - \sqrt{(\lambda q_0 + \mu_0 + \theta)^2 - 4\lambda q_0 (1-p)\mu_0}}{2(1-p)\mu_0} \tag{39}$$

and $0 < r_0 < 1$.

PROOF: Because matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are all upper triangular, $\boldsymbol{R}$ should have the same structure as

$$\boldsymbol{R} = \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}.$$

Substituting $\boldsymbol{R}^2$ and $\boldsymbol{R}$ into (38), we obtain the following set of equations:

$$\begin{cases} (1-p)\mu_0 r_{11}^2 - (\lambda q_0 + \mu_0 + \theta) r_{11} + \lambda q_0 = 0, \\ \mu_1 r_{22}^2 - (\lambda q_1 + \mu_1) r_{22} + \lambda q_1 = 0, \\ p\mu_0 r_{11}^2 + \mu_1 r_{12}(r_{11} + r_{22}) + \theta r_{11} - (\lambda q_1 + \mu_1) r_{12} = 0. \end{cases} \tag{40}$$

Using the discriminant of quadratic equation, we have $r_{11} = r_0$, $0 < r_0 < 1$, where r_0 is one root of the first equation of (40) (the other root is greater than 1). It follows from the second equation of (40) that $r_{22} = \rho_1$ (the other roots is $r_{22} = 1$). Substituting r_0 and ρ_1 into the last equation of (40), we obtain $r_{12} = ((\theta + p\mu_0 r_0) r_0)/(\mu_1(1 - r_0))$. This completes the proof of the Lemma 2. ■

From Lemma 2, we also know that r_0 satisfies the following relationships:

$$\frac{\theta + p\mu_0 r_0}{1-r_0} + \mu_0 = \frac{\lambda q}{r_0}, \tag{41}$$

$$(\lambda q_0 - \mu_0 r_0)(1-r_0) = (\theta + p\mu_0 r_0) r_0 > 0. \tag{42}$$

Using the matrix-geometric solution method, we have

$$\boldsymbol{\pi}_n = (\pi_{n0}, \pi_{n1}) = (\pi_{10}, \pi_{11})\boldsymbol{R}^{n-1}, \quad n \geq 1, \tag{43}$$

and $(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$ satisfies the set of equations:

$$(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}) B[\boldsymbol{R}] = 0, \tag{44}$$

where

$$\begin{aligned} B[\boldsymbol{R}] &= \begin{pmatrix} \boldsymbol{A}_0 & \boldsymbol{C}_0 \\ \boldsymbol{B}_1 & \boldsymbol{A} + \boldsymbol{R}\boldsymbol{B} \end{pmatrix} \\ &= \begin{pmatrix} -(\lambda q_0 + \theta) & \theta & \lambda q_0 & 0 \\ 0 & -\lambda q_1 & 0 & \lambda q_1 \\ \mu_0 & 0 & -\dfrac{\lambda q_0}{r_0} & \dfrac{\theta + p\mu_0 r_0}{1-r_0} \\ \mu_1 & 0 & 0 & -\mu_1 \end{pmatrix}. \end{aligned} \tag{45}$$

In terms of r_0, we can obtain the explicit expressions for the stationary probabilities.

Theorem 3: *Assumed that $\rho_1 < 1$, the stationary probabilities $\{\pi_{ni} : (n,i) \in \Omega_{po}\}$ are as follows:*

$$\begin{cases} \pi_{n0} = K r_0^n, \quad n \geq 0, \\ \pi_{n1} = K \left(\dfrac{\theta}{\mu_1}\rho_1^{n-1} + \dfrac{(\theta + p\mu_0 r_0) r_0}{\mu_1 (1-r_0)} \sum_{j=0}^{n-1} r_0^j \rho_1^{n-1-j} \right), \quad n \geq 1, \end{cases} \tag{46}$$

where

$$K = (1-\rho_1)(1-r_0)\left(\frac{\lambda q_1 + \theta}{\lambda q_1}(1-r_0) + \left(1 - \frac{\mu_0}{\mu_1}\right) r_0 + \frac{\lambda q_0 - \lambda q_1}{\mu_1} \right)^{-1}. \tag{47}$$

Proof: From (44) and (45), we obtain the following equations:

$$\begin{cases} -(\lambda q_1 + \theta)\pi_{00} + \mu_0 \pi_{10} + \mu_1 \pi_{11} = 0, \\ \theta \pi_{00} - \lambda q_1 \pi_{01} = 0, \\ \lambda q_0 \pi_{00} - \dfrac{\lambda q_0}{r_0}\pi_{10} = 0, \\ \lambda q_1 \pi_{01} + \dfrac{\theta + p\mu_0 r_0}{1-r_0}\pi_{10} - \mu_1 \pi_{11} = 0. \end{cases} \tag{48}$$

Solving (48) and letting $\pi_{00} = K$, we obtain

$$(\pi_{00}, \pi_{01}) = K\left(1, \frac{\theta}{\lambda q_1}\right), \tag{49}$$

$$(\pi_{10}, \pi_{11}) = K\left(r_0, \frac{\theta}{\mu_1} + \frac{(\theta + p\mu_0 r_0)r_0}{\mu_1(1-r_0)}\right). \tag{50}$$

Furthermore,

$$\boldsymbol{R}^{n-1} = \begin{pmatrix} r_0^{n-1} & \dfrac{(\theta + p\mu_0 r_0)r_0}{\mu_1(1-r_0)} \displaystyle\sum_{j=0}^{n-2} r_0^j \rho_1^{n-j-2} \\ 0 & \rho_1^{n-1} \end{pmatrix}, \tag{51}$$

Substituting (49), (50), and (51) into (43), we obtain (46). Finally, $\pi_{00} = K$ can be determined by the normalization condition. This completes the proof. ■

From (46), we have the steady-state probability that the server is in state $J(t) = i$, denoted by p_i, as follows:

$$p_0 = \sum_{n=0}^{\infty} \pi_{n0} = \frac{K}{1-r_0}, \tag{52}$$

$$p_1 = \sum_{n=0}^{\infty} \pi_{n1} = K\left(\frac{\theta}{\lambda q_1} + \frac{\theta}{\mu_1(1-\rho_1)} + \frac{(\theta + p\mu_0 r_0)r_0}{\mu_1(1-r_0)^2(1-\rho_1)}\right). \tag{53}$$

The conditional expected sojourn time (waiting plus service) of a joining customer who finds the server in state $i = 0$ or $i = 1$ is given by

$$\begin{aligned} W_0(q_0) &= \frac{\sum_{n=0}^{\infty} T_{n0}\pi_{n0}}{p_0} \\ &= \frac{r_0}{\mu_1(1-r_0)} + \frac{(1-p)\mu_0(\mu_1-\mu_0)}{\mu_1(\mu_0+\theta)} \frac{r_0}{\mu_0+\theta-(1-p)\mu_0 r_0} + \frac{\mu_1+\theta}{\mu_1(\mu_0+\theta)}, \end{aligned} \tag{54}$$

as a function of q_0 or

$$\begin{aligned} W_1(q_0, q_1) &= \frac{\sum_{n=0}^{\infty} T_{n1}\pi_{n1}}{p_1} \\ &= \frac{1}{\mu_1 - \lambda q_1} + \frac{\lambda q_1(\lambda q_0 - \mu_0 r_0)}{\mu_1(1-r_0)(\theta\mu_1(1-r_0) + \lambda q_1(\lambda q_0 - \mu_0 r_0))}, \end{aligned} \tag{55}$$

as a function of q_0 and q_1. Based on the reward-cost structure, the expected net benefit (i.e., utility) of an arriving customer joining the system at server state i is as follows:

$$U_0(q_0) = R - C\left(\frac{r_0}{\mu_1(1-r_0)} + \frac{\mu_0 r_0(1-p)(\mu_1-\mu_0)}{\mu_1(\mu_0+\theta)(\mu_0+\theta-(1-p)\mu_0 r_0)} + \frac{\mu_1+\theta}{\mu_1(\mu_0+\theta)}\right), \tag{56}$$

$$U_1(q_0, q_1) = R - C\left(\frac{1}{\mu_1 - \lambda q_1} + \frac{\lambda q_1(\lambda q_0 - \mu_0 r_0)}{\mu_1(1-r_0)(\theta\mu_1(1-r_0) + \lambda q_1(\lambda q_0 - \mu_0 r_0))}\right). \tag{57}$$

Using (56) and (57), we can determine equilibrium strategies for a customer in this partially observable queue case in the following.

THEOREM 4: *For a partially observable queue, there exists a unique mixed-equilibrium strategy (q_0^e, q_1^e) as follows:*

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

Case (1): $\frac{C(\theta+\mu_1)}{\mu_1(\theta+\mu_0)} < R \le \frac{Cr(1)}{\mu_1(1-r(1))} + \frac{C\mu_0 r(1)(1-p)(\mu_1-\mu_0)}{\mu_1(\mu_0+\theta)(\mu_0+\theta-(1-p)\mu_0 r(1))} + \frac{C(\mu_1+\theta)}{\mu_1(\mu_0+\theta)}$.

$$(q_0^e, q_1^e) = \begin{cases} (x_1, x_2), & \dfrac{C}{\mu_1} \le R \le \dfrac{C}{\mu_1-\lambda} + \dfrac{C\lambda(\lambda x_1 - \mu_0 r(x_1))}{\mu_1(1-r(x_1))(\theta\mu_1(1-r(x_1)) + \lambda(\lambda x_1 - \mu_0 r(x_1)))}, \\ (x_1, 1), & R > \dfrac{C}{\mu_1-\lambda} + \dfrac{C\lambda(\lambda x_1 - \mu_0 r(x_1))}{\mu_1(1-r(x_1))(\theta\mu_1(1-r(x_1)) + \lambda(\lambda x_1 - \mu_0 r(x_1)))}. \end{cases}$$

Case (2): $R > \frac{Cr(1)}{\mu_1(1-r(1))} + \frac{C\mu_0 r(1)(1-p)(\mu_1-\mu_0)}{\mu_1(\mu_0+\theta)(\mu_0+\theta-(1-p)\mu_0 r(1))} + \frac{C(\mu_1+\theta)}{\mu_1(\mu_0+\theta)}$.

$$(q_0^e, q_1^e) = \begin{cases} (1, 0), & R < \dfrac{C}{\mu_1}, \\ (1, x_3), & \dfrac{C}{\mu_1} \le R \le \dfrac{C}{\mu_1-\lambda} + \dfrac{C\lambda(\lambda - \mu_0 r(1))}{\mu_1(1-r(1))(\theta\mu_1(1-r(1)) + \lambda(\lambda - \mu_0 r(1)))}, \\ (1, 1), & R > \dfrac{C}{\mu_1-\lambda} + \dfrac{C\lambda(\lambda - \mu_0 r(1))}{\mu_1(1-r(1))(\theta\mu_1(1-r(1)) + \lambda(\lambda - \mu_0 r(1)))}, \end{cases}$$

where

$$x_1 = \{q_0 | U_0(q_0) = 0, 0 < q_0 < 1\}, x_2 = \{q_1 | U_1(x_1, q_1) = 0, 0 < q_1 < 1\},$$

$$x_3 = \{q_1 | U_1(1, q_1) = 0, 0 < q_1 < 1\}, r_{(1)} = \frac{\lambda + \mu_0 + \theta - \sqrt{(\lambda + \mu_0 + \theta)^2 - 4\lambda(1-p)\mu_0}}{2(1-p)\mu_0},$$

$$r_{(x_1)} = \frac{\lambda x_1 + \mu_0 + \theta - \sqrt{(\lambda x_1 + \mu_0 + \theta)^2 - 4\lambda x_1(1-p)\mu_0}}{2(1-p)\mu_0}.$$

PROOF: Taking the first-order derivative of (39) with respect to q_0, we have

$$\frac{dr_0}{dq_0} = \frac{\lambda}{2(1-p)\mu_0}\left(1 - \frac{\lambda q_0 + \theta - (1-2p)\mu_0}{\sqrt{(\lambda q_0 + \theta - (1-2p)\mu_0)^2 + 4(1-p)\mu_0(\theta + p\mu_0)}}\right) > 0, \tag{58}$$

so r_0 is strictly increasing in $q_0 \in [0, 1]$. Then, we obtain

$$\frac{dW_0}{dq_0} = \frac{\frac{dr_0}{dq_0}}{\mu_1(1-r_0)^2} + \frac{(1-p)\mu_0(\mu_1-\mu_0)}{\mu_1(\mu_0+\theta)} \frac{(\mu_0+\theta)\frac{dr_0}{dq_0}}{(\mu_0+\theta-(1-p)\mu_0 r_0)^2} > 0, \tag{59}$$

$$\frac{dW_1}{dq_1} = \frac{\lambda}{(\mu_1 - \lambda q_1)^2} + \frac{\lambda\theta(\lambda q_0 - \mu_0 r_0)}{(\theta\mu_1(1-r_0) + \lambda q_1(\lambda q_0 - \mu_0 r_0))^2} > 0. \tag{60}$$

Therefore, $W_0(q_0)$ is strictly increasing in $q_0 \in [0, 1]$ and $W_1(q_0, q_1)$ is strictly increasing in $q_1 \in [0, 1]$. Thus, we conclude that $U_0(q_0)$ and $U_1(q_0, q_1)$ are decreasing with q_0 and q_1, respectively.

First, we focus on $U_0(q_0)$. Condition $R > C(\mu_1 + \theta)/(\mu_1(\mu_0 + \theta))$ ensures the existence of q_0^e. Therefore, we have two cases.

Case 1: $U_0(0) > 0$ and $U_0(1) \le 0$. That is, $R > (C(\theta + \mu_1))/(\mu_1(\theta + \mu_0))$ and $R \le CW_0(1)$. In this case, if all other customers finding empty system enter the system, that is $q_0^e = 1$, then the tagged customer receives a negative expected benefit by joining the system. Hence, $q_0^e = 1$ does not lead to an equilibrium. Similarly, if all other customers balk $q_0^e = 0$, then the tagged customer receives a positive expected benefit by joining the system; thus, $q_0^e = 0$ does not lead to an equilibrium either. Therefore, there exists a unique q_0^e satisfying

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

$R - CW_0(q_0^e) = 0$ for which customers are indifferent between joining and balking. This is given by $q_0^e = x_1$.

In this situation, the expected benefit U_1 is given by

$$U_1(x_1, q_1) = R - \frac{C}{\mu_1 - \lambda q_1} - \frac{C\lambda q_1(\lambda x_1 - \mu_0 r(x_1))}{\mu_1(1 - r(x_1))(\theta\mu_1(1 - r(x_1)) + \lambda q_1(\lambda x_1 - \mu_0 r(x_1)))}. \quad \textbf{(61)}$$

Using the equilibrium analysis in an unobservable queue case, we derive from (45) the following equilibria:

(1) if $U_1(x_1, 0) < 0$, then the equilibrium strategy is $q_1^e = 0$.

(2) if $U_1(x_1, 0) \geq 0$ and $U_1(x_1, 1) \leq 0$, then there exists a unique q_1^e satisfying $U_1(x_1, q_1^e) = 0$ and the equilibrium strategy is $q_1^e = x_2$.

(3) if $U_1(x_1, 1) > 0$, then the equilibrium strategy is $q_1^e = 1$.

Case 2: $U_0(1) \geq 0$. That is, $R > \frac{Cr(1)}{\mu_1(1-r(1))} + \frac{C\mu_0 r(1)(1-p)(\mu_1-\mu_0)}{\mu_1(\mu_0+\theta)(\mu_0+\theta-(1-p)\mu_0 r(1))} + \frac{C(\mu_1+\theta)}{\mu_1(\mu_0+\theta)}$. In this case, for every strategy of the other customers, the tagged customer has a positive expected net benefit if he decides to enter. Hence, $q_0^e = 1$.

In this situation, the expected benefit U_1 is given by

$$U_1(1, q_1) = R - \frac{C}{\mu_1 - \lambda q_1} - \frac{C\lambda q_1(\lambda - \mu_0 r(1))}{\mu_1(1 - r(1))(\theta\mu_1(1 - r(1)) + \lambda q_1(\lambda - \mu_0 r(1)))}. \quad \textbf{(62)}$$

Similarly, we obtain the following equilibria of q_1^e:

(1) if $U_1(1, 0) < 0$, then the equilibrium strategy is $q_1^e = 0$.

(2) if $U_1(1, 0) \geq 0$ and $U_1(1, 1) \leq 0$, then there exists a unique q_1^e satisfying $U_1(1, q_1^e) = 0$ and the equilibrium strategy is $q_1^e = x_3$.

(3) if $U_1(1, 1) > 0$, then the equilibrium strategy is $q_1^e = 1$.

By rearranging Cases 1–2, we obtain the results of Theorem 4. This completes the proof. ■

Using (46), we have the mean queue length as a function of joining probabilities:

$$\begin{aligned} L(q_0, q_1) &= \sum_{n=0}^{\infty} n(\pi_{n0} + \pi_{n1}) \\ &= K\left(\frac{r_0}{(1 - r_0)^2} + \frac{\theta}{\mu_1(1 - \rho_1)^2} + \frac{r_0(\theta + p\mu_0 r_0)(1 - r_0\rho_1)}{\mu_1(1 - r_0)^3(1 - \rho_1)^2}\right) \\ &= \frac{K}{(1 - r_0)^2}\left(r_0 + \frac{\theta(1 - r_0)^2 + (\lambda q_0 - \mu_0 r_0)(1 - r_0\rho_1)}{\mu_1(1 - \rho_1)^2}\right), \end{aligned} \quad \textbf{(63)}$$

and the corresponding social benefit for a mixed strategy (q_0, q_1) as

$$\begin{aligned} S_{au}(q_0, q_1) &= \overline{\lambda}R - CL(q_0, q_1) \\ &= \overline{\lambda}R - \frac{CK}{(1 - r_0)^2}\left(r_0 + \frac{\theta(1 - r_0)^2 + (\lambda q_0 - \mu_0 r_0)(1 - r_0\rho_1)}{\mu_1(1 - \rho_1)^2}\right), \end{aligned} \quad \textbf{(64)}$$

where $\overline{\lambda} = \lambda(p_0 q_0 + p_1 q_1)$ is the effective arrival rate.

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

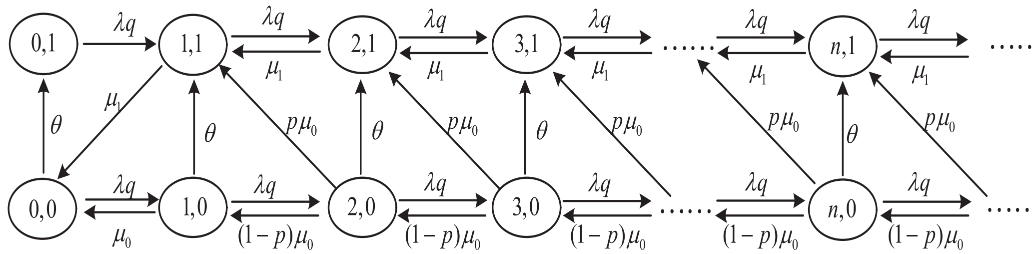

FIGURE 5. Transition rate diagram for fully unobservable queues.

The goal of a social planner is to maximize total social welfare. Let (q_0^*, q_1^*) be the socially optimal mixed strategy. That is,

$$(q_0^*, q_1^*) = \arg \max_{(q_0,q_1)} S_{au}(q_0, q_1).$$

We again use a numerical example to show the discrepancy between customer equilibrium strategy and socially optimal strategy. In a system with $R = 4$, $C = 1$, $\lambda = 0.8$, $\mu_1 = 1$, $\mu_0 = 0.6$, and $\theta = 0.3$, we obtain $(q_0^*, q_1^*) = (0.6221, 0.5242)$. Meanwhile, customers' equilibrium strategy is $(q_0^e, q_1^e) = (1, 0.7621)$. We observe that two types of strategies are inconsistent. This ordering of $q_0^* < q_0^e$ or $q_1^* < q_1^e$ is consistent with the one in a classical unobservable queue situation.

6. FULLY UNOBSERVABLE QUEUES

In this section, an unobservable queue case, where customers cannot observe either the state of the server ($J(t)$ or the number of customers in the system ($N(t)$). In this situation, a pure or mixed strategy can be described by a fraction q ($0 \le q \le 1$), which is the probability of joining the system. The effective arrival rate, or joining rate, is λq. Again, state $(N(t), J(t))$ is a Markov chain with state space $\Omega_{fu} = \{(n, i) \mid n \ge 0,\ i = 0, 1\}$ and the transition diagram is shown in Figure 5.

The stationary distribution of the system when all customers follow a given strategy q can be obtained by simply setting $q_0 = q_1 = q$ in the partially observable queue case. Hence, we have

$$\begin{cases} \pi_{n0} = Kr^n, \quad n \ge 0, \\ \pi_{n1} = K\left(\dfrac{\theta}{\mu_1}\rho_1^{n-1} + \dfrac{(\theta + p\mu_0 r)r}{\mu_1(1-r)} \displaystyle\sum_{j=0}^{n-1} r^j \rho_1^{n-1-j}\right), \quad n \ge 1, \end{cases} \tag{65}$$

where

$$r = \frac{\lambda q + \mu_0 + \theta - \sqrt{(\lambda q + \mu_0 + \theta)^2 - 4\lambda q(1-p)\mu_0}}{2(1-p)\mu_0}, \tag{66}$$

$$K = (1 - \rho_1)(1 - r)K_1, \tag{67}$$

$$K_1 = \left(\frac{\lambda q + \theta}{\lambda q}(1 - r) + \left(1 - \frac{\mu_0}{\mu_1}\right) r\right)^{-1}. \tag{68}$$

From (65), we obtain the probability-generating function of the queue length, denoted by $L(z)$, as follows:

$$\begin{aligned}L(z) &= \sum_{n=0}^{\infty} z^k(\pi_{n0}+\pi_{n1}) \\ &= K\left(\frac{1}{1-rz}+\frac{\theta}{\lambda q}+\frac{(\theta+p\mu_0 r)r}{\mu_1(1-r)}\frac{z}{(1-rz)(1-\rho_1 z)}+\frac{\theta}{\mu_1}\frac{z}{1-\rho_1 z}\right) \\ &= K_1\frac{1-\rho_1}{1-\rho_1 z}\left(\frac{1-r}{1-rz}(1-\rho_1 z)+\frac{\theta}{\lambda q}(1-r)(1-\rho_1 z)\right) \\ &\quad + K_1\frac{1-\rho_1}{1-\rho_1 z}\left(\frac{(\theta+p\mu_0 r)r}{\mu_1(1-r)}\frac{(1-r)z}{(1-rz)(1-\rho_1 z)}+\frac{\theta z}{\mu_1}(1-r)\right).\end{aligned} \quad \textbf{(69)}$$

Using

$$\frac{(\theta+p\mu_0 r)r}{\mu_1(1-r)} = \rho_1-\frac{\mu_0}{\mu_1}r, \quad \frac{1-r}{1-rz}(1-\rho_1 z) = 1-r+(r-\rho_1)\frac{z(1-r)}{1-rz}.$$

in (69), $L(z)$ can be rewritten as

$$L(z) = K_1\frac{1-\rho_1}{1-\rho_1 z}\left(\frac{\lambda q+\theta}{\lambda q}(1-r)+r\left(1-\frac{\mu_0}{\mu_1}\right)\frac{(1-r)z}{1-rz}\right). \quad \textbf{(70)}$$

Equation (70) implies the stochastic decomposition property in this working vacation model. The mean number of customers in the system, denoted by $E[L]$, can be obtained as

$$E[L] = \frac{\rho_1}{1-\rho_1}+K_1\left(1-\frac{\mu_0}{\mu_1}\right)\frac{r}{1-r}. \quad \textbf{(71)}$$

Hence, the mean sojourn time of a joining customer can be obtained by using Little's law:

$$W(q) = \frac{1}{\mu_1-\lambda q}+K_1\left(1-\frac{\mu_0}{\mu_1}\right)\frac{r}{\lambda q(1-r)}. \quad \textbf{(72)}$$

Taking first-order derivatives of (66) and (68), we have $dr/dq>0$ and

$$\frac{dK_1}{dq} = \left(\frac{\lambda q+\theta}{\lambda q}(1-r)+\left(1-\frac{\mu_0}{\mu_1}\right)r\right)^{-2}\left(\frac{\theta}{\lambda q^2}(1-r)+\left(\frac{\mu_0}{\mu_1}+\frac{\theta}{\lambda q}\right)\frac{dr}{dq}\right)>0. \quad \textbf{(73)}$$

Then, we obtain

$$\frac{dW}{dq} = \frac{\lambda}{(\mu_1-\lambda q)^2}+\frac{dK_1}{dq}\left(1-\frac{\mu_0}{\mu_1}\right)\frac{r}{\lambda q(1-r)}+K_1\left(1-\frac{\mu_0}{\mu_1}\right)\frac{\lambda r(1-r+\lambda q\frac{dr}{dq})}{(\lambda q(1-r))^2}>0. \quad \textbf{(74)}$$

Therefore, in the unobservable queue case with $\lambda<\mu_1$, $W(q)$ is strictly increasing in $q\in[0,1]$. If a tagged customer decides to enter the system, his expected net benefit is

$$\begin{aligned}U(q) &= R-CW(q) \\ &= R-\frac{C}{\mu_1-\lambda q}-CK_1\left(1-\frac{\mu_0}{\mu_1}\right)\frac{r}{\lambda q(1-r)},\end{aligned} \quad \textbf{(75)}$$

which is strictly decreasing in q. Thus, $U(q)=0$ has a unique root q_e^*. Consequently, we have the following theorem.

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

THEOREM 5: *In the unobservable queue, the condition $\lambda q < \mu_1$ holds, there exists a unique equilibrium joining strategy q_e, where q_e is given by*

$$q_e = \min\{q_e^*, 1\}. \tag{76}$$

The social-benefit per time unit can be computed as easily computed as

$$\begin{aligned} S_{fu}(q) &= \lambda q R - CE[L] \\ &= \lambda q R - \frac{C\lambda q}{\mu_1 - \lambda q} - CK_1\left(1 - \frac{\mu_0}{\mu_1}\right)\frac{r}{(1-r)}. \end{aligned} \tag{77}$$

Let x^* be the root of the equation $S'_{fu}(q) = 0$ and let q^* be the optimal joining probability. Then, we have the following conclusions: if $0 < x^* < 1$ and $S''_{fu}(q) \leq 0$, then $q^* = x^*$; if $0 < x^* < 1$ and $S''_{fu}(q) > 0$ or $x^* \geq 1$, then $q^* = 1$.

When $R = 4$, $C = 1$,$\lambda = 0.8$, $\mu_1 = 1$, $\mu_0 = 0.6$, and $\theta = 0.3$, then we obtain $q^* = 0.556$. While customers' equilibrium strategy is $q_e = 0.882$. We observe that $q^* < q_e$, which indicates that individual optimization leads queues to be longer than socially desired. Therefore, it is clear that the social planner would like a toll to discourage arrivals in this case.

REMARKS 1: *When $p = 0$, the system studied in this paper becomes a queuing system with a single working vacation. All results for the single working vacation model can be obtained by setting $p = 0$ in our model. Furthermore, when $p = 1$, the system studied in this paper becomes a queuing system with a single working vacation and vacation interruptions which was studied by Doo [9]. Again by setting $p = 1$ in our model, the results are completely consistent with Doo [9].*

7. NUMERICAL EXAMPLES

In this section, we present numerical results to illustrate the effects of the system parameters on customer equilibrium and socially optimal strategies for fully observable, partially observable, and unobservable cases.

For the customer equilibrium strategy in a fully observable queue, while one threshold, $n_{\mathrm{e}}(1)$, has a very simple and explicit expression, the other threshold can only be obtained by solving a complicated equation for the unique root (Theorem 1). Figure 6 shows how these four thresholds change with μ_1 (regular service rate) in a working vacation system with Bernoulli interruptions. It is noticed that all thresholds $n_{\mathrm{e}}(1)$, $n_{\mathrm{e}}(0)$, n_u, n_l increase with μ_1 and the equilibrium threshold n_{e} of the partially observable cases is always between $n_{\mathrm{e}}(1)$ and $n_{\mathrm{e}}(0)$ in the fully observable case. It is intuitive that as the regular service rate increases, arriving customers are more likely to join the system. Figure 7 illustrates how $n_{\mathrm{e}}(0)$ varies with θ (vacation rate) at different interruption probabilities. As θ or p increases, the working vacation duration will decrease, thus customers arriving during a working vacation period are more likely to join.

Figure 8 shows the relation between customer equilibrium strategy and total arrival rate, λ, in a partially observable queue or an unobservable queue. The intuitive decreasing relation reflects the fact that for a given service capacity and a reward-cost structure, the total joining rate is bounded. In Figure 9, the relations between equilibrium joining probabilities and the service reward R for both partially observable and unobservable cases are illustrated. While q_0^e and q_e are increasing in R, q_1^e changes with R in a non-monotonic way. This non-monotonic relation is not entirely intuitive. At the beginning, since the joining

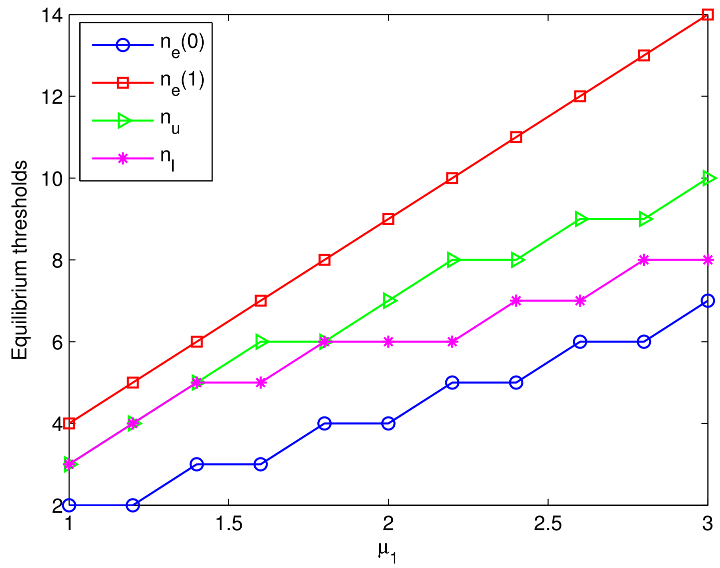

FIGURE 6. Equilibrium threshold strategies for the observable case when $R = 5$, $C = 1$, $\lambda = 1$, $\mu_0 = 0.5$, $\theta = 0.1$, $p = 0.7$.

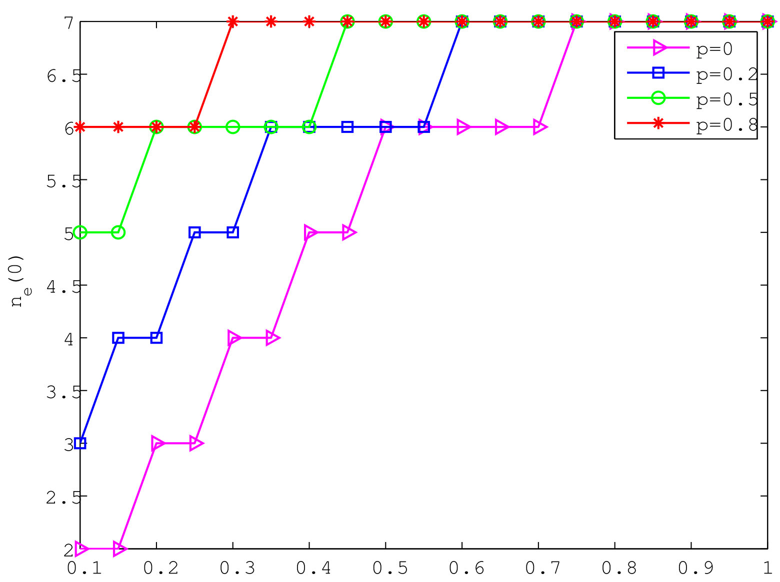

FIGURE 7. Equilibrium threshold strategies for the observable case when $R = 5$, $C = 1$, $\lambda = 1$, $\mu_1 = 2$, $\mu_0 = 0.5$.

probability is small and the system is not congested (low traffic intensity), an increase in R will attract more customers to join (i.e., joining probability increases with R). As the joining probability keeps increasing, the waiting cost is increasing at a faster rate than the linear increase in R, so the joining probability starts to decrease. However, the system is stable $(\lambda = 0.7, \mu_1 = 1)$. Thus, when R keeps increasing to very large value, the reward becomes

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

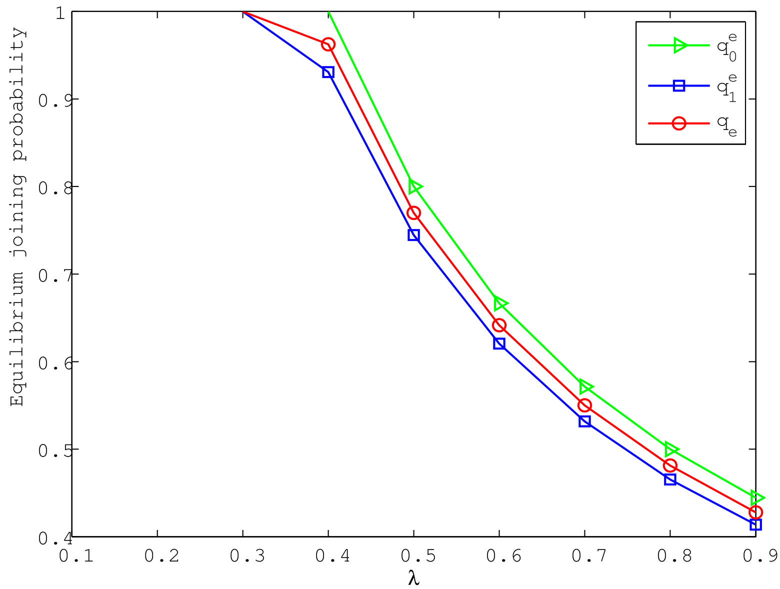

FIGURE 8. Equilibrium joining strategies for the unobservable case when $R = 1$, $C = 1$, $\mu_1 = 1$, $\mu_0 = 0.6$, $\theta = 0.3$, and $p = 0.5$.

dominate again, and the joining probability becomes increasing again. Hence, the relation between q_1 and R is non-monotonic. Furthermore, three functions intersect at one point. At this point, denoted by R^*, the joining probability for the partially observable case is the same as the unobservable case, indicating that the server state information does not affect customer's joining strategy. If $R < R^*$, $q_1^e > q_0^e$, otherwise $q_1^e < q_0^e$. Thus, for higher service value cases, joining probability during a working vacation is higher than that during the regular service period. Figures 10 and 11 show how customer equilibrium strategies depend on service rates, μ_1 and μ_0. The increasing relations between joining probabilities and μ_1 are very intuitive. However, the relations between joining probabilities and μ_0 are not intuitive for the partially observable case. Again, three joining probability functions intersect at one point similarly as in Figure 9. In particular, the relation between q_1^e and μ_0 is non-monotonic with a single minimum. Such a relation is not intuitive but can be explained as follows: Over the lower value range, as μ_0 increases, the increasing rate of q_0^e is quite high. However, for a given service capacity and reward-cost structure, the overall joining probability $q_0^e p_0 + q_1^e p_1$ should be bounded. Thus, q_1^e will decrease over this value range of μ_0 in which q_0^e is quite high. When μ_0 increases to the high value range where q_0^e's increasing rate becomes smaller, q_1^e will start to increase.

Figures 12 and 13 show how equilibrium joining probabilities change with the vacation rate and interruption probability, respectively, for partially observable and unobservable cases. Similar patterns to Figure 10 are shown for these relations. This is because increasing θ or p implies decreasing the mean working vacation duration (i.e., low service rate period). Hence, q_e or $q_0^e p_0 + q_1^e p_1$ (overall joining rate in partially observable case) should be increasing in θ or p although q_1^e is decreasing in θ or p. Similarly, the socially optimal strategies have the same changing trend for p as shown in Figure 14.

Finally, Figures 15 and 16 reveal how equilibrium social benefits and their corresponding social benefits change with the vacation interruption probability p. It is observed that as p increases (i.e., minimizing customer waiting has higher priority than completing the

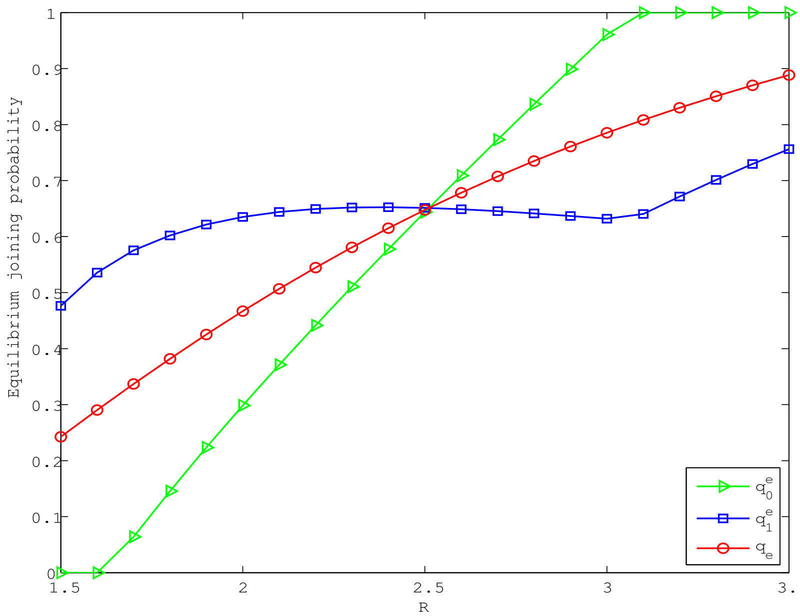

FIGURE 9. Equilibrium joining strategies for the unobservable case when $C = 1$, $\lambda = 0.7$, $\mu_1 = 1$, $\mu_0 = 0.5$, $\theta = 0.3$, and $p = 0.2$.

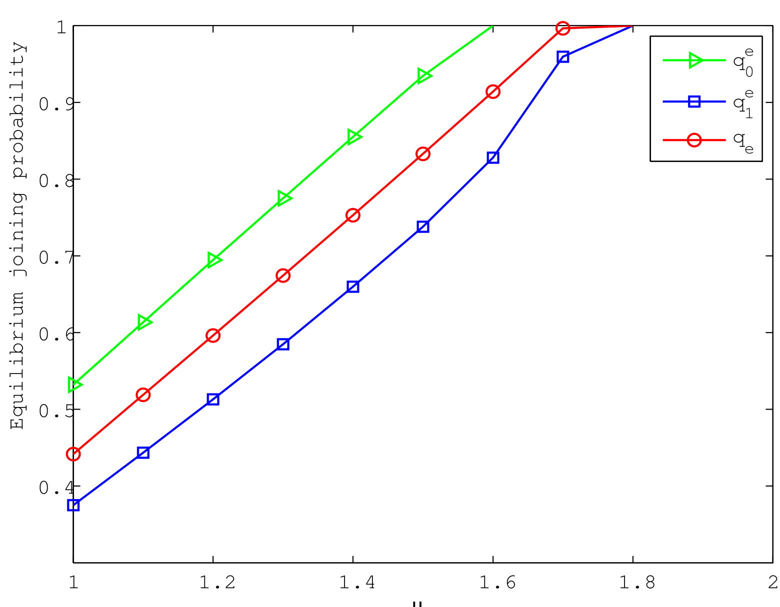

FIGURE 10. Equilibrium joining strategies for the unobservable case when $R = 2$, $C = 1$, $\lambda = 0.9$, $\mu_0 = 0.6$, $\theta = 0.3$, and $p = 0.9$.

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

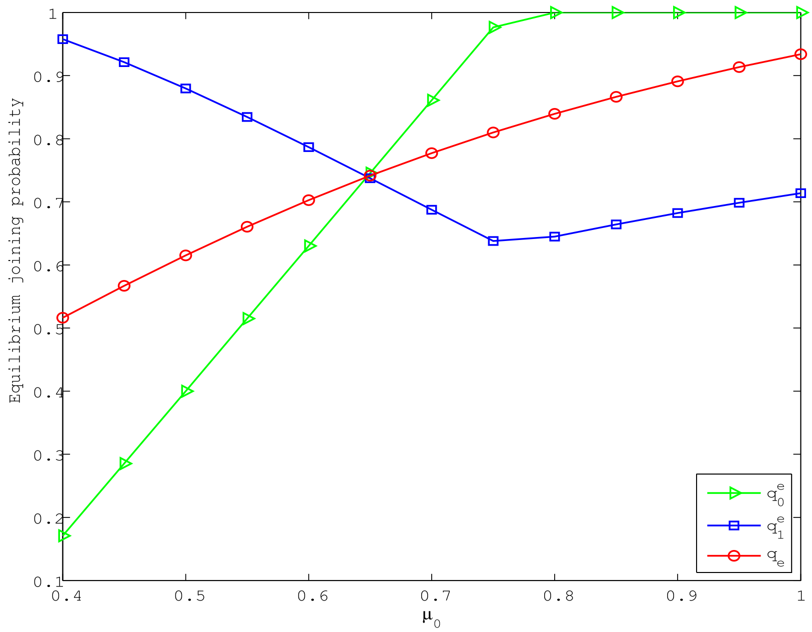

FIGURE **11.** Equilibrium joining strategies for the unobservable case when $R = 2$, $C = 1$, $\lambda = 0.7$, $\mu_1 = 1.2$, $\theta = 0.3$, and $p = 0.3$.

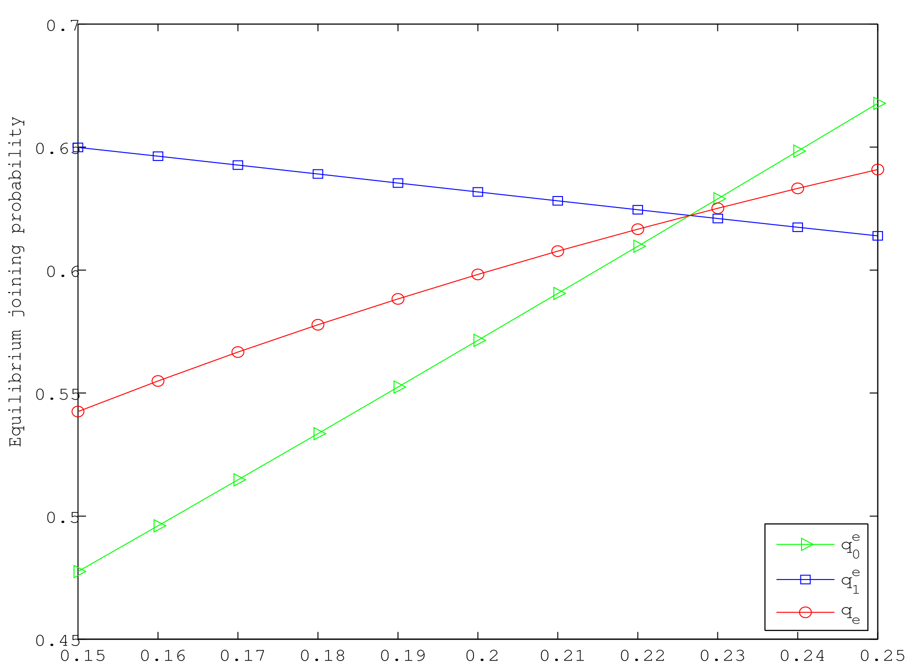

FIGURE **12.** Equilibrium joining strategies for the unobservable case when $R = 2.5$, $C = 1$, $\lambda = 0.8$, $\mu_1 = 1$, $\mu_0 = 0.5$, and $p = 0.5$.

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

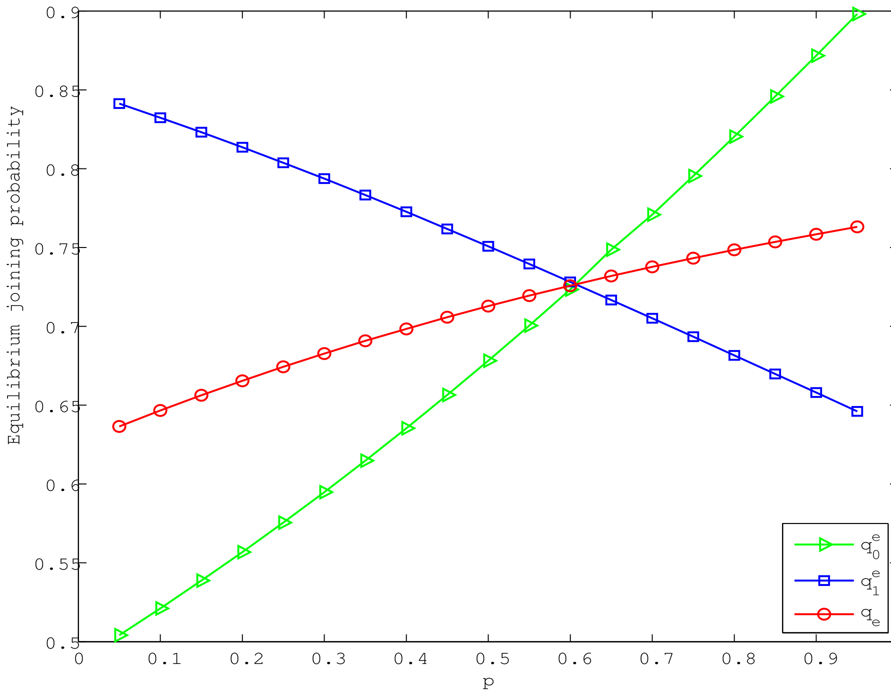

FIGURE **13.** Equilibrium joining strategies for the unobservable case when $R = 2$, $C = 1$, $\lambda = 0.8$, $\mu_1 = 1.3$, $\mu_0 = 0.6$, and $\theta = 0.3$.

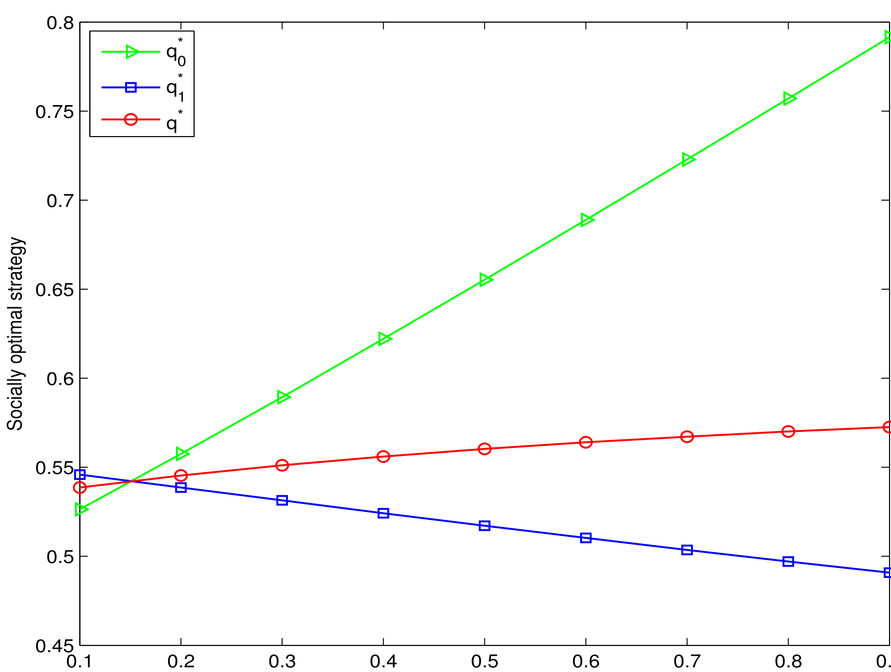

FIGURE **14.** Socially optimal strategies for the unobservable case with various p when $R = 4$, $C = 1$, $\lambda = 0.8$, $\mu_1 = 1$, $\mu_0 = 0.6$, and $\theta = 0.3$.

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

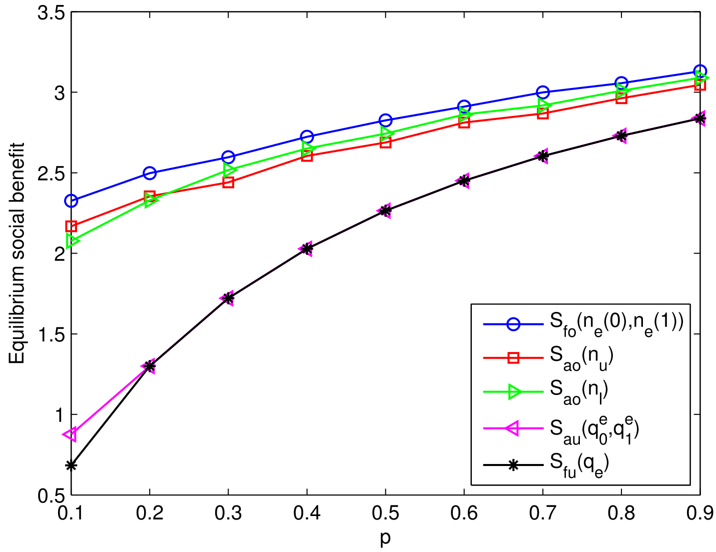

FIGURE 15. Equilibrium social benefit with various p when $R = 5$, $C = 1$, $\lambda = 1.2$, $\mu_1 = 2$, $\mu_0 = 0.5$, and $\theta = 0.15$.

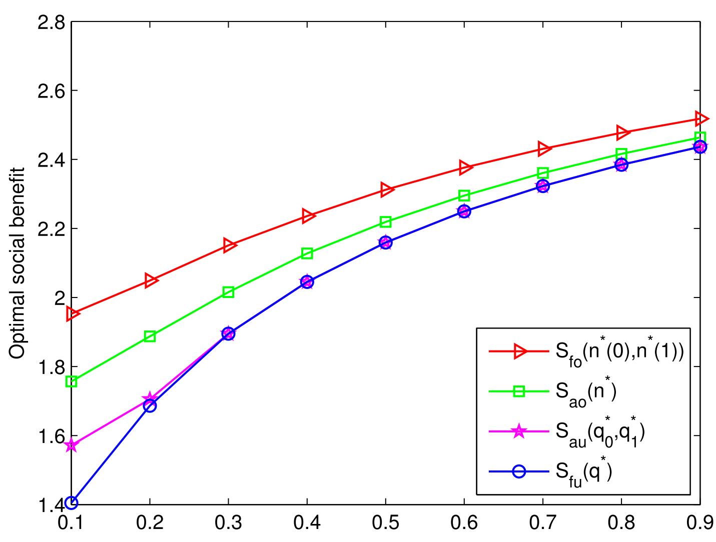

FIGURE 16. Optimal social benefit with various p when $R = 5$, $C = 1$, $\lambda = 0.8$, $\mu_1 = 2$, $\mu_0 = 0.6$, and $\theta = 0.12$.

working vacation), the system should admit more customers into the system. In addition, the additional social benefit of offering the server state information is more significant when p is getting larger as shown in Figures 15 and 16.

https://doi.org/10.1017/S0269964820000704 Published online by Cambridge University Press

8. CONCLUSION

In this study, we analyzed customers' strategic behavior in the M/M/1 queueing system with a single working vacation and Bernoulli interruptions where arriving customers can decide whether to join the system or balk. Four different levels of information disclosed to arriving customers were considered. The customer equilibrium strategies for each information scenario were derived. We compared the equilibrium strategy with the socially optimal strategy numerically for each case and observed that customers tend to overuse the system if they follow the equilibrium strategies. From the perspective of social planners, a toll may be adopted to discourage customers from joining a queue. Moreover, we investigated the effects of system parameters on equilibrium strategies and socially optimal strategies in the almost unobservable and fully unobservable cases.

Based on this work, addressing the issue of using subsidy or price to induce socially optimal strategies under different information scenarios can be a topic of future research. Another direction for future work is to consider non-Markovian queues with various vacation policies.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 11601469) and the Natural Sciences and Engineering Research Council of Canada (No. RGPIN197319).

References

1. Borgs, C., Chayes, J.T., Doroudi, S., Harchol-Balter, M., & Xu, K. (2014). The optimal admission threshold in observable queues with state dependent pricing. *Probability in the Engineering and Informational Sciences* 28: 101–119.
2. Burnetas, A. & Economou, A. (2007). Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing System* 56: 213–228.
3. Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., & Knuth, D.E. (1996). On the Lambert W function. *Advances in Computational Mathematics* 5: 329–359.
4. Economou, A. & Kanta, S. (2008). Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. *Operations Research Letters* 36: 696–699.
5. Gao, S. & Liu, Z. (2013). An M/G/1 queue with single working vacation and vacation interruption under Bernoulli schedule. *Applied Mathematical Modelling* 37: 1564–1579.
6. Guo, P. & Hassin, R. (2011). Strategic behavior and social optimization in Markovian vacation queues. *Operations Research* 59: 986–997.
7. Guo, P. & Hassin, R. (2012). Strategic behavior and social optimization in Markovian vacation queues: the case of heterogeneous customers. *European Journal of Operational Research* 222: 278–286.
8. Lee, D.H. (2017). Optimal pricing strategies and customers' equilibrium behavior in an unobservable M/M/1 queueing system with negative customers and repair. *Mathematical Problems in Engineering* 2017. Article ID 8910819, 11 pages.
9. Lee, D.H. (2018). Equilibrium balking strategies in Markovian queues with a single working vacation and vacation interruption. *Quality Technology and Quantitative Management*. doi:10.1080/16843703.2018.1429805.
10. Li, J., Tian, N., & Ma, Z. (2008). Performance analysis of GI/M/1 queue with working vacations and vacation interruption. *Applied Mathematical Modelling* 32: 2715–2730.
11. Liu, J. & Wang, J. (2017). Strategic joining rules in a single server Markovian queue with Bernoulli vacation. *Operational Research* 17: 413–434.
12. Naor, P. (1969). The regulation of queue size by levying tolls. *Econometrica* 37: 1524.
13. Neuts, M. (1981). *Matrix-geometric solution in stochastic models*. Baltimore: Johns Hopkins University Press.
14. Sun, W. & Li, S. (2014). Equilibrium and optimal behavior of customers in Markovian queues with multiple working vacations. *TOP* 22: 694–715.
15. Sun, W., Guo, P., & Tian, N. (2010). Equilibrium threshold strategies in observable queueing systems with setup/closedown times. *Central European Journal of Operations Research* 18: 241–268.

16. Sun, W., Wang, Y., & Tian, N. (2012). Pricing and setup/closedown policies in unobservable queues with strategic customers. *4OR - A Quarterly Journal of Operations Research* 36: 287–311.
17. Sun, W., Li, S., & Li, Q. (2014). Equilibrium balking strategies of customers in Markovian queues with two-stage working vacations. *Applied Mathematics and Computation* 248: 195–214.
18. Sun, W., Li, S., & Tian, N. (2017). Equilibrium and optimal balking strategies of customers in unobservable queues with double adaptive working vacations. *Quality Technology and Quantitative Management* 14: 94–113.
19. Tao, L., Liu, Z., & Wang, Z. (2011). The GI/M/1 queue with start-up period and single working vacation and Bernoulli vacation interruption. *Applied Mathematics and Computation* 218: 4401–4413.
20. Tao, L., Wang, Z., & Liu, Z. (2013). The GI/M/1 queue with Bernoulli-schedule-controlled vacation and vacation interruption. *Applied Mathematical Modelling* 37: 3724–3735.
21. Tian, N., Li, J., & Zhang, Z.G. (2009). Matrix analysis method and working vacation queues: a survey. *International Journal of Information and Management Sciences* 20: 603–633.
22. Tian, R., Yue, D., & Yue, W. (2015). Optimal balking strategies in an M/G/1 queueing system with a removable server under N-policy. *Journal of Industrial and Management Optimization* 11: 715–731.
23. Tian, R., Hu, L., & Wu, X. (2016). Equilibrium and optimal strategies in M/M/1 queues with working vacations and vacation interruptions. *Mathematical Problems in Engineering* 2016. Article ID 9746962.
24. Wang, F., Wang, J., & Zhang, F. (2014). Equilibrium customer strategies in the Geo/Geo/1 queue with single working vacation. *Discrete Dynamics in Nature and Society* 2014. Article ID 309489.
25. Yu, S., Liu, Z., & Wu, J. (2016). Equilibrium strategies of the unobservable M/M/1 queue with balking and delayed repairs. *Applied Mathematics and Computation* 290: 56–65.
26. Yu, S., Liu, Z., & Wu, J. (2017). Strategic behavior in the partially observable Markovian queues with partial breakdowns. *Operations Research Letters* 45: 471–474.
27. Zhang, F., Wang, J., & Liu, B. (2013). Equilibrium balking strategies in Markovian queues with working vacations. *Applied Mathematical Modelling* 37: 8264–8282.

APPENDIX

Solving the stationary probabilities for fully observable queue case

From (14), it is easy to obtain that $\{\pi_{n0} \mid 1 \leq n \leq n(0)\}$ are solutions of the homogeneous linear difference equation

$$(1-p)\mu_0 x_{n+1} - (\lambda+\mu_0+\theta)x_n + \lambda x_{n-1} = 0, \quad n = 1,2,\ldots,n(0), \tag{A.1}$$

and its corresponding characteristic equation $(1-p)\mu_0 x^2 - (\lambda+\mu_0+\theta)x + \lambda = 0$ has two roots:

$$x_{1,2} = \frac{(\lambda+\mu_0+\theta) \pm \sqrt{(\lambda+\mu_0+\theta)^2 - 4\lambda(1-p)\mu_0}}{2(1-p)\mu_0}, \tag{A.2}$$

Then, the homogeneous solution of (18) is $x_n = A_1 x_1^n + B_1 x_2^n,\ n = 0,1,\ldots,n(0)+1$, where A_1 and B_1 are constants to be determined. Using (13) and (15), we have

$$\begin{cases} A_1(\lambda+\theta-\mu_0 x_1) + B_1(\lambda+\theta-\mu_0 x_2) = \mu_1 \pi_{11}, \\ A_1((\mu_0+\theta)x_1^{n(0)+1} - \lambda x_1^{n(0)}) + B_1((\mu_0+\theta)x_2^{n(0)+1} - \lambda x_2^{n(0)}) = 0. \end{cases} \tag{A.3}$$

Then, we derive

$$\begin{cases} A_1 = \dfrac{\mu_1((\mu_0+\theta)x_2^{n(0)+1} - \lambda x_2^{n(0)})}{(\lambda+\theta-\mu_0 x_1)((\mu_0+\theta)x_2^{n(0)+1} - \lambda x_2^{n(0)}) - (\lambda+\theta-\mu_0 x_2)((\mu_0+\theta)x_1^{n(0)+1} - \lambda x_1^{n(0)})}\pi_{11}, \\ B_1 = \dfrac{\mu_1((\mu_0+\theta)x_1^{n(0)+1} - \lambda x_1^{n(0)})}{(\lambda+\theta-\mu_0 x_2)((\mu_0+\theta)x_1^{n(0)+1} - \lambda x_1^{n(0)}) - (\lambda+\theta-\mu_0 x_1)((\mu_0+\theta)x_2^{n(0)+1} - \lambda x_2^{n(0)})}\pi_{11}. \end{cases} \tag{A.4}$$

Thus,

$$\pi_{n0} = A_1 x_1^n + B_1 x_2^n, \quad n = 0, 1, \ldots, n(0) + 1. \tag{A.5}$$

Then, we consider the probability $\{\pi_{n1} \,|1 \le n \le n(0)\}$. From (17), we find that the probability is equivalent to solutions of the non-homogeneous linear difference equation

$$\begin{aligned} \mu_1 x_{n+1} - (\lambda + \mu_1) x_n + \lambda x_{n-1} &= -\theta \pi_{n0} - p\mu_0 \pi_{n+1,0} \\ &= -(\theta + p\mu_0 x_1) A_1 x_1^n - (\theta + p\mu_0 x_2) B_1 x_2^n, \quad n = 1, 2, \ldots, n(0). \end{aligned} \tag{A.6}$$

Hence, its corresponding characteristic equation $\mu_1 x^2 - (\lambda + \mu_1)x + \lambda = 0$ has two roots at 1 and $\rho = \lambda/\mu_1$. Assume that $\rho \ne 1$; then the homogeneous solution of (A.6) is $x_n^{\text{hom}} = A_2 1^n + B_2 \rho^n$. The general solution of the non-homogeneous equation is given as $x_n^{\text{gen}} = x_n^{\text{hom}} + x_n^{\text{spec}}$, where x_n^{spec} is a specific solution. We find a specific solution $x_n^{\text{spec}} = Cx_1^n + Dx_2^n$. Substituting it into (A.6), we derive

$$\begin{cases} C = \dfrac{A_1 x_1 (\theta + p\mu_0 x_1)}{(1 - x_1)(\mu_1 x_1 - \lambda)}, \\ D = \dfrac{B_1 x_2 (\theta + p\mu_0 x_2)}{(1 - x_2)(\mu_1 x_2 - \lambda)}. \end{cases} \tag{A.7}$$

Therefore, the general solution of the non-homogeneous equation (A.6) is given as

$$x_n^{\text{gen}} = A_2 + B_2 \rho^n + C x_1^n + D x_2^n, \quad n = 1, 2, \ldots, n(0) + 1. \tag{A.8}$$

Taking account of (17), we obtain

$$\begin{cases} A_2 + B_2 \rho = \pi_{11} - (Cx_1 + Dx_2), \\ A_2 \mu_1 + B_2 \mu_1 \rho^2 = (\lambda + \mu_1)\pi_{11} - \lambda \pi_{01} - \theta \pi_{10} - p\mu_0 \pi_{20} - \mu_1 (Cx_1^2 + Dx_2^2). \end{cases} \tag{A.9}$$

Solving the equations, we obtain

$$\begin{cases} A_2 = \dfrac{\mu_1 \pi_{11} + \lambda (Cx_1 + Dx_2) - \mu_1 (Cx_1^2 + Dx_2^2) - \lambda \pi_{01} - \theta \pi_{10} - p\mu_0 \pi_{20}}{\mu_1 - \lambda}, \\ B_2 = \dfrac{\mu_1 (Cx_1^2 + Dx_2^2 - Cx_1 - Dx_2) + \lambda \pi_{01} + \theta \pi_{10} + p\mu_0 \pi_{20} - \lambda \pi_{11}}{\lambda (1 - \rho)}. \end{cases} \tag{A.10}$$

Thus, from (A.8), we obtain

$$\pi_{n1} = A_2 + B_2 \rho^n + C x_1^n + D x_2^n, \quad n = 1, 2, \ldots, n(0) + 1. \tag{A.11}$$

Finally, we compute the probability $\{\pi_{n1} | n(0) + 1 \le n \le n(1) + 1\}$. From (16), we find that they are solutions of the homogeneous linear difference equation of (A.6). Therefore, the homogeneous solution is $x_n = A_3 + B_3 \rho^n$, $n = n(0) + 1, n(0) + 2, \ldots, n(1) + 1$, where A_3 and B_3 are constants to be determined. Using (20) and (A.11), we have

$$\begin{cases} \mu_1 (A_3 + B_3 \rho^{n(1)+1}) = \lambda (A_3 + B_3 \rho^{n(1)}), \\ A_3 + B_3 \rho^{n(0)+2} = A_2 + B_2 \rho^{n(0)+2} + C x_1^{n(0)+2} + D x_2^{n(0)+2}. \end{cases} \tag{A.12}$$

Then, we derive

$$\begin{cases} A_3 = 0, \\ B_3 = \dfrac{A_2 + B_2 \rho^{n(0)+2} + C x_1^{n(0)+2} + D x_2^{n(0)+2}}{\rho^{n(0)+2}}, \end{cases} \tag{A.13}$$

Additionally,

$$\pi_{n1} = B_3 \rho^n, \quad n = n(0) + 1, n(0) + 2, \ldots, n(1) + 1. \tag{A.14}$$

Thus, we have all the stationary probabilities in terms of π_{11}. The remaining probability, π_{11}, can be found from the normalization condition.