Equilibrium balking strategies of customers in Markovian queues with two-stage working vacations

Wei Sun, Shiyong Li *, Quan-Lin Li
School of Economics and Management, Yanshan University, Qinhuangdao 066004, China

ARTICLE INFO

Keywords:
Markovian queues
Two-stage working vacations
Equilibrium balking strategy
Information precision
Social welfare

ABSTRACT

This paper studies the customers’ equilibrium balking behavior in some single-server Markovian queues with two-stage working vacations. That is, the server starts taking two successive working vacations when the system becomes empty, during which he provides low-rate service but maintains different service rates in the two-stage vacations. Based on different precision levels of system information, we discuss observable queues, partially observable queues and unobservable queues, respectively. For each type of queues, we get the customers’ equilibrium balking strategy and equilibrium social welfare per time unit, and numerically observe that their positive equilibrium strategy is unique. Especially, for partially observable queues, the customers’ equilibrium joining probabilities in vacation states are not necessarily smaller than that in busy state.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

For the work studying on customers’ equilibrium balking strategies in queues with classical vacation policies, Burnetas and Economou [1] first presented several Markovian queues with setup times and four precision levels of system information: fully observable, almost observable, almost unobservable and fully unobservable. Then Economou and Kanta [2] discussed an observable queue with server breakdowns. For the fully and partially observable queues, Liu et al. [10] introduced classical single vacation policy, whereas Wang and Zhang [21] and Li et al. [9] focused on server breakdowns and delayed repairs. Then Sun and Li [16] obtained customers’ equilibrium balking strategies with five types of decision criteria in some unobservable queues with multiple vacations. On the other hand, for discrete-time queueing systems, Ma et al. [12] investigated some Geo/Geo/1 queues with multiple vacations. Besides customers’ equilibrium balking strategies, there are also some work concerning socially optimal ones. Sun et al. [15,18] considered fully observable and unobservable queues with several types of setup/closedown policies: interruptible, skippable and insusceptible policies, respectively. Moreover, Guo and Hassin [4,5] elaborately studied fully observable and unobservable queues with homogeneous and heterogeneous customers under N-policy, respectively. Then Economou et al. [3] further discussed the unobservable and partially observable queues with general service and vacation times.

Different from the work with various classical vacation policies above, recently, Sun and Li [17] and Zhang et al. [22] studied the customers’ equilibrium or socially optimal balking strategies in queues with multiple working vacations under different information levels. Moreover, Wang et al. [20] considered customers’ equilibrium strategies in some discrete-time queues with single working vacation. Actually, detailed description of working vacation policy is originally given by Servi http://dx.doi.org/10.1016/j.amc.2014.09.116

* Corresponding author.
E-mail addresses: wsun@ysu.edu.cn (W. Sun), shiyongli@ysu.edu.cn (S. Li), liquanlin@tsinghua.edu.cn (Q.-L. Li).
and Finn [14] and various system performance indices can be consulted in the survey given by Tian and Zhang [19]. In addition, besides [17,20,22], in the other related literature, the most commonly studied policies are also multiple or single working vacation policy, and service rate during a whole vacation period is always uniform. Using matrix analytic method, Li et al. [8] and Liu et al. [11] analyzed stationary queue length distribution and stochastic decomposition results of an Markovian queue and an M/G/1 queue with multiple exponentially distributed working vacations, respectively. Then Li and Tian [7] considered a GI/M/1 queue with single working vacation and derived various performance measures.

However, in some service systems, the server takes neither single nor multiple working vacations, but takes $n$-stage working vacations, where $n$ is a fixed number and $n \geq 2$. During a whole vacation period, the server provides lower-rate service by himself or by his replacements. Moreover, in view of system managers’ specific demands, the server needs to adjust his service rates (or his replacements may be changed) among different vacation states, so service rate during each vacation time can be distinct with each other. One case is that in order to avoid overstocking too many customers at the end of each vacation time, service rate can be arranged to increase progressively with $n$. Alternatively, it can also be arranged to decrease progressively with $n$ if the manager wants to equally attract customers even at the earlier-stage vacation times. To explain the model, here we give a simple real-life application. For example, there are a chef (the owner) and several $(n)$ assistants in a famous snack bar in a small town, and it can also provide takeout to the customers. During the chef’s short vacation period, the assistants, as reserve cooks, take over the duties of the chef in turn. In view of their different levels of cooking skill, they can adjust their order according to the queue length or other factors. In this paper, we assume the original case $n = 2$. That is, we study some queues with two-stage working vacations, where the server provides low-rate service but maintains different service rates in the two successive working vacation times, respectively.

The main objective of our work is to study the customers’ equilibrium balking strategies in some single-server Markovian queues with two-stage exponentially distributed working vacations. When customers arrive at the system, in the light of their acquired different precision levels of system information, they need to make decisions of whether to join the queue or not. Here we study three types of queueing systems: the observable queues, the partially observable queues and the unobservable queues. For the observable queues, we first derive the customers’ equilibrium balking thresholds both in server’s two-stage working vacation states and busy state. In equilibrium, the order of balking thresholds in the two-stage working vacations depends on the relation of service rates in the three states. Then in view of the relation of balking thresholds in the first-stage and second-stage working vacations, we get the stationary queue length distributions and equilibrium social welfare. For the partially observable queues, we derive the customers’ equilibrium joining probabilities and numerically find that their positive equilibrium mixed strategy is unique. Moreover, the customers’ equilibrium joining probabilities in vacation states are not necessarily lower than that in busy state, which depends on the values of the service rates. In addition, we also obtain the stationary queue length distribution and equilibrium social welfare. Finally, for the unobservable queues, we easily get the customers’ equilibrium mixed strategy and equilibrium social welfare based on the results given in the partially observable case.

The paper is organized as follows: In Section 2, we describe the queueing models and different information precision levels. Sections 3–5 are devoted to the observable queues, the partially observable queues and the unobservable queues, respectively. For each type of queue, we derive the customers’ equilibrium balking strategies and equilibrium social welfare. In Section 6, we briefly conclude the paper.

2. Model description

In this paper, we consider some single-server Markovian queueing systems with two-stage working vacations. Assume that customers’ potential arrival rate is $\lambda$ and the server’s regular service rate is $\mu_0$. Once the system becomes empty, the server begins a first-stage working vacation $V_1$, which follows an exponential distribution with parameter $\theta$. During this period, an arriving customer is served at a service rate $\mu_{n}$, which is lower than $\mu_0$. As soon as the first-stage vacation finishes, the server continues to take a second-stage working vacation $V_2$, which follows the same distribution with $V_1$, no matter whether there are customers in the system or not, and then switches service rate from $\mu_{n}$ to $\mu_{2}$, which is also lower than $\mu_0$. After finishing the second-stage vacation, a regular busy period starts if there are still customers waiting in the queue. Otherwise, the server stays idle until one customer arrives. So we denominate this type of queues as $\text{M}/\text{M}/1/(T-S)$ WV queues.

Let $(N(t), I(t))$ represent the system state at time $t$, where $N(t)$ denotes the system occupancy, i.e., the number of customers in the system, and $I(t)$ denotes the server state at time $t$, and

$$
I(t) = \begin{cases} 
0 & \text{the server is busy or stays idle;} \\
1 & \text{the server is taking a first-stage vacation;} \\
2 & \text{the server is taking a second-stage vacation.}
\end{cases}
$$

So the observable case means that arriving customers can observe system information of both $N(t)$ and $I(t)$, and the partially observable case means they can observe $I(t)$ but not $N(t)$, whereas the unobservable case means they can observe neither $N(t)$ nor $I(t)$.

Let us mark an arriving customer, and he can receive a reward $R$ after service completion but has to bear a cost $C$ for waiting a time unit. We adopt a linear cost function, so his expected net benefit, denoted by $U$, is $U = R − CE[W]$, where
$E[W]$ represents his expected sojourn time in the system. Obviously, $U = 0$ if he selects balking. We assume that the inter-arrival times, service times, and the two-stage working vacation times are mutually independent, and the service discipline is first in first out (FIFO).

3. The observable queues

We first consider the customers’ equilibrium balking behavior in the observable case, i.e., arriving customers can observe both $N(t)$ and $l(t)$ at time $t$. Denote the sojourn time of the marked customer who joins the system in case he encounters the system state $(n, i)$ ($i = 0, 1, 2$) as $W_i(n)$, and its mean as $E[W_i(n)]$, and his expected net benefit after service completion as $U_i(n)$. So the customers’ equilibrium balking threshold at state $i$ can be denoted by $n_i(i)$. Now we need to find the customers’ equilibrium integrated balking threshold strategy $(n_0(0), n_0(1), n_0(2))$.

For $E[W_0(n)]$, it is obviously equal to $n + 1$ service times at the regular service rate $\mu_b$, that is,

$$E[W_0(n)] = \frac{n + 1}{\mu_b}. \quad (3.1)$$

On the other hand, in order to derive $E[W_i(n)]$ ($i = 1, 2$), first we need to discuss three cases at state $(n, 1)$: If the residual time of the first-stage working vacation, denoted by $V_{1R}$, is long enough for service completion of $n + 1$ customers, his sojourn time $W_i(n)$ of the marked customer equals $n + 1$ service times at the service rate $\mu_{v_1}$; if there are only $j$ ($0 \leq j \leq n$) customers completing service during $V_{1R}$ and other $n + 1 - j$ customers completing service during the second-stage working vacation $V_{2R}$, $W_i(n)$ equals the sum of $V_{1R}$ and $n + 1 - j$ service times at the service rate $\mu_{v_2}$; Otherwise, if there are $j$ ($0 \leq j \leq n$) customers completing service during $V_{1R}$ and $i$ ($0 \leq i \leq n - j$) customers completing service during $V_{2R}$ and other $n + 1 - i - j$ customers completing service during the next busy period, $W_i(n)$ equals the sum of $V_{1R}$ and $V_{2R}$ and $n + 1 - i - j$ service times at the regular service rate $\mu_b$. Hence, based on the Lemmas 1 and 2 given by Sun and Li [17], we get the Laplace–Stieltjes transform (LST) of $W_i(n)$, denoted by $\tilde{W}_i(s)$, as follows:

$$\tilde{W}_i(s) = P\{ S_{i1}^{(n-1)} < V_{1R} \} \left( \frac{\mu_{v_1} + \theta}{\mu_{v_1} + \theta + s} \right)^{n-1} + \sum_{j=0}^{n} P\{ S_{i1}^{(j)} < V_{1R} < S_{i1}^{(n-j)} \cap S_{i2}^{(n-j)} < V_{2R} \} \right. \right.$$

$$\times \left. \left( \frac{\mu_{v_1} + \theta}{\mu_{v_1} + \theta + s} \right)^{j-1} \frac{\mu_{v_2} + \theta}{\mu_{v_2} + \theta + s} \right)^{n-j} + \sum_{j=0}^{n} P\{ S_{i1}^{(j)} < V_{1R} < S_{i1}^{(n-j)} \cap S_{i2}^{(n-j)} < V_{2R} < S_{i2}^{(n-j)} \} \right. \right.$$

$$\times \left. \left( \frac{\mu_{v_1} + \theta}{\mu_{v_1} + \theta + s} \right)^{j-1} \frac{\mu_{v_2} + \theta}{\mu_{v_2} + \theta + s} \right)^{n-j} \frac{\mu_b}{\mu_b + s} = \left( \frac{\mu_{v_1}}{\mu_{v_1} + \theta + s} \right)^{n-1} \sum_{j=0}^{n} P\{ S_{i1}^{(j)} < V_{1R} < S_{i1}^{(n-j)} \cap S_{i2}^{(n-j)} < V_{2R} < S_{i2}^{(n-j)} \} \right. \right.$$

$$= \left( \frac{\mu_{v_1}}{\mu_{v_1} + \theta + s} \right)^{n-1} \sum_{j=0}^{n} \left( \frac{\mu_{v_1}}{\mu_{v_1} + \theta + s} \right)^{j} \left( \frac{\mu_{v_2}}{\mu_{v_2} + \theta + s} \right)^{n-j} \frac{\mu_b}{\mu_b + s}$$

$$+ \frac{\mu_{v_1}^2}{\mu_b + (\mu_{v_1} - \mu_{v_2})^2} \left( \frac{\mu_b}{\mu_b + (\mu_{v_1} - \mu_{v_2})^2} \right)^{n-1} \left( \frac{\mu_{v_1}}{\mu_{v_1} + \theta + s} \right)^{n-1} \left( \frac{\mu_{v_2}}{\mu_{v_2} + \theta + s} \right)^{n-1} \right). \quad (3.2)$$

Then we can easily get $E[W_i(n)] = -\tilde{W}_i(0)$, that is,

$$E[W_i(n)] = \frac{n + 1}{\mu_b} + \frac{2\mu_b - \mu_{v_1} - \mu_{v_2}}{\mu_b \theta} + \frac{\mu_{v_1}^2 - 2\mu_b \mu_{v_1} + \mu_b \mu_{v_2}}{\mu_b \theta (\mu_{v_1} - \mu_{v_2})} \left( \frac{\mu_{v_1}}{\mu_{v_1} + \theta + s} \right)^{n-1} + \frac{\mu_{v_2} (\mu_b - \mu_{v_2})}{\mu_b \theta (\mu_{v_1} - \mu_{v_2})} \left( \frac{\mu_{v_2}}{\mu_{v_2} + \theta + s} \right)^{n-1}. \quad (3.3)$$

Then for the system state $(n, 2)$, we need to discuss two cases: If the residual time of the second-stage working vacation, denoted by $V_{2R}$, is long enough for service completion of $n + 1$ customers, $W_2(n)$ equals $n + 1$ service times at the service rate $\mu_{v_2}$; Otherwise, if there are only $j$ ($0 \leq j \leq n$) customers completing service during $V_{2R}$ and other $n + 1 - j$ customers completing service during the next busy period, $W_2(n)$ equals the sum of $V_{2R}$ and $n + 1 - j$ service times at the regular service rate $\mu_b$. Hence, we get the LST of $W_2(n)$, denoted by $\tilde{W}_2(s)$, as follows:
is depicted in Fig. 1. So state space of the Markovian process is the marked customer's expected net benefit after service completion, denoted by \( \theta \). Based on (3.1), (3.3) and (3.5), we can get the expressions of \( W_2(s) \),

\[
\begin{align*}
\bar{W}_2(s) &= W_2(s) - W_2(0) = E[W_2(n)] - \bar{W}_2(0), \text{ that is,} \\
E[W_2(n)] &= -W_2(0) = \frac{n+1}{\mu_b} - \frac{\mu_b - \mu_{v_2}}{\theta \mu_b} \left( 1 - \left( \frac{\mu_{v_2}}{\mu_{v_2} + \theta} \right) \right)^{n+1}. 
\end{align*}
\]

Based on (3.1), (3.3) and (3.5), we can get the expressions of the customers' expected net benefit after service completion, denoted by \( U_i(n) \), equals 0, i.e., \( U_i(n) = 0 \). Solving \( U_i(n) = 0 \), we can derive the unique feasible root \( n_i(i) \). Hence the customers' equilibrium integrated balking threshold strategy is \( n_i(0), n_i(1), n_i(2) \) for the first case, the equilibrium transition diagram in Fig. 1 shows that initially \( n_i(1) \geq n_i(2) \), and then \( n_i(2) \geq n_i(1) \) with \( \mu_b \). Summarily, we can numerically observe that for all feasible \( n_i, n_i(0) \geq n_i(2) \geq n_i(1) \) when \( \mu_{v_1} < \mu_{v_2} \) or \( \mu_{v_1} \) does not exceed \( \mu_{v_2} \) too much. However, it is not true when \( \mu_{v_1} \) exceeds \( \mu_{v_2} \), too much and \( \mu_{v_1} \) is close to \( \mu_b \).

Next we will derive the stationary queue length distribution in observable queues. In view of the possible relations of \( n_i(1) \) and \( n_i(2) \), we discuss two cases: \( n_i(1) > n_i(2) \) and \( n_i(1) < n_i(2) \). For the first case, the equilibrium transition diagram is depicted in Fig. 1. So state space of the Markovian process \( \{N(t), I(t)\} \) is

\[
\Omega_{ob_1} = \{(n,0): 0 \leq n \leq n_e(0)+1\} \bigcup \{(n,j): 0 \leq n \leq n_e(1)+1, \quad j = 1, 2\}.
\]

1 Under the model assumptions of FCFS service discipline and independence of arrival process and two-stage vacation times, a customer joining in vacation states only gives negative externality to his following customers. So \( U_i(n) \) is surely monotonously decreasing with \( n \).

2 The case of \( n_i(1) = n_i(2) \) can be regarded as the common special case of \( n_i(1) > n_i(2) \) and \( n_i(1) < n_i(2) \).
Denote the stationary queue length distribution as
\[
\pi_{nj} = P\{N = n, l = j\} = \lim_{t \to \infty} P\{N(t) = n, \quad l(t) = j\}, \quad (n, j) \in \Omega_{ob}. 
\]

Then the stationary transition probability equations can be written as
\[
\begin{align*}
\pi_{n0} &= \pi_{02} \lambda; \\
\pi_{n0}(\lambda + \mu_b) &= \pi_{n-10} \lambda + \pi_{n+10} \mu_b + \pi_{n2} \theta; \quad 1 \leq n \leq n_c(1) + 1 \\
\pi_{n0}(\lambda + \mu_b) &= \pi_{n-10} \lambda + \pi_{n+10} \mu_b; \quad n_c(1) + 2 \leq n \leq n_c(0) \\
\pi_{n0}(0,0; \lambda) &= \pi_{n0}(0,0; \lambda); \\
\pi_{01}(\lambda + \theta) &= \pi_{01} \theta + \pi_{11} \mu_{n1}; \\
\pi_{n1}(\lambda + \theta + \mu_{n1}) &= \pi_{n-11} \lambda + \pi_{n+11} \mu_{n1}; \quad 1 \leq n \leq n_c(1) \\
\pi_{n1}(1,1; \theta + \mu_{n1}) &= \pi_{n1}(1,1; \lambda); \\
\pi_{02}(\lambda + \theta) &= \pi_{01} \theta + \pi_{12} \mu_{n2}; \\
\pi_{n2}(\lambda + \theta + \mu_{n2}) &= \pi_{n-12} \lambda + \pi_{n+12} \mu_{n2} + \pi_{n1} \theta; \quad 1 \leq n \leq n_c(2) \\
\pi_{n2}(2,2; \theta + \mu_{n2}) &= \pi_{n2}(2,2; \lambda) + \pi_{n2}(2,2; \theta) + \pi_{n2}(2,2; \mu_{n2}); \\
\pi_{n2}(2,2; \theta + \mu_{n2}) &= \pi_{n2}(2,2; \lambda) + \pi_{n2}(2,2; \theta) + \pi_{n2}(2,2; \mu_{n2}); \\
\pi_{n1}(1,1; \theta + \mu_{n2}) &= \pi_{n1}(1,1; \lambda). 
\end{align*}
\]

We first consider the probabilities \(\{\pi_{nj}\} 0 \leq n \leq n_c(1) + 1\) in the first-stage vacation state. From (3.11), we notice that they are the solutions of the following homogeneous linear difference equation with constant coefficients
\[
\mu_{n1} x_{n+1} - (\lambda + \theta + \mu_{n1}) x_n + \lambda x_{n-1} = 0, \quad n = 1, 2, \ldots, n_c(1). 
\]  
So its corresponding characteristic equation is
\[
\mu_{n1} x^2 - (\lambda + \theta + \mu_{n1}) x + \lambda = 0.
\]
which has two roots
\[
x_{1,2} = \frac{(\lambda + \theta + \mu_{n1}) \pm \sqrt{(\lambda + \theta + \mu_{n1})^2 - 4 \lambda \mu_{n1}}}{2 \mu_{n1}}. 
\]

The general solution of (3.18), denoted by \(x_{1,2}^{\text{hom}}\), is 
\[
x_{n}^{\text{hom}} = A_1 x_{1}^{n} + B_1 x_{2}^{n} 
\]
(Obviously \(x_{1} \neq x_{2}\), where \(A_1, B_1\) are the coefficients to be determined. From (3.10) and (3.12), we obtain
\[
\begin{align*}
A_1 &= \frac{x_{1}^{(n+1)}((\lambda + \theta + \mu_{n1}) x_{1}^{-1}) \pi_{n10}}{x_{1}^{(n+1)}((\lambda + \theta + \mu_{n1}) x_{1}^{-1}) \pi_{n10} - x_{2}^{(n+1)}((\lambda + \theta + \mu_{n1}) x_{2}^{-1}) \pi_{n10}}, \\
B_1 &= \frac{x_{2}^{(n+1)}((\lambda + \theta + \mu_{n1}) x_{2}^{-1}) \pi_{n10}}{x_{1}^{(n+1)}((\lambda + \theta + \mu_{n1}) x_{1}^{-1}) \pi_{n10} - x_{2}^{(n+1)}((\lambda + \theta + \mu_{n1}) x_{2}^{-1}) \pi_{n10}}. 
\end{align*}
\]
Thus,
\[ \pi_{n_1} = A_1 x_1^{n_1} + B_1 x_2^{n_1}, \quad n = 0, 1, \ldots, n_e(1) + 1, \]  
(3.21)
where \( x_1, x_2 \) and \( A_1, B_1 \) are given by (3.19) and (3.20), respectively.

For the second-stage vacation state, we first consider the probabilities \( \{ \pi_{n_2} | n_2(2) \geq n \leq n_2(1) + 1 \} \). From (3.16), they are the solutions of the following nonhomogeneous linear difference equation
\[ \mu_{n_2} x_{n_2} - (\lambda + \mu_{n_2}) x_{n_2} = -\pi_{n_1} \theta = -\lambda (A_1 x_1^{n_1} + B_1 x_2^{n_1}), \quad n = n_2(2) + 2, \ldots, n_e(1) + 1. \]  
(3.22)

The general solution of the homogeneous version of (3.22) is \( x_2^{\text{hom}} = B_3 x_2^{p} \), where
\[ \hat{x}_2 = \frac{\theta + \mu_{n_2}}{\mu_{n_2}}. \]  
(3.23)

Hence the general solution of (3.22), denoted by \( x_2^{\text{gen}} \), is given as \( x_2^{\text{gen}} = x_2^{\text{hom}} + x_2^{\text{spec}} \), where \( x_2^{\text{spec}} \) is a specific solution of (3.22). Because the nonhomogeneous part of (3.22) is geometric with parameter \( x_1 \) and \( x_2 \), we consider a specific solution of the form \( x_2^{\text{spec}} = C_2 x_1^{n_1} + D_2 x_2^{n_1} \). Substituting it into (3.22), we get
\[ \left\{ \begin{array}{l}
C_2 = -\frac{-\lambda}{\mu_{n_2} x_1^{(\theta + \mu_{n_2})^{-1}}}, \\
D_2 = \frac{-\lambda}{\mu_{n_2} x_2^{(\theta + \mu_{n_2})^{-1}}}
\end{array} \right. \]  
(3.24)

Therefore,
\[ x_2^{\text{gen}} = B_3 x_1^{p} + C_2 x_1^{n_1} + D_2 x_2^{n_1}, \quad n = n_2(2) + 2, \ldots, n_e(1) + 1, \]  
(3.25)
where \( B_3 \) is to be determined. Taking into account (3.17), then obtain
\[ B_3 = \frac{\theta (A_1 x_1^{n_1(1) + 1} + B_1 x_2^{n_1(1) + 1}) - (\theta + \mu_{n_2}) (C_2 x_1^{n_1(1) + 1} + D_2 x_2^{n_1(1) + 1})}{(\theta + \mu_{n_2}) x_2^{n_1(1) + 1}}, \]  
(3.26)

thus, from (3.25),
\[ \pi_{n_2} = B_3 x_1^{p} + C_2 x_1^{n_1} + D_2 x_2^{n_1}, \quad n = n_2(2) + 2, \ldots, n_e(1) + 1 \]  
(3.27)
can be obtained, where \( C_2, D_2 \) and \( B_3 \) are given by (3.24) and (3.26), respectively.

Then we consider the probabilities \( \{ \pi_{n_2} | 0 \leq n \leq n_e(2) + 1 \} \). From (3.14), they are the solutions of the following nonhomogeneous linear difference equation
\[ \mu_{n_2} x_{n_2} - (\lambda + \mu_{n_2}) x_0 + \lambda x_{n_2} = -\pi_{n_1} \theta = -\lambda (A_1 x_1^{n_1} + B_1 x_2^{n_1}), \quad n = 1, 2, \ldots, n_e(2). \]  
(3.28)

The general solution of the homogeneous version of (3.28) is \( x_2^{\text{hom}} = A_2 x_1^{p} + B_2 x_2^{p} \), where
\[ \hat{x}_1 = \frac{\theta + \mu_{n_2}}{2 \mu_{n_2}} \pm \sqrt{\frac{(\lambda + \mu_{n_2})^2}{4 \mu_{n_2}^2} - 4 \lambda \mu_{n_2}}. \]  
(3.29)

So the general solution of (3.28) is \( x_2^{\text{gen}} = x_2^{\text{hom}} + x_2^{\text{spec}} \), where \( x_2^{\text{spec}} = C_1 x_1^{n_1} + D_1 x_2^{n_1} \). Substituting it into (3.28), we get
\[ \left\{ \begin{array}{l}
C_1 = \frac{-\lambda}{\mu_{n_2} x_1^{(\theta + \mu_{n_2})^{-1}}}, \\
D_1 = \frac{-\lambda}{\mu_{n_2} x_2^{(\theta + \mu_{n_2})^{-1}}}
\end{array} \right. \]  
(3.30)

Therefore,
\[ x_2^{\text{gen}} = A_2 x_1^{p} + B_2 x_2^{p} + C_1 x_1^{n_1} + D_1 x_2^{n_1}, \quad n = 0, 1, \ldots, n_e(2) + 1, \]  
(3.31)
where \( A_2, B_2 \) are to be determined. Taking into account (3.13), (3.15) and (3.27), then obtain
\[ \left\{ \begin{array}{l}
A_2 = \left( \left( \lambda + \mu_{n_2} \right) - \mu_{n_2} \hat{x}_1 \right) \left( C_1 x_1^{n_1(1) + 1} \left( \lambda + \mu_{n_2} \right) + D_1 x_2^{n_1(1) + 1} \right) + \lambda \left( A_1 x_1^{n_1(1) + 1} + B_1 x_2^{n_1(1) + 1} \right) \\
\quad + \mu_{n_2} \left( B_3 x_1^{n_1(1) + 1} + C_2 x_1^{n_1(1) + 1} + D_2 x_2^{n_1(1) + 1} \right) - \lambda \left( \hat{x}_1 \right) \left( \left( \lambda + \mu_{n_2} \right) x_2^{p} - \hat{x}_1 \right) \left( \left( \lambda + \mu_{n_2} \right) x_2^{p} - \hat{x}_1 \right)
\end{array} \right. \]  
(3.32)
thus, from (3.31),
\[ \pi_{n2} = A_2 x_2^{n+1} + B_2 x_2^n + C_1 x^n_1 + D_1 x_2^n, \quad n = 0, 1, \ldots, n_s(2) + 1 \] (3.33)
can be obtained, where \( A_2, D_1, B_2 \) are given by (3.30) and (3.32), respectively.

Sequentially analyzing the server’s busy state, we first derive the probabilities \( \{ \pi_{n0} \} \) \( 0 \leq n \leq n_s(2) + 1 \). From (3.7), they are the solutions of the following nonhomogeneous linear difference equation
\[ \mu_b x_{n+1} - (\lambda + \mu_b) x_n + i x_{n-1} = -\pi_{n2} \theta - \theta (A_2 x_2^n + B_2 x_2^n + C_1 x^n_1 + D_1 x_2^n), \quad n = 1, 2, \ldots, n_s(2). \] (3.34)
So the general solution of (3.34) is \( x_n^{A_2} = x_n^{\text{hom}} + x_n^{\text{spec}}, \) where \( x_n^{\text{hom}} = A_4 + B_4 \rho^n \) and \( x_n^{\text{spec}} = E_1 x_1^n + F_1 x_1^n + C_3 x^n_1 + D_3 x_2^n. \) Substituting it into (3.34), we get
\[
\begin{align*}
E_1 &= \frac{-d_1 x_1^n}{\mu_b x_1^n - (\lambda + \mu_b) x_1^n + i x_1^{\text{spec}}}, \\
F_1 &= \frac{-d_2 x_2^n}{\mu_b x_2^n - (\lambda + \mu_b) x_2^n + i x_2^{\text{spec}}}, \\
C_3 &= -\frac{d_3 x_2^n}{\mu_b x_2^n - (\lambda + \mu_b) x_2^n + i x_2^{\text{spec}}}, \\
D_3 &= \frac{-d_4 x_2^n}{\mu_b x_2^n - (\lambda + \mu_b) x_2^n + i x_2^{\text{spec}}},
\end{align*}
\] (3.35)
Therefore,
\[ x_n^{A_2} = A_4 + B_4 \rho^n + E_1 x_1^n + F_1 x_1^n + C_3 x^n_1 + D_3 x_2^n, \quad n = 0, 1, \ldots, n_s(2) + 1, \] (3.36)
where \( A_4, B_4 \) are to be determined. Taking into account (3.6), then obtain
\[
\begin{align*}
A_4 &= \left( \frac{\pi_{n0}(n-1) - \pi_{n0}(n-2) - \pi_{n0}(n-3) + \pi_{n0}(n-4)}{\rho^{n+2} - \rho^n} \right), \\
B_4 &= \left( \frac{\pi_{n0}(n-1) - \pi_{n0}(n-2) - \pi_{n0}(n-3) + \pi_{n0}(n-4)}{\rho^{n+2} - \rho^n} \right),
\end{align*}
\] (3.37)
thus, from (3.36),
\[ \pi_{n0} = A_5 + B_5 \rho^n + E_1 x_1^n + F_1 x_1^n + C_4 x^n_1 + D_4 x_2^n, \quad n = 0, 1, \ldots, n_s(2) + 1, \] (3.38)
can be obtained, where \( E_1, F_1, C_3, D_4 \) are given by (3.35) and \( A_5, B_4 \) are given by (3.37), respectively.

Then we derive the probabilities \( \{ \pi_{n0} \} \) \( 2 \leq n \leq n_s(1) + 1 \). They are the solutions of the following nonhomogeneous linear difference equation
\[ \mu_b x_{n+1} - (\lambda + \mu_b) x_n + i x_{n-1} = -\pi_{n2} \theta - \theta (B_5 x_2^n + C_2 x^n_1 + D_2 x_2^n), \quad n = n_s(2) + 2, \ldots, n_s(1) + 1. \] (3.39)
So the general solution of (3.39) is \( x_n^{B_5} = x_n^{\text{hom}} + x_n^{\text{spec}}, \) where \( x_n^{\text{hom}} = A_5 + B_5 \rho^n \) and \( x_n^{\text{spec}} = F_2 x_2^n + C_4 x^n_1 + D_4 x_2^n. \) Substituting it into (3.39), we get
\[
\begin{align*}
F_2 &= \frac{-d_2 x_2^n}{\mu_b x_2^n - (\lambda + \mu_b) x_2^n + i x_2^{\text{spec}}}, \\
C_4 &= -\frac{d_3 x_2^n}{\mu_b x_2^n - (\lambda + \mu_b) x_2^n + i x_2^{\text{spec}}}, \\
D_4 &= \frac{-d_4 x_2^n}{\mu_b x_2^n - (\lambda + \mu_b) x_2^n + i x_2^{\text{spec}}},
\end{align*}
\] (3.40)
Therefore,
\[ x_n^{B_5} = A_5 + B_5 \rho^n + F_2 x_2^n + C_4 x^n_1 + D_4 x_2^n, \quad n = n_s(2) + 2, \ldots, n_s(1) + 1, \] (3.41)
where \( A_5, B_5 \) are to be determined. Taking into account (3.38), then obtain
\[
\begin{align*}
A_5 &= A_4 + \frac{E_1 x_1^{n+1} + F_1 x_1^{n+1} + C_3 x_1^{n+1} + D_3 x_2^{n+1}}{\mu_b x_1^n - (\lambda + \mu_b) x_1^n + i x_1^{\text{spec}} + \frac{x_1^{n+1} - \rho (x_1^n - \rho)}{\rho^{n+1} - \rho^-} (D_3 - D_4)}, \\
B_5 &= B_4 + \frac{E_1 x_1^{n+1} + F_1 x_1^{n+1} + C_3 x_1^{n+1} + D_3 x_2^{n+1}}{\mu_b x_1^n - (\lambda + \mu_b) x_1^n + i x_1^{\text{spec}} + \frac{x_1^{n+1} - \rho (x_1^n - \rho)}{\rho^{n+1} - \rho^-} (D_3 - D_4)}
\end{align*}
\] (3.42)
thus, from (3.41),
\[ \pi_{n0} = A_5 + B_5 \rho^n + F_2 x_2^n + C_4 x^n_1 + D_4 x_2^n, \quad n = n_s(2) + 2, \ldots, n_s(1) + 1 \] (3.43)
can be obtained, where \( F_2, C_4, D_4 \) are given by (3.40) and \( A_5, B_5 \) are given by (3.42), respectively.

Finally, we derive the probabilities \( \{ \pi_{n0} \} \) \( 1 \leq n \leq n_s(0) + 1 \). From (3.8), we get \( x_n^{A_6} = A_6 + B_6 \rho^n, \) where \( A_6, B_6 \) are to be determined. Taking into account (3.9) and (3.43), we obtain \( A_6 = 0 \) and
\[ B_6 = A_5 + B_5 \rho^n + F_2 x_2^n + C_4 x^n_1 + D_4 x_2^n \] (3.44)
Therefore,
\[ \pi_{n0} = B_0 \rho^n, \quad n = n_e(1) + 2, \ldots, n_e(0) + 1, \]
where \( B_0 \) are given by (3.44).

In summary, we find that the obtained stationary state probabilities \( \{ \pi_{nj}(n,j) \in \Omega_{ob} \} \) are all related to \( \pi_{10} \). Using the normalization condition, we get the results in the following theorem.

**Theorem 3.1.** For the observable Markovian queue with two-stage working vacations and state space \( \Omega_{ob} = \{(n,0) : 0 \leq n \leq n_e(0) + 1 \} \cup \{(n,j) : 0 \leq n \leq n_e(1) + 1, j = 1,2 \} \), where \( (n_e(0), n_e(1), n_e(2)) \) are the customers’ equilibrium threshold strategies, the stationary queue length distribution \( \{ \pi_{nj}(n,j) \in \Omega_{ob} \} \) is:

\[ \pi_{n1} = A_1 x_1^n + B_1 x_2^n; \quad n = 0, 1, \ldots, n_e(1) + 1 \]
\[ \pi_{n2} = A_2 x_1^n + B_2 x_2^n + C_1 x_1^n + D_1 x_2^n; \quad n = 0, 1, \ldots, n_e(2) + 1 \]
\[ \pi_{n2} = B_3 x_1^n + C_2 x_1^n + D_2 x_2^n; \quad n = n_e(2) + 2, \ldots, n_e(1) + 1 \]
\[ \pi_{n0} = A_4 + B_4 \rho^n + E_1 x_1^n + F_1 x_2^n + C_3 x_1^n + D_3 x_2^n; \quad n = 0, 1, \ldots, n_e(2) + 1 \]
\[ \pi_{n0} = A_5 + B_5 \rho^n + F_2 x_2^n + C_4 x_1^n + D_4 x_2^n; \quad n = n_e(2) + 2, \ldots, n_e(1) + 1 \]
\[ \pi_{n0} = B_6 \rho^n; \quad n = n_e(1) + 2, \ldots, n_e(0) + 1 \]

where \( \rho = \lambda / \mu_h \neq 1 \) and \( x_1^i, x_2^j, \mu_i, A_i (i = 1,2,4,5), B_i (1 \leq j \leq 6), C_i, D_i (1 \leq k \leq 4), E_1, F_1, F_2 \) are given by (3.19), (3.20), (3.23), (3.24), (3.26), (3.29), (3.30), (3.32), (3.35), (3.37), (3.40), (3.42), (3.44), respectively, and \( \pi_{10} \) can be solved by the normalization equation:

\[ \sum_{(n,j) \in \Omega_{ob}} \pi_{nj} = 1. \] (3.52)

Based on Fig. 3 and Theorem 3.1, an arriving customer will select balking when he finds the system is at state \( (n_e(0) + 1,0), (n_e(1) + 1,1) \) or \( (n_e(2) + 1 \leq n \leq n_e(1) + 1) \). So the equilibrium social welfare per time unit, denoted by \( U_s(n_e(0), n_e(1), n_e(2)) \), equals

\[ U_s(n_e(0), n_e(1), n_e(2)) = \lambda R \left( 1 - \pi_{n_e(0) + 10} - \pi_{n_e(1) + 11} - \sum_{n=n_e(2)+1}^{n_e(1)+1} \pi_{n2} \right) - C \left( \sum_{n=1}^{n_e(0)+1} n \pi_{n0} + \sum_{n=1}^{n_e(1)+1} n \pi_{n1} + \pi_{n2} \right). \] (3.53)

As for the second case \( n_e(1) < n_e(2) \), the equilibrium transition diagram is depicted in Fig. 4. So the state space of the Markovian process \( \{ N(t), I(t) \} \) is

\( \Omega_{ob} = \{(n,j) : 0 \leq n \leq n(j) + 1, j = 0,1,2 \}. \)

![Fig. 3. Equilibrium transition rate diagram for observable queues in case \( n_e(1) > n_e(2) \).](image-url)
Fig. 4. Equilibrium transition rate diagram for observable queues in case $n_e(1) < n_e(2)$.

Fig. 5. Equilibrium social welfare for case $n_e(1) > n_e(2)$ when $R = 10, C = 1, \mu_b = 3, \mu_{p_1} = 2.5, \mu_{p_2} = 1, \theta = 0.2$.

Then the stationary transition probability equations can be written as

\[
\pi_{00}\lambda = \pi_{02}\theta; \\
\pi_{0n}(\lambda + \mu_b) = \pi_{n-10}\lambda + \pi_{n+10}\mu_b + \pi_{n2}\theta; \quad 1 \leq n \leq n_e(2) + 1 \\
\pi_{0n}(\lambda + \mu_b) = \pi_{n-10}\lambda + \pi_{n+10}\mu_b; \quad n_e(2) + 2 \leq n \leq n_e(0) \\
\pi_{0n}(0,10)\mu_b = \pi_{n0}(0)\lambda; \\
\pi_{01}(\lambda + \theta) = \pi_{10}\mu_b + \pi_{11}\mu_{p_1}; \\
\pi_{n1}(\lambda + \theta + \mu_{p_1}) = \pi_{n-11}\lambda + \pi_{n+11}\mu_{p_1}; \quad 1 \leq n \leq n_e(1) \\
\pi_{n1}(1,11)(\theta + \mu_{p_1}) = \pi_{n1}(1)\lambda; \\
\pi_{02}(\lambda + \theta) = \pi_{01}\theta + \pi_{12}\mu_{p_2}; \\
\pi_{n2}(\lambda + \theta + \mu_{p_2}) = \pi_{n-12}\lambda + \pi_{n+12}\mu_{p_2} + \pi_{n1}\theta; \quad 1 \leq n \leq n_e(1) + 1 \\
\pi_{n2}(\lambda + \theta + \mu_{p_2}) = \pi_{n-12}\lambda + \pi_{n+12}\mu_{p_2}; \quad n_e(1) + 2 \leq n \leq n_e(2) \\
\pi_{n2}(2,12)\mu_{p_2} = \pi_{n2}(2)\lambda. 
\]

Using the similar analysis method with that in the first case, we can also get the stationary queue length distribution $\{\pi_{nj} | (n, j) \in \Omega_{nk}\}$ for the second case. Based on Fig. 4, the customers’ balking states are $(n_e(0) + 1, 0), (n_e(1) + 1, 1)$ or $(n_e(2) + 1, 2)$. So the equilibrium social welfare per time unit equals

\[W(n_e(0), n_e(1), n_e(2)).\]
Us(ne(0), ne(1), ne(2)) = kR1/C0 + \frac{10}{C0} + \frac{11}{C1/C0/C0} C \times ne(0) + \frac{1}{C0} + \frac{1}{C1/C0/C0} C \times ne(1) = q0, q1, q2 \times ne(2).

Figs. 5 and 6 show the tendency of the equilibrium social welfare with \lambda for the cases ne(1) > ne(2) and ne(1) < ne(2), respectively. We observe that Us(ne(0), ne(1), ne(2)) first gradually increases and achieves a maximum, then decreases as \lambda continues increasing no matter whichever case. The reason for this behavior is that the system is rarely crowded and joining customers have no need to wait long time for service when \lambda is small, which makes the social welfare improve. However, customers face longer waiting delay and more waiting cost if they decide to join as \lambda keeps increasing, which has a negative effect on the equilibrium social welfare.

4. The partially observable queues

Next we consider the partially observable case, that is, arriving customers only know the server’s state I(t) but cannot observe the system occupancy N(t) at time t. So the customers’ decision problem can be expressed by a set of joining probabilities (q0, q1, q2) (0 \leq q_i \leq 1, i = 0, 1, 2), and their equilibrium mixed strategy is denoted by (q_e(0), q_e(1), q_e(2)).

In order to derive (q_e(0), q_e(1), q_e(2)), we first try to get the stationary queue length distribution. The transition diagram is depicted in Fig. 7. Based on Fig. 7, we can obviously observe that {N(t), I(t)} is a quasi-birth-and-death (QBD) process with the state space

\Omega_{po} = \{(k, j) : k \geq 0, \ j = 0, 1, 2\}.
If \( \rho = q_0/\mu_b < 1 \), let \((N,I)\) be the stationary limit of the QBD process \( \{N(t),I(t)\} \). Denote the stationary queue length distribution as

\[
\pi = (\pi_0, \pi_1, \pi_2, \ldots), \quad \pi_k = (\pi_{k0}, \pi_{k1}, \pi_{k2}), \quad k \geq 0,
\]

\[
\pi_{ij} = P(N = k, I = j) = \lim_{t \to \infty} P(N(t) = k, \ I(t) = j), \quad (k,j) \in \Omega_{po}
\]

and the infinitesimal generator of the process as \( Q \), then we get the stationary transition probability equations \( \pi Q = 0 \) as follows

\[
\pi_{00} \lambda q_0 = \pi_{02} \theta;
\]

\[
\pi_{k0} (\lambda q_0 + \mu_b) = \pi_{k-10} \lambda q_0 + \pi_{k2} \theta + \pi_{k10} \mu_b, \quad k \geq 1;
\]

\[
\pi_{01} (\lambda q_1 + \theta) = \pi_{10} \mu_b + \pi_{11} \mu_{\nu_1};
\]

\[
\pi_{k1} (\lambda q_1 + \theta + \mu_{\nu_1}) = \pi_{k-11} \lambda q_1 + \pi_{k11} \mu_{\nu_1}, \quad k \geq 1;
\]

\[
\pi_{02} (\lambda q_2 + \theta) = \pi_{01} \theta + \pi_{12} \mu_{\nu_2};
\]

\[
\pi_{k2} (\lambda q_2 + \theta + \mu_{\nu_2}) = \pi_{k-12} \lambda q_2 + \pi_{k22} \mu_{\nu_2} + \pi_{k12} \theta, \quad k \geq 1.
\]

So using the lexicographical ordering for the system states, \( Q \) can be written in the block-diagonal form as follows:

\[
Q = \begin{pmatrix}
A_{00} & C \\
B_{10} & A & C \\
B & A & C \\
\end{pmatrix},
\]

where

\[
A_{00} = \begin{pmatrix}
-\lambda q_0 & 0 & 0 \\
0 & -\theta & \lambda q_1 \\
\theta & 0 & -\theta - \lambda q_2
\end{pmatrix}, \quad B_{10} = \begin{pmatrix}
0 & \mu_b & 0 \\
0 & \mu_{\nu_1} & 0 \\
0 & 0 & \mu_{\nu_2}
\end{pmatrix};
\]

\[
B = \begin{pmatrix}
\mu_b & 0 & 0 \\
0 & \mu_{\nu_1} & 0 \\
0 & 0 & \mu_{\nu_2}
\end{pmatrix}, \quad C = \begin{pmatrix}
\lambda q_0 & 0 & 0 \\
0 & \lambda q_1 & 0 \\
0 & 0 & \lambda q_2
\end{pmatrix};
\]

\[
A = \begin{pmatrix}
-\lambda q_0 - \mu_b & 0 & 0 \\
0 & -\lambda q_1 - \theta - \mu_{\nu_1} & \theta \\
\theta & 0 & -\lambda q_2 - \theta - \mu_{\nu_2}
\end{pmatrix}.
\]

Now we need to solve the following matrix quadratic equation

\[
R^2 B + RA + C = 0
\]

for its minimal non-negative solution which is called the rate matrix \( R \). Based on the special structure of \( A, B, C \), we conclude that \( R \) has the similar form. So we assume

\[
R = \begin{pmatrix}
r_{11} & 0 & 0 \\
r_{21} & r_{22} & r_{23} \\
r_{31} & 0 & r_{33}
\end{pmatrix},
\]

Substituting it into (4.8), we get the set of equations

\[
\begin{align}
(r_{11} \mu_b - r_{11} (\lambda q_0 + \mu_b) + \lambda q_0) &= 0, \\
(r_{21} r_{11} + r_{22} r_{21} + r_{23} r_{31}) \mu_b - r_{21} (\lambda q_0 + \mu_b) + r_{23} \theta &= 0, \\
(r_{22} \mu_{\nu_1} - r_{22} (\theta + \lambda q_1 + \mu_{\nu_1}) + \lambda q_1) &= 0, \\
(r_{22} r_{23} + r_{23} r_{33}) \mu_{\nu_2} + r_{22} \theta - r_{23} (\theta + \lambda q_2 + \mu_{\nu_2}) &= 0, \\
(r_{31} r_{11} + r_{33} r_{31}) \mu_b - r_{31} (\lambda q_0 + \mu_b) + r_{33} \theta &= 0, \\
(r_{33} \mu_{\nu_2} - r_{33} (\theta + \lambda q_2 + \mu_{\nu_2}) + \lambda q_2) &= 0.
\end{align}
\]
Solving (4.10), we get

\[
\begin{align*}
 r_{11} &= \frac{\lambda_0}{\mu_0}, \\
r_{21} &= \mu_0 \frac{2\mu_1 r_1 r_2}{r_1 r_2}, \\
r_{22} &= r_1 r_2, \\
r_{23} &= r_1 r_2, \\
r_{31} &= \mu_0 \left( \frac{\theta^2 \lambda_j}{\lambda_0} + \lambda_j \right) \left( r_1 r_2 - 2 \mu_1 r_1 + \mu_2 \right), \\
r_{33} &= \frac{\lambda_0}{r_1 r_2},
\end{align*}
\]

(4.11)

where

\[
\begin{align*}
 r_{q_1} &= \theta + \lambda q_1 + \mu r_1 - \sqrt{(\theta + \lambda q_1 + \mu r_1)^2 - 4\lambda q_1 \mu r_1}, \\
r_{q_2} &= \theta + \lambda q_2 + \mu r_2 - \sqrt{(\theta + \lambda q_2 + \mu r_2)^2 - 4\lambda q_2 \mu r_2}, \\
r_{q_3} &= \theta + \lambda q_3 + \mu r_3 + \sqrt{(\theta + \lambda q_3 + \mu r_3)^2 - 4\lambda q_3 \mu r_3}.
\end{align*}
\]

Using the matrix analysis method (see [6,13]), we have

\[
\pi_k = (\pi_{00}, \pi_{01}, \pi_{02}) = (\pi_{00}, \pi_{01}, \pi_{02}) R^k, \quad k \geq 0.
\]

(4.12)

and \((\pi_{00}, \pi_{01}, \pi_{02}) (RB_{10} + A_{00}) = 0\),

(4.13)

and the normalization condition is

\[
(\pi_{00}, \pi_{01}, \pi_{02}) (I - R)^{-1} e = 1.
\]

(4.14)

where \(I\) is identity matrix, \(e\) is unit vector, and

\[
RB_{10} + A_{00} = \begin{pmatrix}
-\lambda q_0 & r_{11} \mu_0 & 0 \\
0 & r_{21} \mu_0 + r_{22} \mu r_1 - \theta - \lambda q_1 & \theta + r_{23} \mu r_2 \\
\theta & r_{31} \mu_0 & r_{33} \mu_2 - \theta - \lambda q_2
\end{pmatrix}.
\]

(4.15)

Substituting (4.15) into (4.13), we obtain

\[
\begin{align*}
-\lambda q_0 \pi_{00} + \theta \pi_{02} &= 0, \\
r_{11} \mu_0 \pi_{00} + (r_{21} \mu_0 + r_{22} \mu r_1 - \theta - \lambda q_1) \pi_{01} + r_{31} \mu_0 \pi_{02} &= 0, \\
(\theta + r_{23} \mu r_2) \pi_{01} + (r_{33} \mu_2 - \theta - \lambda q_2) \pi_{02} &= 0.
\end{align*}
\]

(4.16)

Solving (4.16), we get

\[
(\pi_{00}, \pi_{01}, \pi_{02}) = \pi_{00} \left(1, \frac{\lambda q_0 \left(\theta + \lambda q_2 - r_{33} \mu_2\right)}{\theta (\theta + r_{23} \mu r_2),} \lambda q_0 \right).
\]

Furthermore,

\[
R^k = \begin{pmatrix}
 r_{11}^k & 0 & 0 \\
r_{21}^k \sum_{j=0}^{k-1} r_{11}^{k-1-j} r_{22}^j & r_{22}^k & r_{23}^k \sum_{j=0}^{k-1} r_{22}^{k-1-j} \\
r_{31}^k \sum_{j=0}^{k-1} r_{11}^{k-1-j} & r_{33}^k & r_{33}^k
\end{pmatrix}, \quad k \geq 1,
\]
Substituting \((\pi_{00}, \pi_{01}, \pi_{02})\) and \(R^k\) into (4.12), we can obtain the stationary queue length distribution as follows:

\[
\begin{align*}
\pi_{00} &= \frac{\rho_{00}}{\rho_{00} + \rho_{10}(\rho_{20} + \rho_{21} + \rho_{11} + \rho_{12} + \rho_{13}) + \rho_{20} + \rho_{21} + \rho_{11} + \rho_{12} + \rho_{13})}, \quad k = 0, \\
\pi_{01} &= \frac{\rho_{01}}{\rho_{00} + \rho_{10}(\rho_{20} + \rho_{21} + \rho_{11} + \rho_{12} + \rho_{13}) + \rho_{20} + \rho_{21} + \rho_{11} + \rho_{12} + \rho_{13})}, \quad k = 0, \\
\pi_{02} &= \frac{\rho_{02}}{\rho_{00} + \rho_{10}(\rho_{20} + \rho_{21} + \rho_{11} + \rho_{12} + \rho_{13}) + \rho_{20} + \rho_{21} + \rho_{11} + \rho_{12} + \rho_{13})}, \quad k = 0.
\end{align*}
\]

(4.17)

where \(\pi_{00}\) can be determined by (4.14).

According to (4.17), the probability that the system is in state \(i (i = 0, 1, 2)\), denoted by \(p_i\), can be derived as follows:

\[
\begin{align*}
p_0 &= P(I = 0) = \sum_{k=0}^{\infty} \pi_{00} = \pi_{00} \left( \frac{\theta - r_{23} \mu_{p_2}}{(1 - r_{11})(\theta + r_{23} \mu_{p_2})} - \frac{\lambda q_0 r_{21}(\theta + r_{23} \mu_{p_2})}{(1 - r_{11})(1 - r_{22})(\theta + r_{23} \mu_{p_2})} + \frac{\lambda q_0 r_{31}(\theta - r_{33} \mu_{p_3})}{(1 - r_{11})(1 - r_{22})(\theta + r_{23} \mu_{p_2})} \right), \\
p_1 &= P(I = 1) = \sum_{k=0}^{\infty} \pi_{01} = \frac{\lambda q_0(\theta + 2 \theta q_2 - r_{33} \mu_{p_3})}{\theta(1 - r_{22})(\theta + r_{23} \mu_{p_2})}, \\
p_2 &= P(I = 2) = \sum_{k=0}^{\infty} \pi_{02} = \frac{\lambda q_0(\theta + 2 \theta q_2 - r_{33} \mu_{p_3})}{\theta(1 - r_{22})(\theta + r_{23} \mu_{p_2})} \left( \frac{1}{1 - r_{33}} - \frac{\lambda q_0 r_{33}(\theta + r_{23} \mu_{p_2})}{(1 - r_{22})(\theta + r_{23} \mu_{p_2})} \right).
\end{align*}
\]

(4.18) (4.19) (4.20)

So we can get the conditional expected queue length in state \(i (i = 0, 1, 2)\), denoted by \(E[L_i]\), as

\[
\begin{align*}
E[L_0] &= \sum_{k=0}^{\infty} k \pi_{00} = \pi_{00} \left( \frac{(\theta - r_{23} \mu_{p_2})(1 - r_{11} r_{22})}{(\theta + r_{23} \mu_{p_2})(1 - r_{11})^2(1 - r_{22})^2} + \frac{r_{31}(1 - r_{11} r_{33})}{(1 - r_{11})^2(1 - r_{33})^2} + \frac{\theta r_{11}}{\lambda q_0(1 - r_{11})^2} \right. \\
&\quad \left. + \frac{r_{31} r_{23}(\theta + \lambda q_2 - r_{33} \mu_{p_3})}{(1 - r_{11})^2(1 - r_{22})^2} + \frac{r_{31} r_{23}(2 - (r_{11} + r_{22}))}{(1 - r_{11})^2(1 - r_{22})^2} + \frac{r_{33}(2 - (r_{11} + r_{22}))}{(1 - r_{11})^2(1 - r_{22})^2} \right) \\
&\quad \times \left( \frac{(\theta + \lambda q_2 - r_{33} \mu_{p_3})}{(1 - r_{23} \mu_{p_2})(1 - r_{11})(1 - r_{22})} + \frac{r_{31}}{(1 - r_{11})(1 - r_{33})} + \frac{\theta}{\lambda q_0(1 - r_{11})} \right) \\
&\quad - \frac{r_{31} r_{23}(\theta + \lambda q_2 - r_{33} \mu_{p_3})}{(1 - r_{11})(1 - r_{22})} \left( \frac{1 + r_{33} - (r_{11} + r_{22})}{(1 - r_{22})(1 - r_{33})} \right)^{-1} \\
E[L_1] &= \sum_{k=0}^{\infty} k \pi_{01} = \pi_{00} \lambda q_0(\theta + \lambda q_2 - r_{33} \mu_{p_3}) \left( \frac{r_{22}}{(\theta + r_{23} \mu_{p_2})(1 - r_{22})^2} \right)^{-1} = \frac{r_{22}}{1 - r_{22}}. \\
E[L_2] &= \sum_{k=0}^{\infty} k \pi_{02} = \frac{\lambda q_0(\theta + \lambda q_2 - r_{33} \mu_{p_3})}{\theta} \left( \frac{r_{33}}{(1 - r_{33})} + \frac{r_{22} r_{23} (\theta + \lambda q_2 - r_{33} \mu_{p_3})}{(\theta + r_{23} \mu_{p_2})(1 - r_{22})} \right)^{-1} \\
&\quad \times \left( \frac{\lambda q_0(\theta + \lambda q_2 - r_{33} \mu_{p_3})}{\theta} \left( \frac{1}{1 - r_{33}} + \frac{r_{22} r_{23} (\theta + \lambda q_2 - r_{33} \mu_{p_3})}{(\theta + r_{23} \mu_{p_2})(1 - r_{22})} \right)^{-1} \right)^{-1} \\
&= \frac{r_{33}}{1 - r_{33}} + \frac{r_{22}(\theta + \lambda q_2 - r_{33} \mu_{p_3})}{(1 - r_{22})(\theta + r_{23} \mu_{p_2})}.
\end{align*}
\]

(4.21) (4.22) (4.23)

Replacing \(n\) with \(E[L_0]\) in (3.1), we can easily get the conditional expected sojourn time of the customers in busy state, denoted by \(E[W_{0i}]\), as
\[
E[W_0] = \frac{E[L_0]}{\mu_b} + \frac{1}{\mu_b} \\
= \left(\frac{\theta + \lambda q_2 - r_{33} \mu_{r_2}}{(\theta + r_{23} \mu_{r_2})(1 - r_{11})(1 - r_{22})}ight) \\
+ \frac{r_{31} \lambda r_{23}(\theta + \lambda q_2 - r_{33} \mu_{r_2})}{(\theta + r_{23} \mu_{r_2})(1 - r_{11})(1 - r_{22})} \\
+ \frac{r_{31} \lambda r_{23}(\theta + \lambda q_2 - r_{33} \mu_{r_2})}{(\theta + r_{23} \mu_{r_2})(1 - r_{11})(1 - r_{22})} \\
- \frac{1 + r_{33} - (r_{11} + r_{22})}{(1 - r_{11})(1 - r_{22})} \\
- \frac{\mu_b \lambda r_{31} r_{23}(\theta + \lambda q_2 - r_{33} \mu_{r_2})}{(\theta + r_{23} \mu_{r_2})(1 - r_{11})(1 - r_{22})} \\
- \frac{\mu_b \lambda r_{31} r_{23}(\theta + \lambda q_2 - r_{33} \mu_{r_2})}{(\theta + r_{23} \mu_{r_2})(1 - r_{11})(1 - r_{22})}^{-1}.
\]

\[W^*_i(s) = \sum_{k=0}^{\infty} \pi_{ki} \left(\frac{\mu_{r_1}}{\mu_{r_1} + \theta + s}\right)^{k+1} \left(\frac{\theta \mu_{r_2}}{(\theta + \mu_{r_2})(1 - r_{22}) + \theta + s}\right)^{k+1} \]

\[= \frac{\lambda q_0(\theta + \lambda q_2 - r_{33} \mu_{r_2})}{\theta (\theta + r_{23} \mu_{r_2})(1 - r_{11})(1 - r_{22})} \]

Therefore, the LST of the conditional sojourn time in the first-stage working vacation state is

\[\tilde{W}_i(s) = \frac{1}{p_i} W_i(s)\]

\[= \frac{1 - r_{22}}{\mu_{r_1}(1 - r_{22}) + \theta + s} \left(\frac{\theta^2 \mu_{r_2}}{(\theta + \mu_{r_2})(1 - r_{22}) + \theta + s} + \frac{\theta \mu_{r_2}}{\mu_{r_2}(1 - r_{22}) + \theta + s} + \mu_{r_1}\right).\]

Then we get \(E[W_1] = -\tilde{W}_i(0)\), that is,

\[E[W_1] = \frac{\lambda q_0}{\theta} \left(\frac{\theta (\mu_{r_2}(1 - r_{22}) + \theta + s)}{\mu_{r_2}(1 - r_{22}) + \theta + s} + \frac{\theta \mu_{r_2}}{\mu_{r_2}(1 - r_{22}) + \theta + s} + \mu_{r_1}\right)\]

On the other hand, for the conditional expected sojourn time in the second-stage working vacation state, denoted by \(E[W_2]\), we can also discuss the same two cases at state \((k, 2)\) \((k \geq 0)\) with those in the observable queues. So the LS of the customers’ sojourn time, denoted by \(W^*_2(s)\), is

\[W^*_2(s) = \sum_{k=0}^{\infty} \pi_{2k} \left(\frac{\mu_{r_2}}{\mu_{r_2} + \theta + s}\right)^{k+1} \left(\frac{\theta \mu_{r_2}}{(\theta + \mu_{r_2})(1 - r_{22}) + \theta + s}\right)^{k+1} \]

\[= \frac{\mu_b \lambda r_{32} r_{23}(\theta + \lambda q_2 - r_{33} \mu_{r_2})}{(\theta + r_{23} \mu_{r_2})(1 - r_{11})(1 - r_{22})} \]

\[\times \frac{\mu_b \lambda r_{32} r_{23}(\theta + \lambda q_2 - r_{33} \mu_{r_2})}{(\theta + r_{23} \mu_{r_2})(1 - r_{11})(1 - r_{22})}^{-1} \]

So the LST of the conditional sojourn time in the second-stage working vacation state is
\[
\begin{align*}
\bar{W}_2(s) &= \frac{1}{p_2} W_2(s) \left( \frac{\mu_b(\mu_{v_2}(1 - r_{33}) + \theta) + \mu_{v_2}s}{(\mu_{v_2}(1 - r_{33}) + \theta)(\mu_b(1 - r_{33}) + s)} - \frac{\mu_{v_2}^2(\theta + \lambda q_2 - 3r_{33}\mu_{v_2})r_{23}}{(\theta + r_{23}\mu_{v_2})^2(\mu_{v_2}s - \mu_b(\theta + s))} \right) \\
&\times \left(1 - \frac{(\mu_b - \mu_{v_2})s}{(1 - r_{22})\mu_{v_2} + \theta + s}(1 - \frac{(1 - r_{33})\mu_{v_2} + \theta + s}{\mu_{v_2}}) \right) - \frac{\mu_b^2(\theta + \lambda q_2 - r_{33}\mu_{v_2})r_{23}}{(\theta + r_{23}\mu_{v_2})(\mu_{v_2}s - \mu_b(\theta + s))} \\
&\times \left(1 - \frac{(\mu_b - \mu_{v_2})s}{(1 - r_{22})\mu_{v_2} + \theta + s}(1 - \frac{(1 - r_{33})\mu_{v_2} + \theta + s}{\mu_{v_2}}) \right)
\end{align*}
\]

(4.29)

Then we get \( E[W_2] = -\bar{W}_2(0) \), that is,

\[
E[W_2] = \left( \frac{\mu_{v_2}(\theta + r_{23}\mu_{v_2})(1 - r_{22})(1 - r_{33})}{(\mu_{v_2}(1 - r_{33}) + \theta)^2(\mu_b(1 - r_{33}))} + \frac{\theta(\theta + r_{23}\mu_{v_2})(1 - r_{22})(1 - r_{33})(\mu_b + \mu_{v_2})}{(\mu_{v_2}(1 - r_{33}) + \theta)(\mu_b(1 - r_{33}))} \right) \\
- \frac{\mu_{v_2}(\theta + \lambda q_2 - r_{33}\mu_{v_2})r_{23}(\mu_b - \mu_{v_2})(1 - r_{22})(1 - r_{33})}{(\mu_{v_2}(1 - r_{33}) + \theta)(\mu_b(1 - r_{33}))} + \frac{(\theta + \lambda q_2 - r_{33}\mu_{v_2})r_{23}(\mu_b - \mu_{v_2})}{(\mu_b(1 - r_{22})(1 - r_{33}))}
\]

(4.30)

Hence, from (4.27), the expected net benefit of a customer who joins in the first-stage working vacation state, denoted by \( U(1;q_0), U(1;q_1) = \) \( R - CE[W_i] \), which does not depend on \( q_0 \) and \( q_2 \). Then from (4.30), he expected net benefit in the second-stage working vacation state, denoted by \( U(2;q_1, q_2) = R - CE[W_2] \), which only depends on \( q_1 \) and \( q_2 \) but does not depend on \( q_0 \). On the other hand, from (4.24), his expected net benefit in busy state, denoted by \( U(0; q_0, q_1, q_2) = R - CE[W_0] \), which depends on all of \( q_0, q_1 \) and \( q_2 \). Solving the equations \( U(1;q_1) = 0, U(2;q_1, q_2) = 0 \) and \( U(0; q_0, q_1, q_2) = 0 \), we can get the root \( (q^*(0), q^*(1), q^*(2)) \). As for the issue of uniqueness of \( (q^*(0), q^*(1), q^*(2)) \), it is nearly impossible to take theoretical analysis because of complexity of the expressions. Therefore, we make some numerical experiments and observe that there indeed exists a unique feasible and positive root for a fixed \( \lambda \). This indicates that large customer arrivals just mainly exists negative externality. So the customers’ equilibrium strategy is \( (q_0(0), q_1(1), q_2(2)) = (\min(q^*(0), 1), \min(q^*(1), 1), \min(q^*(2), 1)) \) in partially observable case.

Although the server’s service rates have the relations of \( \mu_{v_1} < \mu_b < \mu_{v_2} \). \( q_0(2) \) is the lowest joining probability for a fixed \( \lambda \) and \( q_0(0) \) is not necessarily higher than \( q_0(1) \). The reason is that \( \mu_{v_1} \) is not so slow compared with \( \mu_b \) so that it can satisfy customers’ service requirement. Moreover, shorter queue length also attracts customers to join although the system is in vacation state. However, during the second-stage vacation, \( \mu_{v_2} \) is too slow and the queue length maybe contains some residual customers who do not complete service during the first-stage vacation. Of course, the cases of \( q_2(2) > q_0(2) \) or \( q_1(1) < q_2(2) \) also possibly occur, which depends on the values \( \mu_b, \mu_{v_1} \) and \( \mu_{v_2} \).

Besides \( q_0(2) \), \( q_1(1) \) and \( q_2(2) \) show the sensitivity of equilibrium mixed strategies with respect to \( \theta, \mu_{v_1} \) and \( \mu_{v_2} \) respectively. \( q_0(2) \) shows that both \( q_0(1) \) and \( q_0(2) \) increase with \( \theta \) but the increasing rate of \( q_0(1) \) is much faster than \( q_0(2) \) (Fig. 9). which is resulted by \( \mu_{v_1} = 2\mu_{v_2} \) and the cumulated customers in the first-stage vacation. \( q_0(2) \) shows that \( q_0(1) \) increases with \( \mu_{v_1} \), whereas \( q_0(2) \) decreases with \( \mu_{v_1} \). Moreover, \( q_1(1) \) may be greater than \( q_0(2) \) although \( \mu_{v_1} < \mu_{v_2} \) because of the shorter expected queue length in the first-stage vacation. Similar to \( q_0(2) \), \( q_1(1) \) also shows that both \( q_1(1) \) and \( q_1(2) \) increase with \( \mu_{v_1} \). The reason is that the higher value of \( \mu_{v_1} \) benefits customers arriving in the two-stage vacations. As for \( q_0(0) \), Figs. 9–11 all show the sensitivity of equilibrium mixed strategies with respect to \( \theta, \mu_{v_1} \) and \( \mu_{v_2} \), respectively.
show that $q_{e}(0)$ nearly has little change compared with $q_{e}(1)$ and $q_{e}(2)$, and $q_{e}(0)$ is not necessarily higher than $q_{e}(1)$ or $q_{e}(2)$ in case the vacation environment is better enough.

Then, we consider the equilibrium social welfare. Based on (4.17), the unconditional expected queue length, denoted by $E[L_{po}]$ is

$$E[L_{po}] = \sum_{k=0}^{\infty} k(\pi_{k0} + \pi_{k1} + \pi_{k2})$$

$$= \frac{\lambda q_{0}\pi_{00}}{\theta} \left( \frac{(\theta + \lambda q_{2} - r_{33}\mu_{p})}{(\theta + r_{23}\mu_{p})^2(1-r_{11})^2(1-r_{22})^2} r_{23}r_{31}r_{22}(2 - r_{11} - r_{22})(2 - r_{11} - r_{22} - r_{33}) + r_{21}(1 - r_{11}r_{22}) \right)$$

$$+ \frac{\theta r_{11}}{\lambda q_{0}(1 - r_{11})^2} \left( \frac{r_{31}(1 - r_{11}r_{33})}{(1-r_{11})^2(1-r_{33})} + \frac{r_{22}(\theta + \lambda q_{2} - r_{33}\mu_{p})}{(\theta + r_{23}\mu_{p})^2(1-r_{22})^2} + \frac{r_{33}}{(1-r_{33})} + \frac{r_{23}(\theta + \lambda q_{2} - r_{33}\mu_{p}) (1 - r_{22}r_{33})}{(\theta + r_{23}\mu_{p})(1-r_{22})^2(1-r_{33})^2} \right).$$

Hence, the social welfare per time unit for $(q_{0}, q_{1}, q_{2})$, denoted by $U_{i}(q_{0}, q_{1}, q_{2})$, is $U_{i}(q_{0}, q_{1}, q_{2}) = \lambda (p_{0}q_{0} + p_{1}q_{1} + p_{2}q_{2}) R - CE[L_{po}]$, where $p_{0}, p_{1}, p_{2}$ are given by (4.18), (4.19), (4.20). When all the customers follow the equilibrium mixed strategy $(q_{e}(0), q_{e}(1), q_{e}(2))$, the equilibrium social welfare per time unit can be expressed as $U_{i}(q_{e}(0), q_{e}(1), q_{e}(2))$. Fig. 12 shows the tendency of $U_{i}(q_{e}(0), q_{e}(1), q_{e}(2))$ with $\lambda$ for the partially observable case. Similar to $U_{i}(n_{e}(0), n_{e}(1), n_{e}(2))$, $U_{i}(q_{e}(0), q_{e}(1), q_{e}(2))$ also first increases then decreases.

\[4\] The fitting curve in Fig. 9 is not so smooth because of the calculating error.
Fig. 11. Sensitivity of joining probabilities for the partially observable case with respect to $\mu_{v_2}$ when $R = 10, C = 1, \mu_b = 3, \lambda = 3, \mu_{v_1} = 1, \theta = 0.1$.

Fig. 12. Equilibrium social welfare for the partially observable case when $R = 10, C = 1, \mu_b = 3, \mu_{v_1} = 2, \mu_{v_2} = 1, \theta = 0.1$.

Fig. 13. Transition rate diagram for unobservable queues.
Fig. 14. Equilibrium mixed strategy for the unobservable case when $R = 10, C = 1, \mu_b = 3, \mu_{v_1} = 2, \mu_{v_2} = 1, \theta = 0.1$.

Fig. 15. Sensitivity of joining probability for the unobservable case with respect to $\theta$ when $R = 5, C = 1, \mu_b = 3, \lambda = 3, \mu_{v_1} = 1, \mu_{v_2} = 0.5$.

Fig. 16. Sensitivity of joining probability for the unobservable case with respect to $\mu_{v_1}$ when $R = 10, C = 1, \mu_b = 3, \lambda = 3, \mu_{v_2} = 1, \theta = 0.1$. 
select a joining probability on the customers' equilibrium mixed strategy is depicted in Fig. 13. Hence, the customer's expected net benefit, denoted by $E[L_{\text{un}}]$, in unobservable queues. Therefore, the expected sojourn time of a joining customer, denoted by $E[W]$, is

$$E[W] = \frac{E[L_{\text{un}}]}{\lambda q} = \frac{r_{11}}{\theta} \left( \frac{(\theta + iq - r_{33}\mu_{v_2})}{(\theta + r_{23}\mu_{v_2})(1-r_{11})^2} \frac{r_{33}}{(1-r_{11})^2} + \frac{r_{31}(1-r_{11}r_{33})}{(1-r_{11})^2(1-r_{33})^2} + \frac{r_{22}(\theta + iq - r_{33}\mu_{v_2})}{(\theta + r_{23}\mu_{v_2})(1-r_{22})^2} + \frac{r_{33}}{(1-r_{33})^2} + \frac{r_{23}(\theta + iq - r_{33}\mu_{v_2})(1-r_{22}r_{33})}{(\theta + r_{23}\mu_{v_2})(1-r_{22})^2(1-r_{33})^2} \right).$$

Hence, the customer's expected net benefit, denoted by $U(q)$, is $U(q) = R - CE[W]$. Solving $U(q) = 0$, we get the root $q^*$, and the customers' equilibrium mixed strategy is $q_e = \min(q^*, 1)$. Similar to the partially observable case, we also numerically observe that $U(q) = 0$ has a unique positive root for a fixed $\lambda$ on condition that $\rho = iq/\mu_b < 1$. Fig. 14 shows the decreasing

5. The unobservable queues

Finally, we discuss the customers' equilibrium behavior in the unobservable case, i.e., an arriving customer can observe neither the server's current state $I(t)$ nor the system occupancy $N(t)$ at time $t$. So the customers' decision is equivalent to select a joining probability $q (0 \leq q \leq 1)$, and their equilibrium mixed strategy is denoted by $q_{e}$. The transition diagram is depicted in Fig. 13.

Replacing all $q_i$ ($i = 0, 1, 2$) with $q$ in (4.31), including those in $r_{10}$ and $r_{ij} \in R[i, j = 1, 2, 3]$, we can directly get the expression of the expected queue length, denoted by $E[L_{\text{un}}]$, in unobservable queues. Therefore, the expected sojourn time of a joining customer, denoted by $E[W]$, is

$$E[W] = \frac{E[L_{\text{un}}]}{\lambda q} = \frac{r_{11}}{\theta} \left( \frac{(\theta + iq - r_{33}\mu_{v_2})}{(\theta + r_{23}\mu_{v_2})(1-r_{11})^2} \frac{r_{33}}{(1-r_{11})^2} + \frac{r_{31}(1-r_{11}r_{33})}{(1-r_{11})^2(1-r_{33})^2} + \frac{r_{22}(\theta + iq - r_{33}\mu_{v_2})}{(\theta + r_{23}\mu_{v_2})(1-r_{22})^2} + \frac{r_{33}}{(1-r_{33})^2} + \frac{r_{23}(\theta + iq - r_{33}\mu_{v_2})(1-r_{22}r_{33})}{(\theta + r_{23}\mu_{v_2})(1-r_{22})^2(1-r_{33})^2} \right).$$

Fig. 17. Sensitivity of joining probability for the unobservable case with respect to $\mu_{v_2}$ when $R = 10, C = 1, \mu_b = 3, \lambda = 3, \mu_{v_1} = 1, \theta = 0.1$.

Fig. 18. Equilibrium social welfare for the unobservable case when $R = 10, C = 1, \mu_b = 3, \mu_{v_1} = 2, \mu_{v_2} = 1, \theta = 0.1$. 
trend of $q_e$ with $\lambda$ for the unobservable case. Then Figs. 15–17 show that $q_e$ increases with respect to $\theta$, $\mu_{r_1}$ and $\mu_{r_2}$, respectively, which are consistent with intuition.

Then according to (5.1), we can get the social welfare per time unit for $q_e$ denoted by $U_s(q_e)$, as $U_s(q_e) = 2q(R - CE[W])$. When all customers follow the above equilibrium mixed strategy $q_e$, the equilibrium social welfare per time unit can be denoted by $U_s(q_e)$. Similar to $U_s(n_e(0), n_e(1), n_e(2))$ and $U_s(q_e(0), q_e(1), q_e(2))$, Fig. 18 also shows the first-increase-then-decrease trend of $U_s(q_e)$ with $\lambda$ for the unobservable case.

6. Conclusions

Weighing the cost of state conversion and the actual requirement of assistant work, usually in practice, the server has to take a fix number of vacations, and the lower service rates in successive vacation times can be adjusted but not always variable. Therefore, we presented a new two-stage working vacation policy in this paper, which is an original case. Based on different precision levels of system information, we studied observable, partially observable and unobservable Markovian queues with two-stage working vacations, respectively, and derived customers’ equilibrium balking strategies and equilibrium social welfare. No matter which type of system information, we observed that customers’ positive equilibrium strategy is unique. By comparisons, customers’ equilibrium joining probabilities in vacation states are not necessarily smaller than that in busy state for the partially observable queues. Moreover, in observable/partially observable queues, the order of balking thresholds/joining probabilities in the two-stage working vacations depends on the relation of service rates.

Acknowledgements

The authors would like to thank the anonymous reviewers for the useful comments on this work, and the support from the National Natural Science Foundation of China (Nos. 71101124, 71301139, 71271187), the China Postdoctoral Science Foundation (No. 2014T70231), the Research Fund for the Doctoral Program of Higher Education (No. 20131333120001), the Humanity and Social Science Foundation of Ministry of Education of China (No. 12YJC790101), and the Natural Science Foundation of Hebei Province (Nos. G2012203068, F2013203136).

References