Applied Mathematical Modelling 37 (2013) 3861–3878

Contents lists available at SciVerse ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Equilibrium balking behavior in the $Geo/Geo/1$ queueing system with multiple vacations

Yan Ma [b,a], Wei-qi Liu [b,c], Ji-hong Li [c,*]

[a] School of Mathematics and Statistics, Central South University, Changsha 410075, China
[b] College of Mathematical Sciences, Shanxi University, Taiyuan 030006, China
[c] College of Management, Shanxi University, Taiyuan 030006, China

ARTICLE INFO

Article history:
Received 12 November 2011
Received in revised form 7 July 2012
Accepted 16 August 2012
Available online 4 September 2012

Keywords:
Economics of queues
Multiple vacations
Equilibrium balking strategies
Stationary distribution
Social benefit
Matrix-geometric solution method

ABSTRACT

This paper studies the equilibrium behavior of customers in the $Geo/Geo/1$ queueing system under multiple vacation policy. The server leaves for repeated vacations as soon as the system becomes empty. Customers decide for themselves whether to join or to balk, which is more sensible than the classical viewpoint in queueing theory. Equilibrium customer behavior is considered under four cases: fully observable, almost observable, almost unobservable and fully unobservable, which cover all the levels of information. Based on the reward-cost structure, we obtain the equilibrium balking strategies in all cases. Furthermore, the stationary system behavior is analyzed and a variety of performance measures are developed under the corresponding strategies. Finally, we present several numerical experiments that demonstrate the effect of the information level as well as several parameters on the equilibrium behavior and social benefit. The research results not only offer the customers optimal strategies but also provide the managers with a good reference to discuss the pricing problem in the queueing system.

1. Introduction

Due to the widely applications for management in service system and electronic commerce, there exists an emerging tendency to study customers' behavior in queueing models. In these models, customers are allowed to make decisions as to whether to join or to balk, buy priority or not etc., which is more sensible to describe queueing models. Traditionally, queueing systems are divided into the observable model and the unobservable model regarding whether the information of the queue length is available to customers or not prior to their actions. The observable queueing system was first analyzed by Naor [1], who studied equilibrium and social optimal strategies in an $M/M/1$ queue with a simple linear reward-cost structure. Afterward, Naor's model and results had been extended in several literatures, see e.g. [2–4]. Chen and Frank [5] generalized Naor's model assuming that both the customers and the server maximize their expected discounted utility using a common discount rate. Erlichman and Hassin [6] discussed a priority queue in which customers have the option of overtaking some or all of the customers. On the other hand, Edelson and Hildebrand [7] presented the pioneering literature on the unobservable queue in which the properties of the basic unobservable $M/M/1$ queue were discovered. Littlechild [8] extended the model of Edelson and Hildebrand assuming that customers have different service values. Chen and Frank [9] discussed the robustness of the main result of Edelson and Hildebrand, that a profit maximizer chooses a socially optimal admission fee, when the assumption of a linear utility function is removed. Balachandran [10] considered an unobservable

* Corresponding author. Tel.: +86 13485330831.
E-mail addresses: myhicharge@126.com (Y. Ma), liuwq@sxu.edu.cn (W.-q. Liu), lijh1982@sxu.edu.cn (J.-h. Li).

0307-904X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.apm.2012.08.017

$M/G/1$ model with a fixed cost of running the service facility. Subsequently, several authors had investigated the equilibrium strategies in various unobservable models incorporating many diverse characteristics. The fundamental results on this subject in both the observable and unobservable queueing systems can be found in the comprehensive monograph of Hassin and Haviv [11].

Discrete-time queueing systems with vacations have been widely studied in the past because of their extensively use in digital communication and telecommunication networks. An excellent and complete study on discrete-time vacation models had been presented by Takagi [12]. Zhang and Tian [13] presented the detailed analysis on the $Geo/G/1$ queue with multiple adaptive vacations and further in [14] Tian and Zhang dealt with a $GI/Geo/1$ queue with multiple vacations and exhaustive service. Recently, Samanta et al. [15] and Tang et al. [16] did research on the discrete-time $Geo^X/G/1$ vacation queue with different characteristics.

As for the research on the equilibrium customer behavior in vacation queue models, the first was presented by Burnetas and Economou [17], who explored both the observable and unobservable cases in a single server Markovian queue with set up times. Then, Economou and Kanta [18] analyzed the equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. Recently, Sun et al. [19] considered the equilibrium behavior of customers in an observable $M/M/1$ queue under interruptible and insusceptible setup/closedown policies. Economou et al. [20] extended the analysis done for the almost and fully unobservable queues in [17] to the non-Markovian case. Liu et al. [21] studied the equilibrium threshold strategies in observable queues under single vacation policy. However, there was no work concerning the equilibrium balking behavior in the discrete-time queues with multiple vacations.

In the present paper we analyze the equilibrium balking strategies in the discrete-time $Geo/Geo/1$ queue with multiple vacations. To the authors' knowledge, this is the first time that the multiple vacation policy is introduced into the economics of queues. The customers' dilemma is whether to join the system or balk. They make decisions based on a nature reward-cost structure, which incorporates their desire for service as well as their unwillingness to wait. We explore various cases with regard to the level of information available to customers upon arrival. More specifically, at his arrival epoch a customer may or may not know the number of customers present and/or the state of the server. Therefore, four combinations emerge, ranging from full to no information. In each of the four cases we discuss customer equilibrium strategies, analyze the stationary behavior of the corresponding system and derive the equilibrium social benefit for all customers. Furthermore, we present several numerical experiments to explore the effect of the information level as well as several parameters on the equilibrium behavior and the social benefit.

This paper is organized as follows. In Section 2, we give the model description and the reward-cost structure. Section 3 discusses the observable queue in which customers observe the length of the queue. We distinguish two subcases depending on the additional information, or lack thereof, of the server state. We determine equilibrium threshold strategies and analyze the resulting stationary system behavior. Then Section 4 studies the unobservable queue where the queue length is not available to customers. We derive the corresponding mixed equilibrium strategies and investigate the stationary behavior for the almost and fully unobservable models. In Section 5, we illustrate the effect of the information level on the equilibrium behavior and the social benefit via analytical and numerical comparisons. Finally, in Section 6, we give a necessary conclusion.

2. Model description

We consider a single server queueing system with multiple vacations. The server takes a vacation immediately at the end of each busy period. If he finds the system still empty upon returning from the vacation, he will take another vacation and so on. In this paper, for any real number $x \in [0,1]$, we denote $\bar{x} = 1 - x$.

Assume that customer arrivals occur at the end of slot $t = n^-, n = 0, 1, \ldots$. The inter-arrival times are independent and identically distributed sequences following a geometric distribution with rate p.

$$P(T = k) = p\bar{p}^{k-1}, \quad k \geqslant 1, \quad 0 < p < 1, \quad \bar{p} = 1 - p.$$

The service starting and service ending occur at slot division point $t = n, \ n = 0, 1, \ldots$. The service times are independent each other and geometrically distributed with rate μ.

$$P(S = k) = \mu\bar{\mu}^{k-1}, \quad k \geqslant 1, \quad 0 < \mu < 1, \quad \bar{\mu} = 1 - \mu.$$

The beginning and ending of vacation occur at the epoch which is similar to $t = n^-$ in shape. The vacation time is an independent and identically distributed random variable following a geometric distribution with rate θ.

$$P(V = k) = \theta\bar{\theta}^{k-1}, \quad k \geqslant 1, \quad 0 < \theta < 1, \quad \bar{\theta} = 1 - \theta.$$

We assume that inter-arrival times, service times, and vacation times are mutually independent. The queueing system follows the First-Come-First-Served (FCFS) service discipline. Moreover, suppose the systems considered are stable. The service rate exceeds the arrival rate so that the server can accommodate all arrivals.

Let L_n be the number of customers in the system at time n^+. According to the assumptions above, a customer who finishes service and leaves at $t = n^+$ does not reckon on L_n while arrives at $t = n^-$ should reckon on L_n. We assume

$$J_n = \begin{cases} 0, & \text{the system is in a vacation period at time } n^+, \\ 1, & \text{the system is in a service period at time } n^+. \end{cases}$$

It is clear that $\{L_n, J_n\}$ is a Markov chain with state space

$$\Omega = \{(k,j) | k \geqslant j, j = 0, 1\},$$

where state $(k, 0), k \geqslant 0$, indicates that the system is in the vacation period and there are k customers; state $(k, 1), k \geqslant 1$, indicates that the system is in the busy period and there are k customers. Furthermore, we distinguish four cases depending on the information provided to customers before making decisions, which cover all the levels of information.

- Fully observable case: Customers observe (L_n, J_n);
- Almost observable case: Customers observe only L_n;
- Almost unobservable case: Customers observe only J_n;
- Fully unobservable case: Customers do not observe the system state.

Our interest is in the behavior of customers when they decide whether to join or to balk at their arrival instant. Suppose Se be the mean sojourn time of a customer in equilibrium and Be be the expected net benefit. To model the decision process, we assume that every customer receives a reward of R units for completing service. This may reflect his satisfaction and the added value of being served. On the other hand, there exists a waiting cost of C units per time unit that the customer remains in the system (in queue or in service). Customers are risk neutral and maximize their expected net benefit. From now on, we assume the condition

$$R > \frac{C}{\mu} + \frac{C}{\theta}. \tag{1}$$

This condition ensures that the reward for service exceeds the expected cost for a customer who finds the system empty. Otherwise, after the system becomes empty for the first time no customers will ever enter. Finally, we stress that the decisions are irrevocable: retrials of balking customers and reneging of entering customers are not allowed.

3. Analysis of the observable queues

We first consider the observable queues in which customers are informed of the queue length upon arrival. We show that there exist equilibrium balking strategies of thresholds type. In the fully observable queue, a pure threshold strategy is specified by a pair $(L_e(0), L_e(1))$ and has the form 'observe (L_n, J_n) at arrival instant; enter if $L_n \leqslant L_e(J_n)$ and balk otherwise'. In the almost observable queue, a pure threshold strategy is specified by a single number L_e and has the form 'observe L_n; enter if $L_n \leqslant L_e$ and balk otherwise'.

3.1. Fully observable queue

We begin with the fully observable case in which arriving customers know both the number of present customers L_n and the state of the server J_n. In equilibrium, a customer who joins the system when he observes state (k,j) has mean sojourn time

$$Se = \frac{k+1}{\mu} + \frac{1-j}{\theta}.$$

Thus his expected net benefit is

$$Be = R - \frac{C(k+1)}{\mu} - \frac{C(1-j)}{\theta}.$$

The customer strictly prefers to enter if Be is positive and is indifferent between entering and balking if it equals zero. We thus conclude the following theorem.

Theorem 3.1. *In the fully observable Geo/Geo/1 queue with multiple vacations, there exist thresholds*

$$(L_e(0), L_e(1)) = \left(\left\lfloor \frac{R\mu}{C} - \frac{\mu}{\theta} \right\rfloor - 1, \left\lfloor \frac{R\mu}{C} \right\rfloor - 1 \right) \tag{2}$$

such that the strategy 'observe (L_n, J_n), enter if $L_n \leqslant L_e(J_n)$ and balk otherwise' is a unique equilibrium in the class of the threshold strategies.

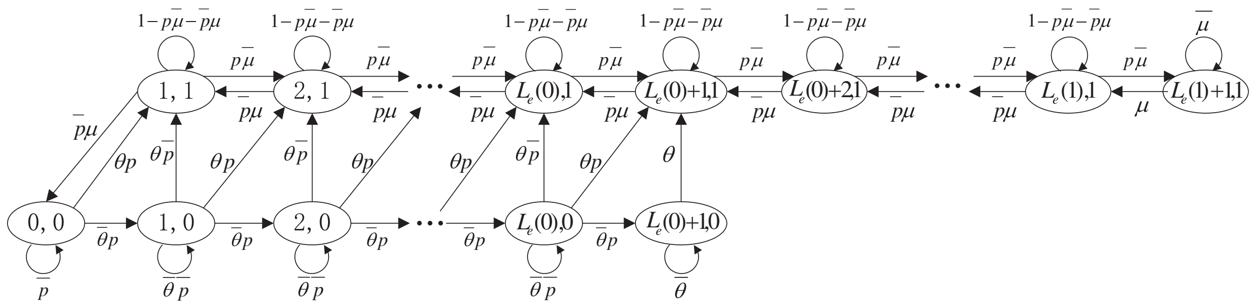

Fig. 1. Transition rate diagram for the $(L_e(0), L_e(1))$ threshold strategy in the fully observable queue.

Remark 1. $L_e(0)$ is the threshold when an arriving customer finds the system is in a vacation period and $L_e(1)$ is the threshold when it is in a regular busy period. We get $L_e(0)$ and $L_e(1)$ from the condition $Be > 0$ when $i = 0$ and 1 respectively. The symbol $\lfloor\rfloor$ indicates rounding down.

For the stationary analysis of the system, note that if all customers follow the threshold strategy in (2), the system follows a Markov chain with state space restricted to $\Omega_{fo} = \{(k,0)|0 \leqslant k \leqslant L_e(0)+1\} \cup \{(k,1)|1 \leqslant k \leqslant L_e(1)+1\}$. The transition rate diagram is depicted in Fig. 1. The one-step transition probabilities of (L_n, J_n) are as follows:

Case 1: if $X_n = (0,0)$,

$$X_{n+1} = \begin{cases} (0,0), & \text{with probability } \bar{p}, \\ (1,0), & \text{with probability } \bar{\theta}p, \\ (1,1), & \text{with probability } \theta p. \end{cases}$$

Case 2: if $X_n = (k,0), 1 \leqslant k \leqslant L_e(0)$,

$$X_{n+1} = \begin{cases} (k,0), & \text{with probability } \bar{\theta}\bar{p}, \\ (k,1), & \text{with probability } \theta\bar{p}, \\ (k+1,0), & \text{with probability } \bar{\theta}p, \\ (k+1,1), & \text{with probability } \theta p. \end{cases}$$

Case 3: if $X_n = (L_e(0)+1,0)$,

$$X_{n+1} = \begin{cases} (L_e(0)+1,0), & \text{with probability } \bar{\theta}, \\ (L_e(0)+1,1), & \text{with probability } \theta. \end{cases}$$

Case 4: if $X_n = (1,1)$,

$$X_{n+1} = \begin{cases} (0,0), & \text{with probability } \bar{p}\mu, \\ (1,1), & \text{with probability } 1 - p\bar{\mu} - \bar{p}\mu, \\ (2,1), & \text{with probability } p\bar{\mu}. \end{cases}$$

Case 5: if $X_n = (k,1), 2 \leqslant k \leqslant L_e(1)$

$$X_{n+1} = \begin{cases} (k-1,1), & \text{with probability } \bar{p}\mu, \\ (k,1), & \text{with probability } 1 - p\bar{\mu} - \bar{p}\mu, \\ (k+1,1), & \text{with probability } p\bar{\mu}. \end{cases}$$

Case 6: if $X_n = (L_e(1)+1,1)$,

$$X_{n+1} = \begin{cases} (L_e(1),1), & \text{with probability } \mu, \\ (L_e(1)+1,1), & \text{with probability } \bar{\mu}. \end{cases}$$

Based on the one-step transition situation analysis, using the lexicographical sequence for the states, the transition probability matrix can be written as

$$\widetilde{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{B}_0 & \boldsymbol{A}_0 & & & & & & & & & & \\ \boldsymbol{B}_1 & \boldsymbol{A}_1 & \boldsymbol{C}_1 & & & & & & & & & \\ & \boldsymbol{B}_2 & \boldsymbol{A}_1 & \boldsymbol{C}_1 & & & & & & & & \\ & & \ddots & \ddots & \ddots & & & & & & & \\ & & & \boldsymbol{B}_2 & \boldsymbol{A}_1 & \boldsymbol{C}_1 & & & & & & \\ & & & & \boldsymbol{B}_2 & \boldsymbol{A}_2 & \boldsymbol{C}_2 & & & & & \\ & & & & & \boldsymbol{B}_3 & \boldsymbol{A}_3 & \boldsymbol{C}_3 & & & & \\ & & & & & & \boldsymbol{B}_4 & \boldsymbol{A}_3 & \boldsymbol{C}_3 & & & \\ & & & & & & & \ddots & \ddots & \ddots & & \\ & & & & & & & & \boldsymbol{B}_4 & \boldsymbol{A}_3 & \boldsymbol{C}_3 & \\ & & & & & & & & & \boldsymbol{B}_5 & \boldsymbol{A}_4 & \end{bmatrix}, \tag{3}$$

where

$$\boldsymbol{B}_0 = \bar{p}, \quad \boldsymbol{A}_0 = (\bar{\theta}p, \theta p),$$
$$\boldsymbol{B}_1 = \begin{bmatrix} 0 \\ \bar{p}\mu \end{bmatrix}, \quad \boldsymbol{A}_1 = \begin{bmatrix} \bar{\theta}\bar{p} & \theta\bar{p} \\ 0 & 1 - p\bar{\mu} - \bar{p}\mu \end{bmatrix}, \quad \boldsymbol{C}_1 = \begin{bmatrix} \bar{\theta}p & \theta p \\ 0 & p\bar{\mu} \end{bmatrix},$$
$$\boldsymbol{B}_2 = \begin{bmatrix} 0 & 0 \\ 0 & \bar{p}\mu \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} \bar{\theta} & \theta \\ 0 & 1 - p\bar{\mu} - \bar{p}\mu \end{bmatrix}, \quad \boldsymbol{C}_2 = \begin{bmatrix} 0 \\ p\bar{\mu} \end{bmatrix}$$

and

$$\boldsymbol{B}_3 = (0, \bar{p}\mu), \quad \boldsymbol{A}_3 = 1 - p\bar{\mu} - \bar{p}\mu, \quad \boldsymbol{C}_3 = p\bar{\mu}, \quad \boldsymbol{B}_4 = \bar{p}\mu, \quad \boldsymbol{A}_4 = \bar{\mu}, \quad \boldsymbol{B}_5 = \mu.$$

Let (L, J) be the stationary limit of (L_n, J_n) and its distribution is denoted as

$$\pi_{kj} = P\{L = k, J = j\}, \quad (k, j) \in \Omega_{fo}$$
$$\pi_k = \begin{cases} \pi_{00}, & k = 0, \\ (\pi_{k0}, \pi_{k1}), & 1 \leqslant k \leqslant L_e(0) + 1, \\ \pi_{k1}, & L_e(0) + 2 \leqslant k \leqslant L_e(1) + 1. \end{cases}$$
$$\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_{L_e(1)+1}).$$

We solve for the stationary distribution π_{kj} by noting that the vector $\boldsymbol{\pi}$ satisfies the equation $\boldsymbol{\pi}\widetilde{\boldsymbol{P}} = \boldsymbol{\pi}$ and have the following system of steady-state equations:

$$\pi_{00} = \bar{p}\pi_{00} + \bar{p}\mu\pi_{11}, \tag{4}$$
$$\pi_{k0} = \bar{\theta}p\pi_{k-1,0} + \bar{\theta}\bar{p}\pi_{k0}, \quad k = 1, 2, 3, \ldots, L_e(0), \tag{5}$$
$$\pi_{L_e(0)+1,0} = \bar{\theta}p\pi_{L_e(0),0} + \bar{\theta}\pi_{L_e(0)+1,0}, \tag{6}$$
$$\pi_{11} = \theta p\pi_{00} + \theta\bar{p}\pi_{10} + (1 - p\bar{\mu} - \bar{p}\mu)\pi_{11} + \bar{p}\mu\pi_{21}, \tag{7}$$
$$\pi_{k1} = \theta p\pi_{k-1,0} + p\bar{\mu}\pi_{k-1,1} + \theta\bar{p}\pi_{k0} + (1 - p\bar{\mu} - \bar{p}\mu)\pi_{k1} + \bar{p}\mu\pi_{k+1,1}, \quad k = 2, 3, \ldots, L_e(0), \tag{8}$$
$$\pi_{L_e(0)+1,1} = \theta p\pi_{L_e(0),0} + p\bar{\mu}\pi_{L_e(0),1} + \theta\pi_{L_e(0)+1,0} + (1 - p\bar{\mu} - \bar{p}\mu)\pi_{L_e(0)+1,1} + \bar{p}\mu\pi_{L_e(0)+2,1}, \tag{9}$$
$$\pi_{k1} = p\bar{\mu}\pi_{k-1,1} + (1 - p\bar{\mu} - \bar{p}\mu)\pi_{k1} + \bar{p}\mu\pi_{k+1,1}, \quad k = L_e(0) + 2, \ldots, L_e(1) - 1, \tag{10}$$
$$\pi_{L_e(1),1} = p\bar{\mu}\pi_{L_e(1)-1,1} + (1 - p\bar{\mu} - \bar{p}\mu)\pi_{L_e(1),1} + \mu\pi_{L_e(1)+1,1}, \tag{11}$$
$$\pi_{L_e(1)+1,1} = p\bar{\mu}\pi_{L_e(1),1} + \bar{\mu}\pi_{L_e(1)+1,1}. \tag{12}$$

Define $\alpha = \frac{p\bar{\mu}}{\bar{p}\mu}, \beta = \frac{\bar{\theta}p}{1-\bar{\theta}\bar{p}}$. Thus $\alpha < 1$ and $\beta < 1$.

By iterating (5) and (10), taking into account (6), (11) and (12), we obtain

$$\pi_{k0} = \beta^k\pi_{00}, \quad k = 1, 2, \ldots, L_e(0), \tag{13}$$
$$\pi_{L_e(0)+1,0} = \frac{1 - \bar{\theta}\bar{p}}{\theta}\beta^{L_e(0)+1}\pi_{00}, \tag{14}$$
$$\pi_{k1} = \alpha^{k-L_e(0)-1}\pi_{L_e(0)+1,1}, \quad k = L_e(0) + 2, \ldots, L_e(1), \tag{15}$$
$$\pi_{L_e(1)+1,1} = \bar{p}\alpha^{L_e(1)-L_e(0)}\pi_{L_e(0)+1,1}. \tag{16}$$

From (8) we observe that $\{\pi_{k1} | k = 1, 2, \ldots, L_e(0) + 1\}$ is a solution of the nonhomogeneous linear difference equation with constant coefficients

$$\bar{p}\mu x_{k+1} - (p\bar{\mu} + \bar{p}\mu)x_k + p\bar{\mu}x_{k-1} = -\theta p\pi_{k-1,0} - \theta\bar{p}\pi_{k0} = -\frac{\theta}{\bar{\theta}}\beta^k\pi_{00}, \quad k = 2, 3, \ldots, L_e(0), \tag{17}$$

where the last equation is due to (13). Using the standard approach for solving such equations, see e.g. [22], we consider the corresponding characteristic equation

$$\bar{p}\mu x^2 - (p\bar{\mu} + \bar{p}\mu)x + p\bar{\mu} = 0,$$

which has two roots at 1 and α. Then the general solution of the homogeneous version of (17) is $x_k^{hom} = A1^k + B\alpha^k$.

The general solution x_k^{gen} of (17) is given as $x_k^{gen} = x_k^{hom} + x_k^{spec}$, where x_k^{spec} is a specific solution of (17). Because the nonhomogeneous part of (17) is geometric with parameter β, we can find a specific solution $D\beta^k$ (we assume $\alpha \neq \beta$). Substituting $x_k^{gen} = D\beta^k$ into (17) we obtain

$$D = \frac{1 - \bar{\theta}\bar{p}}{\bar{\theta}\bar{p} - \bar{\mu}}\pi_{00}. \tag{18}$$

Hence the general solution of (17) is given as

$$x_k^{gen} = A1^k + B\alpha^k + D\beta^k, \quad k = 1, 2, \ldots, L_e(0) + 1, \tag{19}$$

where D is given by (18) and A, B are to be determined.

From (19) for $k = 1$, we obtain

$$A + B\alpha + D\beta = \pi_{11} = \frac{p}{\bar{p}\mu}\pi_{00}. \tag{20}$$

Furthermore, substituting (19) into (7) for $k = 2$, it follows after some rather tedious algebra that

$$A + B\alpha^2 + D\beta^2 = \pi_{21} = \frac{p^2(p\bar{\mu} + \bar{p}\mu\bar{\theta} + \bar{p}\theta\bar{\mu})}{(\bar{p}\mu)^2(1 - \bar{\theta}\bar{p})}\pi_{00}. \tag{21}$$

Solving the system of (20) and (21), we obtain

$$A = 0, \quad B = \frac{-\theta - \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\pi_{00}.$$

Then, from (19),

$$\pi_{k1} = \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}(\beta^k - \alpha^k)\pi_{00}, \quad k = 1, 2, \ldots, L_e(0) + 1. \tag{22}$$

We have thus expressed all stationary probabilities in terms of π_{00} in relations see (13)–(16) and (22). The remaining probability π_{00} can be found from the normalization equation

$$\sum_{k=0}^{L_e(0)+1}\pi_{k0} + \sum_{k=1}^{L_e(1)+1}\pi_{k1} = 1.$$

After some algebraic simplification, we can express all stationary probabilities in the following theorem.

Theorem 3.2. *Consider a fully observable Geo/Geo/1 queue with multiple vacations and $\alpha \neq \beta$, in which customers follow the threshold policy $(L_e(0), L_e(1))$ given in Theorem 3.1. The stationary probabilities $\{\pi_{kj} | (k, j) \in \Omega_{fo}\}$ are as follows:*

$$\pi_{00} = \left\{\frac{\theta + \bar{\theta}p}{\theta} + \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left[\frac{\bar{\theta}p}{\theta} - \frac{p\bar{\mu}}{\mu - p} + \frac{p\bar{p}}{\mu - p}\alpha^{L_e(1)+1} + \left(\frac{p\bar{\mu}}{\mu - p} - \frac{\bar{\theta}p}{\theta} - \frac{p\bar{p}}{\mu - p}\alpha^{L_e(1)-L_e(0)}\right)\beta^{L_e(0)+1}\right]\right\}^{-1}, \tag{23}$$

$$\pi_{k0} = \beta^k\pi_{00}, \quad k = 1, 2, \ldots, L_e(0), \tag{24}$$

$$\pi_{L_e(0)+1,0} = \frac{\beta^{L_e(0)+1}}{1 - \beta}\pi_{00}, \tag{25}$$

$$\pi_{k1} = \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}(\beta^k - \alpha^k)\pi_{00}, \quad k = 1, 2, \ldots, L_e(0) + 1, \tag{26}$$

$$\pi_{k1} = \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left[\left(\frac{\beta}{\alpha}\right)^{L_e(0)+1} - 1\right]\alpha^k\pi_{00}, \quad k = L_e(0) + 2, \ldots, L_e(1), \tag{27}$$

$$\pi_{L_e(1)+1,1} = \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left[\left(\frac{\beta}{\alpha}\right)^{L_e(0)+1} - 1\right]\bar{p}\alpha^{L_e(1)+1}\pi_{00}. \tag{28}$$

Because the probability of balking is equal to $\pi_{L_e(0)+1,0} + \pi_{L_e(1)+1,1}$, the social benefit per time unit when all customers follow the threshold policy $(L_e(0), L_e(1))$ given in Theorem 3.1 equals

$$SB_{fo} = Rp(1 - \pi_{L_e(0)+1,0} - \pi_{L_e(1)+1,1}) - C\left(\sum_{k=0}^{L_e(0)+1} k\pi_{k0} + \sum_{k=1}^{L_e(1)+1} k\pi_{k1}\right).$$

3.2. Almost observable queue

We next consider the almost observable case, where arriving customers only know the queue length L_n before making decisions. Hence the stationary distribution of the corresponding Markov chain is from Theorem 3.2 with $L_e(0) = L_e(1) = L_e$ and state space $\Omega_{ao} = \{k|0 \leqslant k \leqslant L_e + 1\}$. The transition rate diagram is depicted in Fig. 2.

Theorem 3.3. *Consider an almost observable Geo/Geo/1 queue with multiple vacations and $\alpha \neq \beta$, in which customers follow the threshold policy L_e. The stationary probabilities $\{\pi'_k|k \in \Omega_{ao}\}$ are as follows:*

$$\pi'_0 = \left[\frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left(\frac{\mu^2 - \mu p - \mu\theta\bar{p}}{\theta(\mu - p)} + \frac{\bar{p}\mu}{\mu - p}\alpha^{L_e+1} - \frac{\mu}{\theta}\beta^{L_e+1}\right)\right]^{-1},$$

$$\pi'_k = \left(\frac{\mu}{\mu - \theta - \bar{\theta}p}\beta^k - \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\alpha^k\right)\pi'_0, \quad k = 1, 2, \ldots, L_e,$$

$$\pi'_{L_e+1} = \left[\frac{(\mu - \theta)(\theta + \bar{\theta}p)}{\theta(\mu - \theta - \bar{\theta}p)}\beta^{L_e+1} - \frac{\bar{p}(\theta + \bar{\theta}p)}{\mu - \theta - \bar{\theta}p}\alpha^{L_e+1}\right]\pi'_0.$$

Because the expected net benefit of a customer who finds k customers in the system, if he decides to enter, is

$$Be = R - \frac{C(k+1)}{\mu} - \frac{C\pi_{\bar{J}|L}(0|k)}{\theta}, \tag{29}$$

where $\pi_{\bar{J}|L}(0|k)$ is the probability that an arriving customer finds the server in a vacation time, given that there are k customers. Using the various forms of π_{kj} from (23)–(28), we get

$$\begin{cases} \pi_{\bar{J}|L}(0|k) = \left[1 + \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left(1 - \left(\frac{\alpha}{\beta}\right)^k\right)\right]^{-1}, & k = 0, 1, 2, \ldots, L_e, \\ \pi_{\bar{J}|L}(0|L_e + 1) = \left[1 + (1 - \beta)\bar{p}\frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left(1 - \left(\frac{\alpha}{\beta}\right)^{L_e+1}\right)\right]^{-1}. \end{cases} \tag{30}$$

In light of (29) and (30), we introduce the function

$$g(k, y) = R - \frac{C(k+1)}{\mu} - \frac{C}{\theta}\left[1 + y\frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left(1 - \left(\frac{\alpha}{\beta}\right)^k\right)\right]^{-1}, \quad y \in [(1 - \beta)\bar{p}, 1], \quad k = 0, 1, 2, \ldots, \tag{31}$$

which will allow us to prove the existence of equilibrium threshold strategies and derive the corresponding thresholds. Let

$$g_U(k) = g(k, 1) = R - \frac{C(k+1)}{\mu} - \frac{C}{\theta}\left[1 + \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left(1 - \left(\frac{\alpha}{\beta}\right)^k\right)\right]^{-1}, \quad k = 0, 1, 2, \ldots, \tag{32}$$

$$g_L(k) = g(k, (1 - \beta)\bar{p}) = R - \frac{C(k+1)}{\mu} - \frac{C}{\theta}\left[1 + (1 - \beta)\bar{p}\frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left(1 - \left(\frac{\alpha}{\beta}\right)^k\right)\right]^{-1}, \quad k = 0, 1, 2, \ldots, \tag{33}$$

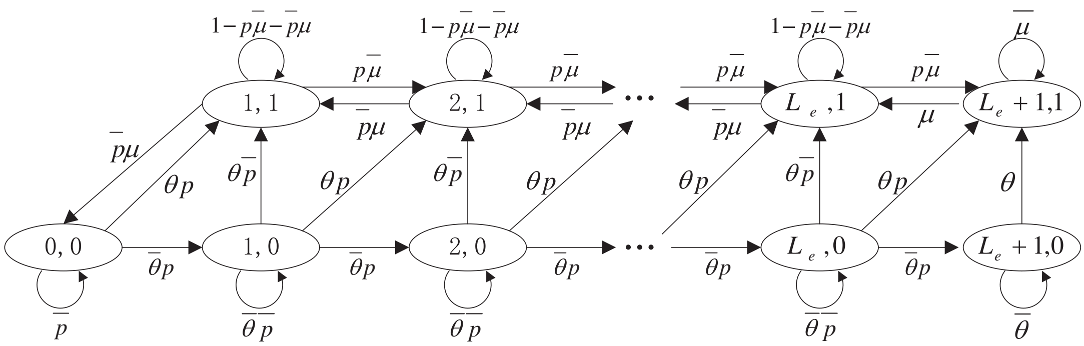

Fig. 2. Transition rate diagram for the L_e threshold strategy in the almost observable queue.

It is easy to see that

$$g_U(0) = g_L(0) = R - \frac{C}{\mu} - \frac{C}{\theta} > 0.$$

In addition,

$$\lim_{k\to\infty} g_U(k) = \lim_{k\to\infty} g_L(k) = -\infty.$$

Hence there exists k_U such that

$$g_U(0), g_U(1), g_U(2), \ldots, g_U(k_U) > 0 \quad \text{and } g_U(k_U+1) \leqslant 0. \tag{34}$$

Because the function $g(k,y)$ is increasing with respect to y for every fixed k, we get the relation $g_L(k) \leqslant g_U(k),\ k = 0,1,2,\ldots$. In particular, $g_L(k_U+1) \leqslant 0$ while $g_L(0) > 0$. Hence, there exists $k_L \leqslant k_U$ such that

$$g_L(k_L) > 0 \quad \text{and } g_L(k_L+1), \ldots, g_L(k_U), g_L(k_U+1) \leqslant 0. \tag{35}$$

We can now establish the existence of the equilibrium threshold policies in the almost observable case and give the following theorem.

Theorem 3.4. *In the almost observable Geo/Geo/1 queue with multiple vacations, all pure threshold strategies 'observe L_n, enter if $L_n \leqslant L_e$ and balk otherwise' for $L_e = k_L, k_L+1, \ldots, k_U$ are equilibrium balking strategies.*

Proof. Consider a tagged customer at his arrival instant and assume all other customers follow the same threshold strategy 'observe L_n, enter if $L_n \leqslant L_e$ and balk otherwise' for some fixed $L_e \in \{k_L, k_L+1, \ldots, k_U\}$. Then $\pi^-_{J|L}(0|k)$ is given by (30).

If the tagged customer finds $k \leqslant L_e$ customers and decides to enter, his expected net benefit is equal to

$$R - \frac{C(k+1)}{\mu} - \frac{C}{\theta}\left[1 + \frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left(1 - \left(\frac{\alpha}{\beta}\right)^k\right)\right]^{-1} = g_U(k) > 0,$$

because of (29)–(32) and (34). So in this case the customer prefers to enter.

If the tagged customer finds $k = L_e + 1$ customers and decides to enter, his expected net benefit is

$$R - \frac{C(L_e+2)}{\mu} - \frac{C}{\theta}\left[1 + (1-\beta)\bar{p}\frac{\theta + \bar{\theta}p}{\mu - \theta - \bar{\theta}p}\left(1 - \left(\frac{\alpha}{\beta}\right)^{L_e+1}\right)\right]^{-1} = g_L(L_e+1) \leqslant 0,$$

because of (29)–(31), (33) and (35). Therefore in this case the customer prefers to balk. □

Remark: There exist equilibrium mixed threshold strategies. In Theorem 3.4, we restrict our attention to the pure threshold strategies because they are evolutionarily stable strategy (ESS) while the mixed are not ESS.

Because the probability of balking is equal to π'_{L_e+1}, the social benefit per time unit when all customers follow the threshold policy L_e given in Theorem 3.4 equals

$$SB_{ao} = Rp(1 - \pi'_{L_e+1}) - C\left(\sum_{k=0}^{L_e+1} k\pi'_k\right).$$

4. Analysis of the unobservable queues

In this section we turn attention to the unobservable queues, where arriving customers do not observe the length of the queue. We prove that there exist equilibrium mixed strategies. In the almost unobservable queue, a mixed strategy is specified by a vector $(q(0), q(1))$, where $q(j)$ is the probability of joining when the server is in state j. In the fully unobservable queue, where customers are provided with no information, a mixed strategy is specified by the probability q of entering.

4.1. Almost unobservable queue

We begin with the almost unobservable case in which arriving customers observe the state j of the server upon arrival. If all customers follow the same mixed strategy $(q(0), q(1))$, then the system follows a Markov chain in which the arrival rate equals $p(j) = pq(j)$ when the server is in state j. The state space is $\Omega_{au} = \{(k,j)|k \geqslant j, j = 0,1\}$ and the transition rate diagram is illustrated in Fig. 3.

Using the lexicographical sequence for the states, the transition probability matrix can be written as

$$\widetilde{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{B}_0 & \boldsymbol{A}_0 & & & & \\ \boldsymbol{B}_1 & \boldsymbol{A}_1 & \boldsymbol{C}_1 & & & \\ & \boldsymbol{B}_2 & \boldsymbol{A}_1 & \boldsymbol{C}_1 & & \\ & & \boldsymbol{B}_2 & \boldsymbol{A}_1 & \boldsymbol{C}_1 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}, \tag{36}$$

where

$$\boldsymbol{B}_0 = \overline{p(0)}, \quad \boldsymbol{A}_0 = (\bar{\theta}p(0), \theta p(0)), \quad \boldsymbol{B}_1 = \begin{bmatrix} 0 \\ \overline{p(1)}\mu \end{bmatrix},$$
$$\boldsymbol{B}_2 = \begin{bmatrix} 0 & 0 \\ 0 & \overline{p(1)}\mu \end{bmatrix}, \quad \boldsymbol{A}_1 = \begin{bmatrix} \bar{\theta}\overline{p(0)} & \theta\overline{p(0)} \\ 0 & 1 - p(1)\bar{\mu} - \overline{p(1)}\mu \end{bmatrix}, \quad \boldsymbol{C}_1 = \begin{bmatrix} \bar{\theta}p(0) & \theta p(0) \\ 0 & p(1)\bar{\mu} \end{bmatrix}.$$

Due to the block tridiagonal structure of transition probability matrix, $\{L_n, J_n\}$ is a quasi birth and death chain. Let (L, J) be the stationary limit of (L_n, J_n) and its distribution is denoted as

$$\pi''_{kj} = P\{L = k, J = j\}, \quad (k, j) \in \Omega_{au},$$
$$\pi''_0 = \pi''_{00}, \quad \pi''_k = (\pi''_{k0}, \pi''_{k1}), \quad k \geqslant 1,$$
$$\boldsymbol{\pi}'' = (\pi''_0, \pi''_1, \pi''_2, \ldots).$$

We solve for (π''_{10}, π''_{11}) by the equation $\boldsymbol{\pi}''\widetilde{\boldsymbol{P}} = \boldsymbol{\pi}''$ and obtain that:

$$(\pi''_{10}, \pi''_{11}) = \left(\frac{\bar{\theta}p(0)}{1 - \bar{\theta}\overline{p(0)}}, \frac{p(0)}{\overline{p(1)}\mu} \right) \pi''_{00}. \tag{37}$$

Using the matrix-geometric solution method, the rate matric $\boldsymbol{R}$ is the solution of the matrix quadratic equation

$$\boldsymbol{R} = \boldsymbol{R}^2 B_2 + \boldsymbol{R} A_1 + C_1.$$

After some rather tedious algebra, we obtain that:

$$\boldsymbol{R} = \begin{bmatrix} \beta(0) & \frac{p(0)}{\overline{p(1)}\mu} \\ 0 & \alpha(1) \end{bmatrix}, \tag{38}$$

where $\beta(0) = \frac{\bar{\theta}p(0)}{1 - \bar{\theta}\overline{p(0)}}, \alpha(1) = \frac{p(1)\bar{\mu}}{\overline{p(1)}\mu}$.
Thus from

$$\pi''_k = (\pi''_{10}, \pi''_{11}) \boldsymbol{R}^{k-1}, \quad k \geqslant 1, \tag{39}$$

we can express all stationary probabilities in terms of π''_{00} in relations see (37)–(39). The remaining probability π''_{00} can be found from the normalization equation

$$\sum_{k=0}^{\infty} \pi''_{k0} + \sum_{k=1}^{\infty} \pi''_{k1} = 1.$$

After some algebraic simplification, we can express all stationary probabilities in the following theorem.

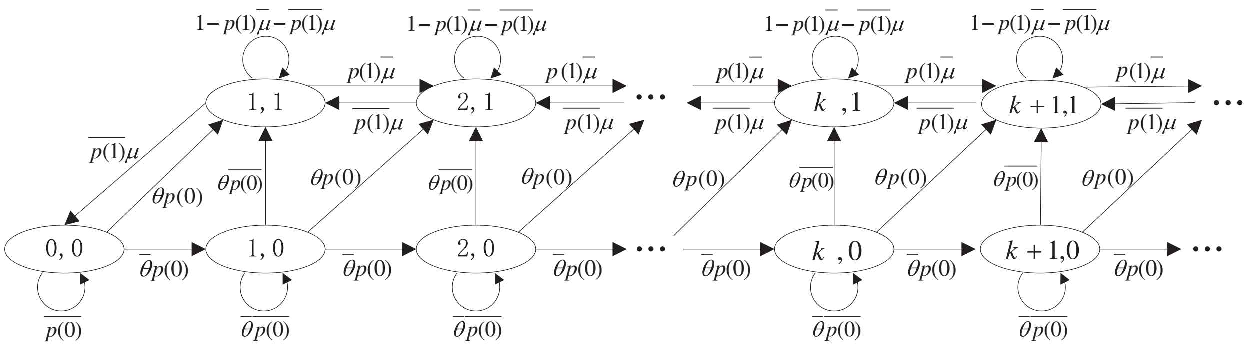

Fig. 3. Transition rate diagram for the $(q(0), q(1))$ mixed strategy in the almost unobservable queue.

Theorem 4.1. *Consider a Geo/Geo/1 queue with multiple vacations and* $\beta(0) \neq \alpha(1)$, *in which customers observe the state j of the server upon arrival and enter with probability* $q(j)$, *i.e. they follow the mixed policy* $(q(0), q(1))$. *The stationary probabilities* $\{\pi''_{kj} | (k,j) \in \Omega_{au}\}$ *are*

$$\pi''_{00} = \frac{\theta(\mu - p(1))}{(\theta + \bar{\theta}p(0))(\mu - p(1) + p(0))}, \tag{40}$$

$$(\pi''_{10}, \pi''_{11}) = \left(\frac{\bar{\theta}p(0)}{1 - \bar{\theta}p(0)}, \frac{p(0)}{p(1)\mu} \right) \pi''_{00}, \tag{41}$$

$$(\pi''_{k0}, \pi''_{k1}) = (\pi''_{10}, \pi''_{11}) \boldsymbol{R}^{k-1}, \quad k \geqslant 1. \tag{42}$$

From Theorem 4.1, we can easily obtain the state probabilities of a server in steady-state

$$P(J=0) = \sum_{k=0}^{\infty} \pi''_{k0} = \frac{\mu - p(1)}{\mu - p(1) + p(0)}, \tag{43}$$

$$P(J=1) = \sum_{k=1}^{\infty} \pi''_{k1} = \frac{p(0)}{\mu - p(1) + p(0)}. \tag{44}$$

Now we consider a customer who finds the server at state j upon arrival, thus his mean sojourn time is

$$Se(j) = \frac{E[L^- | j] + 1}{\mu} + \frac{1-j}{\theta},$$

where $E[L^-|j]$ is the expected number of customers in system found by an arrival, given that the server is found at state j. The expected net benefit of such a customer who decides to enter is

$$Be(j) = R - \frac{C(E[L^-|j] + 1)}{\mu} - \frac{C(1-j)}{\theta}. \tag{45}$$

We thus need to compute $E[L^-|j]$ when all customers follow the same mixed strategy $(q(0), q(1))$. Since the probability $\pi^-_{L|J}(k|j)$, that an arrival finds k customers in system, given that the server is found at state j is

$$\begin{cases} \pi^-_{L|0}(k|0) = \frac{\pi''_{k0}}{P(J=0)}, & k = 0, 1, 2, \ldots, \\ \pi^-_{L|1}(k|1) = \frac{\pi''_{k1}}{P(J=1)}, & k = 1, 2, \ldots. \end{cases} \tag{46}$$

Substituting (40)–(44) into (46) and from $E[L^-|j] = \sum_{k=j}^{\infty} k \pi^-_{L|J}(k|j)$, we get

$$E[L^-|0] = \frac{\bar{\theta}p(0)}{\theta}, \tag{47}$$

$$E[L^-|1] = \frac{\bar{\theta}p(0)}{\theta} + \frac{\mu\overline{p(1)}}{\mu - p(1)}. \tag{48}$$

Then the social benefit per time unit when all customers follow the mixed policy $(q(0), q(1))$ can be easily computed as

$$SB_{au} = p\frac{\mu - p(1)}{\mu - p(1) + p(0)} q(0) \left[R - \frac{C}{\mu} \frac{\theta + \bar{\theta}p(0)}{\theta} - \frac{C}{\theta} \right] + p \frac{p(0)}{\mu - p(1) + p(0)} q(1) \left[R - \frac{C}{\mu} \left(\frac{\theta + \bar{\theta}p(0)}{\theta} + \frac{\mu\overline{p(1)}}{\mu - p(1)} \right) \right].$$

Substituting (47) and (48) into (45) we can identify mixed equilibrium strategies for the almost unobservable model.

Theorem 4.2. *In the almost unobservable Geo/Geo/1 queue with multiple vacations, there exists a unique mixed strategy* $(q_e(0), q_e(1))$ *'observe* J_n *and enter with probability* $q_e(J_n)$*' where the vector* $(q_e(0), q_e(1))$ *is given as follows:*

Case I: $\frac{1}{\theta} < \frac{1}{\mu}$,

$$(q_e(0), q_e(1)) = \begin{cases} \left(\frac{1}{\bar{\theta}p} \left(\frac{\mu\theta R}{C} - \mu - \theta \right), 0 \right), & R \in \left(\frac{C}{\mu} + \frac{C}{\theta}, \frac{C(\theta + \bar{\theta}p)}{\mu\theta} + \frac{C}{\theta} \right), \\ (1, 0), & R \in \left[\frac{C(\theta + \bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}, \frac{C}{\mu} + \frac{C(\theta + \bar{\theta}p)}{\mu\theta} \right), \\ \left(1, \frac{\mu(Cp + 2C\theta - Cp\theta - \mu\theta R)}{p(Cp + C\theta - Cp\theta + C\mu\theta - \mu\theta R)} \right), & R \in \left[\frac{C}{\mu} + \frac{C(\theta + \bar{\theta}p)}{\mu\theta}, \frac{C\bar{p}}{\mu - p} + \frac{C(\theta + \bar{\theta}p)}{\mu\theta} \right), \\ (1, 1), & R \in \left[\frac{C\bar{p}}{\mu - p} + \frac{C(\theta + \bar{\theta}p)}{\mu\theta}, \infty \right). \end{cases}$$

Case II: $\frac{1}{\mu} \leqslant \frac{1}{\theta} \leqslant \frac{\bar{p}}{\mu-p}$,

$$(q_e(0), q_e(1)) = \begin{cases} \left(\frac{1}{\bar{\theta}p}\left(\frac{\mu\theta R}{C} - \mu - \theta\right), \frac{\mu-\theta}{\bar{\theta}p}\right), & R \in \left(\frac{C}{\mu} + \frac{C}{\theta}, \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}\right), \\ \left(1, \frac{\mu(Cp+2C\theta-Cp\theta-\mu\theta R)}{p(Cp+C\theta-Cp\theta+C\mu\theta-\mu\theta R)}\right), & R \in \left[\frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}, \frac{C\bar{p}}{\mu-p} + \frac{C(\theta+\bar{\theta}p)}{\mu\theta}\right), \\ (1,1), & R \in \left[\frac{C\bar{p}}{\mu-p} + \frac{C(\theta+\bar{\theta}p)}{\mu\theta}, \infty\right). \end{cases}$$

Case III: $\frac{\bar{p}}{\mu-p} < \frac{1}{\theta}$,

$$(q_e(0), q_e(1)) = \begin{cases} \left(\frac{1}{\bar{\theta}p}\left(\frac{\mu\theta R}{C} - \mu - \theta\right), 1\right), & R \in \left(\frac{C}{\mu} + \frac{C}{\theta}, \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}\right), \\ (1,1) & R \in \left[\frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}, \infty\right). \end{cases}$$

Proof. Consider a tagged customer who finds the server at state 0 upon arrival. If he decides to enter, his expected net benefit is

$$Be(0) = R - \frac{C(E[L^-|0] + 1)}{\mu} - \frac{C}{\theta} = R - \frac{C(\theta + \bar{\theta}p(0))}{\mu\theta} - \frac{C}{\theta}.$$

Therefore we have two cases:

Case 1: $Be(0) \leqslant 0$ *i.e.* $\frac{C}{\mu} + \frac{C}{\theta} < R \leqslant \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}$.

In this case if all customers who find the system empty enter with probability $q_e(0) = 1$, then the tagged customer suffers a negative expected benefit if he decides to enter. Hence $q_e(0) = 1$ does not lead to an equilibrium. Similarly, if all customers use $q_e(0) = 0$, then the tagged customer receives a positive benefit from entering. Thus $q_e(0) = 0$ also cannot be part of an equilibrium mixed strategy. Therefore, there exists a unique $q_e(0)$ satisfying

$$R - \frac{C(\theta + \bar{\theta}pq_e(0))}{\mu\theta} - \frac{C}{\theta} = 0,$$

for which customers are indifferent between entering and balking. This is given by

$$q_e(0) = \frac{1}{\bar{\theta}p}\left(\frac{\mu\theta R}{C} - \mu - \theta\right). \tag{49}$$

Case 2: $Be(0) > 0$ *i.e.* $R > \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}$.

In this case, for every strategy of the other customers, the tagged customer has a positive expected net benefit if he decides to enter. Therefore $q_e(0) = 1$.

We next consider $q_e(1)$ and tag a customer who finds the server at state 1 upon arrival. If he decides to enter, his expected net benefit is

$$Be(1) = R - \frac{C(E[L^-|1] + 1)}{\mu} = R - \frac{C\overline{p(1)}}{\mu - p(1)} - \frac{C(\theta + \bar{\theta}p(0))}{\mu\theta} = \begin{cases} \frac{C}{\theta} - \frac{C\overline{p(1)}}{\mu-p(1)}, & \text{in case 1}, \\ R - \frac{C\overline{p(1)}}{\mu-p(1)} - \frac{C(\theta+\bar{\theta}p)}{\mu\theta}, & \text{in case 2}. \end{cases} \tag{50}$$

Therefore to find $q_e(1)$ in equilibrium, we must examine Case 1 and Case 2 separately and consider the following subcases in each:

Case 1a: $\frac{C}{\mu} + \frac{C}{\theta} < R \leqslant \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}$ *and* $\frac{C}{\theta} < \frac{C}{\mu}$.

$$(q_e(0), q_e(1)) = \left(\frac{1}{\bar{\theta}p}\left(\frac{\mu\theta R}{C} - \mu - \theta\right), 0\right).$$

Case 1b: $\frac{C}{\mu} + \frac{C}{\theta} < R \leqslant \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}$ *and* $\frac{C}{\mu} \leqslant \frac{C}{\theta} \leqslant \frac{C\bar{p}}{\mu-p}$.

$$(q_e(0), q_e(1)) = \left(\frac{1}{\bar{\theta}p}\left(\frac{\mu\theta R}{C} - \mu - \theta\right), \frac{\mu - \theta}{\bar{\theta}p}\right).$$

Case 1c: $\frac{C}{\mu} + \frac{C}{\theta} < R \leqslant \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}$ *and* $\frac{C\bar{p}}{\mu-p} < \frac{C}{\theta}$.

$$(q_e(0), q_e(1)) = \left(\frac{1}{\bar{\theta}p}\left(\frac{\mu\theta R}{C} - \mu - \theta\right), 1\right).$$

Case 2a: $R > \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}$ *and* $R < \frac{C}{\mu} + \frac{C(\theta+\bar{\theta}p)}{\mu\theta}$.

$$(q_e(0), q_e(1)) = (1, 0).$$

Case 2b: $R > \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}$ *and* $\frac{C}{\mu} + \frac{C(\theta+\bar{\theta}p)}{\mu\theta} \leqslant R \leqslant \frac{C\bar{p}}{\mu-p} + \frac{C(\theta+\bar{\theta}p)}{\mu\theta}$.

$$(q_e(0), q_e(1)) = \left(1, \frac{\mu(Cp + 2C\theta - Cp\theta - \mu\theta R)}{p(Cp + C\theta - Cp\theta + C\mu\theta - \mu\theta R)}\right).$$

Case 2c: $R > \frac{C(\theta+\bar{\theta}p)}{\mu\theta} + \frac{C}{\theta}$ *and* $R > \frac{C\bar{p}}{\mu-p} + \frac{C(\theta+\bar{\theta}p)}{\mu\theta}$.

$$(q_e(0), q_e(1)) = (1, 1).$$

By rearranging Cases 1a–2c as R varies from $\frac{C}{\mu} + \frac{C}{\theta}$ to infinity, keeping the operating parameters p, μ, θ and the waiting cost rate C fixed, we obtain Case I-III in the theorem statement. □

It seems reasonable at first glance that arriving customers are less willing to enter the system when they find the server on vacation, since they have to wait for the left vacation time, i.e., we might expect that $q_e(0) \leqslant q_e(1)$. However, this is not generally true. As Theorem 4.2 shows: in case I it is always true that $q_e(0) \geqslant q_e(1)$. In fact, consider a system with small mean vacation time and concentrate on a tagged customer. If he is given the information that the server is on vacation, then he knows that he must wait for the left vacation time. On the other hand, he expects that few customers are ahead of him, because the system is on vacation and the mean vacation time is small. Thus it is optimal for the tagged customer to enter.

4.2. Fully unobservable queue

We finally consider the fully unobservable case, where customers observe neither the state of the system nor the queue length. Here a mixed strategy for a customer is specified by the probability q of entering. The stationary distribution of the system state is from Theorem 4.1 by taking $q(0) = q(1) = q$ and the state space Ω_{fu} is identical with Ω_{au}. The transition rate diagram is depicted in Fig. 4, where $p' = pq$.

Theorem 4.3. *Consider a fully unobservable Geo/Geo/1 queue with multiple vacations and $\beta \neq \alpha$, in which customers enter with probability q, i.e. they follow the mixed policy q. The stationary probabilities $\{\pi'''_{kj} | (k,j) \in \Omega_{fu}\}$ are*

$$\pi'''_{00} = \frac{\theta(\mu - p')}{(\theta + \bar{\theta}p')\mu},$$

$$(\pi'''_{10}, \pi'''_{11}) = \left(\frac{\bar{\theta}p'}{1 - \bar{\theta}\bar{p}'}, \frac{p'}{\bar{p}'\mu}\right)\pi'''_{00},$$

$$(\pi'''_{k0}, \pi'''_{k1}) = (\pi'''_{10}, \pi'''_{11})\widetilde{\boldsymbol{R}}^{k-1}, \quad k \geqslant 1,$$

where

$$\alpha = \frac{p'\bar{\mu}}{\bar{p}'\mu}, \quad \beta = \frac{\bar{\theta}p'}{1 - \bar{\theta}\bar{p}'}, \widetilde{\boldsymbol{R}} = \begin{bmatrix} \beta & \frac{p'}{\bar{p}'\mu} \\ 0 & \alpha \end{bmatrix}.$$

From (43), (44), (47) and (48) with $q(0) = q(1) = q$, we can easily get:

$$P(J = 0) = \frac{\mu - p'}{\mu}, \quad P(J = 1) = \frac{p'}{\mu}, \tag{51}$$

$$E[L^-|0] = \frac{\bar{\theta}p'}{\theta}, \quad E[L^-|1] = \frac{\bar{\theta}p'}{\theta} + \frac{\mu\bar{p}'}{\mu - p'}. \tag{52}$$

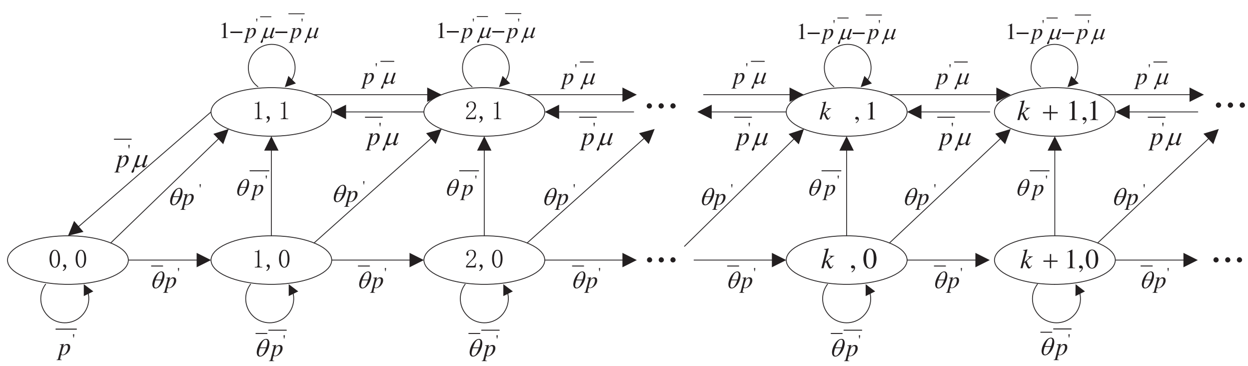

Fig. 4. Transition rate diagram for the q mixed strategy in the fully unobservable queue.

Hence,

$$SB_{fu} = p'\left[R - \frac{C}{\mu}\left(\frac{\mu + \theta - \theta p'}{\theta} + \frac{p'\overline{p'}}{\mu - p'}\right)\right].$$

Theorem 4.4. *In the fully unobservable Geo/Geo/1 queue with multiple vacations, there exists a unique mixed equilibrium strategy 'enter with probability q_e', where q_e is given by*

$$q_e = \begin{cases} \frac{\mu\theta R - C\mu - C\theta}{p(R\theta - C - C\theta)} & R \in \left(\frac{C}{\theta} + \frac{C}{\mu}, \frac{C}{\theta} + \frac{C\overline{p}}{\mu - p}\right), \\ 1 & R \in \left[\frac{C}{\theta} + \frac{C\overline{p}}{\mu - p}, \infty\right). \end{cases} \tag{53}$$

Proof. We consider a tagged customer at his arrival instant. If he decides to enter his mean sojourn time

$$Se = \frac{E[L^-] + 1}{\mu} + \frac{P(J = 0)}{\theta}.$$

From (51) and (52), we get

$$E[L^-] = \frac{\overline{\theta}p'}{\theta} + \frac{p'\overline{p'}}{\mu - p'}.$$

Hence the expected net benefit

$$Be = R - \frac{C(1 - pq)}{\mu - pq} - \frac{C}{\theta}. \tag{54}$$

When $R \in \left(\frac{C}{\theta} + \frac{C}{\mu}, \frac{C}{\theta} + \frac{C\overline{p}}{\mu - p}\right)$, we find (54) has a unique root in $(0, 1)$ which gives the first branch of (53). When $R \in \left[\frac{C}{\theta} + \frac{C\overline{p}}{\mu - p}, \infty\right)$, Be is positive for every q, thus the best response is 1 and the unique equilibrium point is $q_e = 1$, which gives the second branch of (53). □

5. Numerical examples

In this section, according to the analysis above, we firstly present a set of numerical experiments to show the effect of the information level as well as several parameters on the behavior of the system. And then we give an example to show the use of results.

5.1. Numerical experiments

Here we concern about the values of the equilibrium thresholds for the observable systems and the values of the equilibrium entrance probabilities for the unobservable systems as well as the social benefit per unit time when customers follow the corresponding equilibrium strategies.

We first consider the observable models and explore the sensitivity of the equilibrium thresholds with respect to the arrival rate p, vacation rate θ and service reward R. From the three sub-figures (a_1, b_1, c_1) on the left of Fig. 5, we can make an interesting conjecture that the equilibrium thresholds $\{k_L, \ldots, k_U\}$ for the almost observable case always locate in $(L_e(0), L_e(1))$ for the fully observable case. In other words, the thresholds in the almost observable model have intermediate values between the two separate thresholds when arriving customers observe the state of the server.

Concerning the sensitivity of the equilibrium thresholds, we can make the following observations. When the arrival rate p varies, the fully observable thresholds remain fixed since the arrival rate is irrelevant to the customer's decision when he has full state information. On the other hand, the almost observable thresholds increase with the arrival rate, which means that if an arriving customer is told the information of the present queue length, then he is more likely to enter when the arrival rate is higher. This phenomenon in the almost observable case that customers in equilibrium tend to imitate the behavior of other customers is of the 'Follow-The-Crowd' (FTC) type. The reason for this phenomenon is that when the arrival rate is high, it is probably that the server is active, therefore the expected delay from server vacation is reduced. However, all thresholds increase when the vacation rate θ varies, except that $L_e(1)$ remains constant. This is certainly intuitive, because when the server vacation is shorter, customers generally have a greater incentive to enter both in the fully and almost observable systems. Finally, along with the increasing of the service reward R, no matter in the fully observable model or in the almost observable model, the thresholds increase in a linear fashion, which is expected from Theorems 3.1 and 3.4.

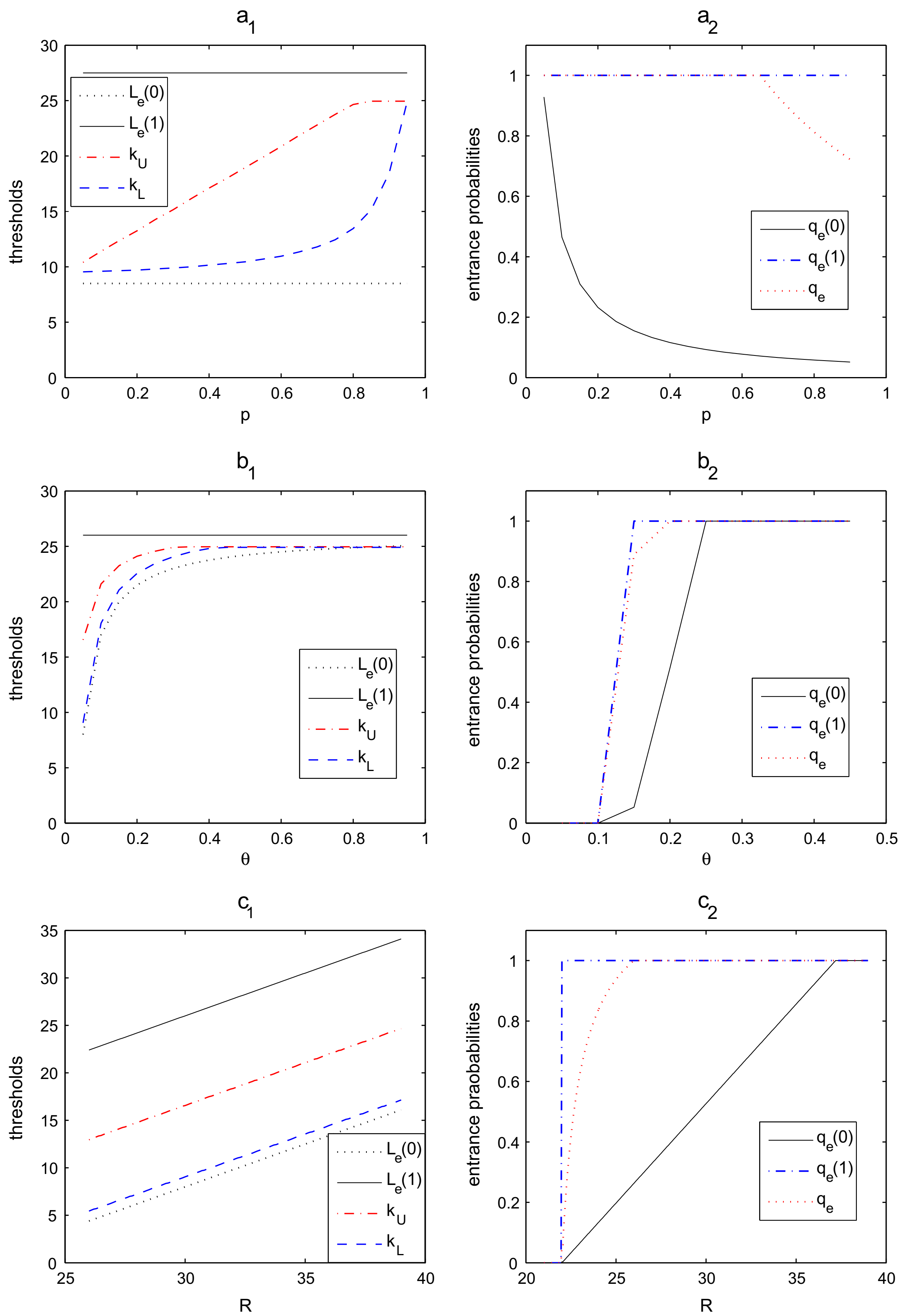

Fig. 5. Equilibrium thresholds for the observable systems. Sensitivity with respect to: a_1. p, for $\mu = 0.95$, $\theta = 0.05$, $C = 1$, $R = 30$; b_1. θ, for $p = 0.4$, $\mu = 0.9$, $C = 1$, $R = 30$; c_1. R, for $p = 0.4$, $\mu = 0.9$, $\theta = 0.05$, $C = 1$. Equilibrium entrance probabilities for the unobservable systems. Sensitivity with respect to: a_2. p, for $\mu = 0.95, \theta = 0.3, C = 1, R = 4.5$; b_2. θ, for $p = 0.9, \mu = 0.95, C = 1, R = 8$; c_2. R, for $p = 0.4, \mu = 0.5, \theta = 0.05, C = 1$.

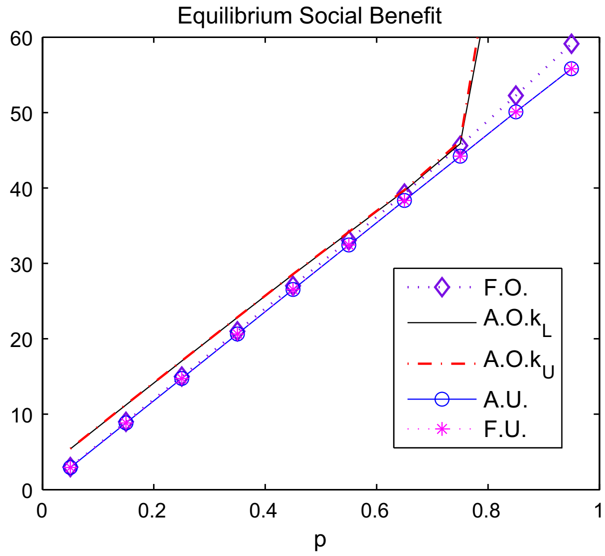

Fig. 6. Social benefit for different information levels. Sensitivity with respect to p, for $\mu = 0.99, \theta = 0.05, R = 80, C = 1$.

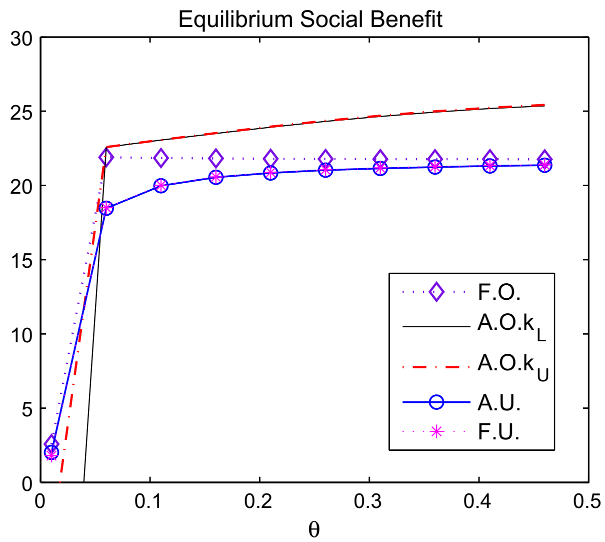

Fig. 7. Social benefit for different information levels. Sensitivity with respect to θ, for $\mu = 0.99, p = 0.2, R = 110, C = 1$.

Next we turn to the unobservable systems and explore the sensitivity of the equilibrium entrance probabilities. The results are presented in the three sub-figures (a_2, b_2, c_2) on the right of Fig. 5. A general observation is that the entrance probability in the fully unobservable queueing system is always inside the interval formed by the two entrance probabilities in the almost unobservable queueing system, which is similar to the results in the observable models. Therefore when customers are not told the server state, they join the queue with a probability between those in the two separate cases provided that the state of the server can be observed by arriving customers. With regard to the sensitivity of entrance probabilities, we observe that they are decreasing with respect to p. Therefore when the arrival rate increases, customers are less willing to enter the system. This is in contrast to the observable models, where thresholds are increasing with p. The reason for the difference is that here the information about the present queue length is not available, when the arrival rate is higher, arriving customers expect that the system is more loaded and are less inclined to enter. Furthermore the entrance probabilities are all increasing with respect to θ and R, which is intuitive.

Figs. 6–8 are concerned with the social benefit under the equilibrium balking strategy for the different information levels. For the almost observable case, we only present the social benefit under the two extreme thresholds k_L and k_U. In figures we notice that the social benefit in the almost observable model is generally more than those in other cases while the value for the fully observable model is always in an intermediate position. In addition, the difference in social benefit is small between the fully unobservable and almost unobservable systems. These phenomena indicate that the customers as a whole are generally better off when they are informed of the number of present customers while the information about the server state is

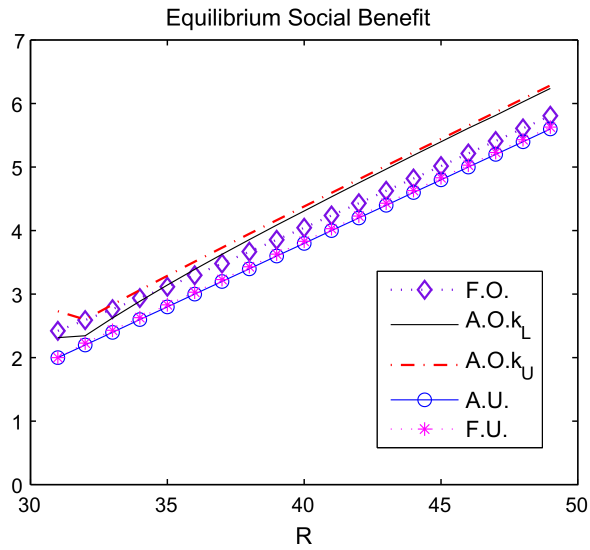

Fig. 8. Social benefit for different information levels. Sensitivity with respect to R, for $\mu = 0.99, p = 0.2, \theta = 0.05, C = 1$.

not so beneficial. Moreover, the full information about the queueing system may actually hurt the social benefit. In other words, the additional information on the server state is not very helpful for increasing equilibrium social benefit. As to the sensitivity of the equilibrium social benefit, we find that it is increasing with the arrival rate, vacation rate and service reward, which is intuitive. Now we pay more attention to Fig. 7 in which from a certain value of θ, the trend of the equilibrium social benefit becomes smooth and steady. This is because that with the vacation rate increasing, on the one hand the vacation time reduces which benefits the social benefit, on the other hand since more people left for service, the service time increases which has a detrimental effect on the social benefit. When the vacation rate increases to a certain point, the positive part counterbalances the negative part, therefore the equilibrium social benefit tends to reach a steady value.

5.2. An example

In daily life, it is common to see the queueing phenomenon, especially in the banks where lots of people are waiting for service. Now banks generally use the automatical calling system, from which the arriving customer takes a queue number. The data series recorded by bank queueing machines are used to analyze the queueing rules of service counters. We get data series from a regional bank and show it in Table 1, where k represents the data number and T_k represents the data arrival time. From Table 1, we obtain that the mean arrival rate is 0.6, i.e. $p = 0.6$. If the automatical calling system is in work order, customers could observe the queue length, who are thus in an observable queue, there exist equilibrium pure threshold strategies. However, if not, there exist equilibrium mixed strategies. The equilibrium strategies both in observable and unobservable queues are shown in Table 2, which are obtained from Theorems 3.1, 3.4, 4.4 and the case III of Theorem 4.2. In an observable model, customers have a unique equilibrium strategy when the server state is provided. More concretely, an arriving customer could follow the equilibrium strategy that 'enter if $L_n \leqslant 8$ and balk otherwise' when the server is on vacation, or follow the equilibrium strategy that 'enter if $L_n \leqslant 27$ and balk otherwise' when the server is busy. However, if the information of server state is not provided, all pure threshold strategies 'enter if $L_n \leqslant L_e$ for some fixed $L_e \in \{9, 10, \cdots, 20\}$ and balk otherwise' are equilibrium. We could find that L_e locates between $L_e(0)$ and $L_e(1)$. On the other thing, in the unobservable case, if a customer is only informed of the state of the server upon arrival, there exists a mixed strategy specified by $(q_e(0), q_e(1))$. When the server is on vacation, the equilibrium strategy is 'enter with probability 0.7456'. However, when the server is busy, the equilibrium strategy is 'always enter'. If customers could observe neither the server state nor the queue length, there exists a unique mixed strategy 'enter with probability 1'. It is less convenient to use the equilibrium strategies as the entrance probabilities in unobservable case than as the thresholds in observable case.

Figs. 9 and 10 are concerned with the sensitivity of equilibrium strategies and social benefits with respect to μ. We can make the following observations. No matter in the observable case nor in the unobservable case, customers are more likely

Table 1
Data arrival time in an hour.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T_k	0	2	3	5	6	8	11	12	14	15	16	17	19	22	23	25	26	29
k	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
T_k	30	32	33	35	36	38	41	42	44	45	46	47	49	52	53	55	56	60

Table 2
Equilibrium strategies. ($p = 0.6, \mu = 0.95, \theta = 0.05, C = 1, R = 30$).

Observable case			Unobservable case		
$L_e(0)$	$L_e(1)$	$L_e\{k_L, k_L + 1, \ldots, k_U\}$	$q_e(0)$	$q_e(1)$	q_e
8	27	9, 10,. . ., 20	0.7456	1	1

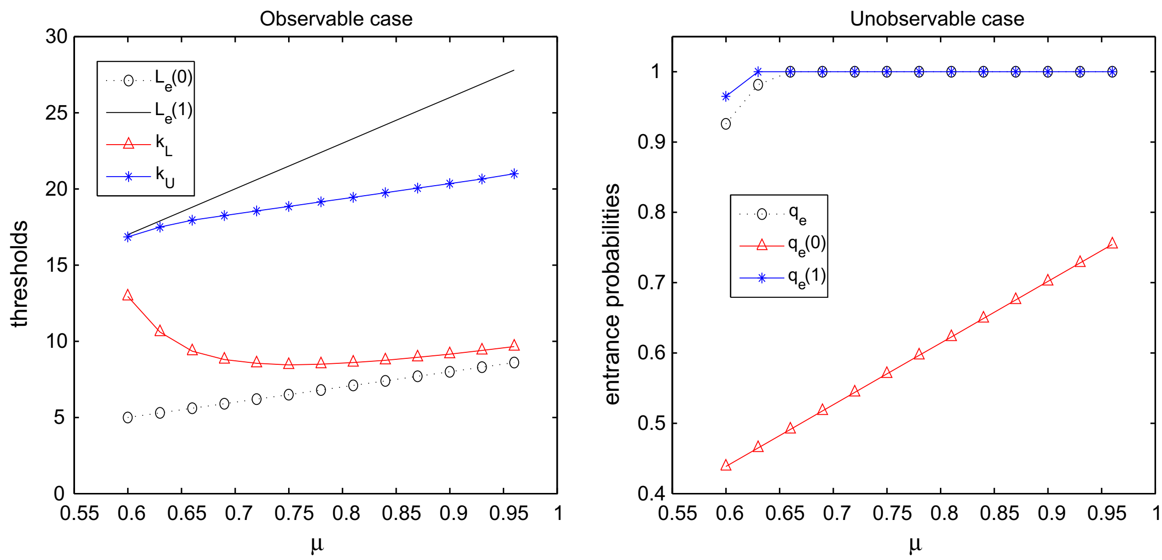

Fig. 9. Equilibrium strategies. Sensitivity with respect to μ. ($p = 0.6, \theta = 0.05, C = 1, R = 30$).

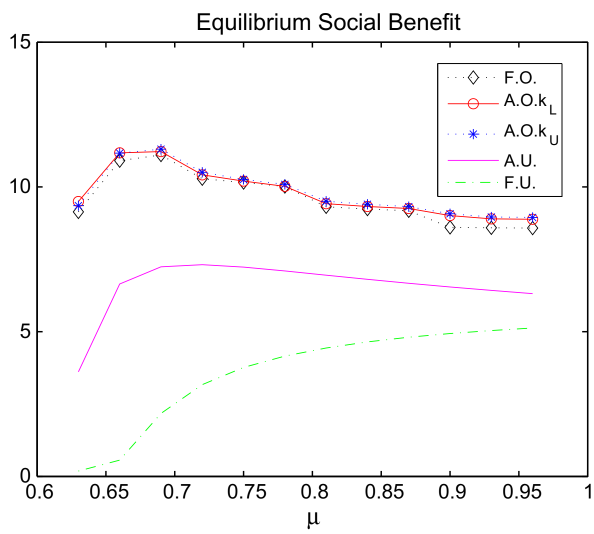

Fig. 10. Social benefit for different information levels. Sensitivity with respect to μ. ($p = 0.6, \theta = 0.05, C = 1, R = 30$).

to enter when the service rate is higher, which is intuitive. The difference in social benefit is small between the fully and almost observable case, while there may be significant differences between the observable and the unobservable models. Thus, it may be argued that customers as a whole are more better off when they are informed of the queue length than the information about the server state. Moreover, we notice that the social benefits in the almost observable model are generally more than those in the fully observable case. This might mean the full information about the queueing system may actually hurt the social benefit. In other words, in an observable model the additional information on the server state is not very helpful for increasing equilibrium social benefit. However, in an unobservable queue, the social benefit when customers could observe the server state is obviously more than the case provided with no information. Those observations indicate that when the service rate μ varies, the information of the queue length is the most beneficial factor to increase the social benefit, while the server state is helpful only in the unobservable case.

6. Conclusion

In this paper, we have explored the equilibrium behavior in the discrete-time $Geo/Geo/1$ queue under multiple vacation policy. On the one hand, the discrete-time queue where the inter-arrival time and service time are positive integer random variables is more suitable to model and analyze the digital communication system. On the other hand, customers have the right to make decisions according to the accurate situation, which is more sensible than the classical viewpoint in queueing theory that decisions are made by the servers and customers are forced to follow them. To the best of the authors' knowledge, there is no work concerning the equilibrium balking strategies in discrete-time queues with multiple vacations. Besides, this is the first time that the multiple vacation policy is introduced into the economics of queues.

We have classified four cases with respect to the level of information provided to arriving customers and obtained the equilibrium strategies for each case. The stationary system behavior has been analyzed and a variety of stationary performance measures have been developed under the corresponding strategies. Furthermore, we have discussed the sensitivity of equilibrium behavior with respect to various parameters as well as the effect of the information level on the equilibrium social benefit. The research results could instruct the customers to take optimal strategies to reduce the loss of queueing. In addition, the study could provide the managers with a good reference to discuss the pricing issues in queuing systems, such as the entrance fee, the priority fee and so on. Further extensions to the work would be to consider the equilibrium customer behavior in models under single vacation policy.

Acknowledgements

The authors thank the support from Humanities and Social Science Fund in Ministry of Education in China (Nos. 10YJC630114; 07JA630027) and Post-Doctoral Research Project of China (No. 92169).

References

[1] P. Naor, The regulation of queue size by levying tolls, Econometrica 37 (1969) 15–24.
[2] S.G. Johansen, S. Stidham Jr., Control of arrivals to a stochastic input–output system, Adv. Appl. Probab. 12 (1980) 972–999.
[3] S. Stidham Jr., Optimal control of admissions to a queueing system, IEEE Trans. Automat. Control 30 (1985) 705–713.
[4] H. Mendelson, S. Whang, Optimal incentive-compatible priority pricing for the $M/M/1$ queue, Oper. Res. 38 (1990) 870–883.
[5] H. Chen, M. Frank, State dependent pricing with a queue, IIE Trans. 33 (2001) 847–860.
[6] J. Erlichman, R. Hassin, Equilibrium solutions in the observable $M/M/1$ queue with overtaking, Naval Res. Logist. 58 (2011) 107–122.
[7] N.M. Edelson, K. Hildebrand, Congestion tolls for Poisson queueing processes, Econometrica 43 (1975) 81–92.
[8] S.C. Littlechild, Optimal arrival rate in a simple queueing system, Int. J. Product. Res. 12 (1974) 391–397.
[9] H. Chen, M. Frank, Monopoly pricing when customers queue, IIE Trans. 36 (2004) 569–581.
[10] K.R. Balachandran, Incentive and regulation in queues, in: M.J. Beckmann, M.N. Gopalan, R. Subramanian (Eds.), Lecture Notes in Economics and Mathematical Systems, vol. 370, Springer-Verlag, Berlin-Heidelberg, 1991, pp. 162–176.
[11] R. Hassin, M. Haviv, To Queue or not to Queue: Equilibrium Behavior in Queueing Systems, Kluwer Academic Publishers, Dordrecht, 2003.
[12] H. Takagi, Queueing Analysis: A Foundation of Performance Evaluation, vol. 3, North-Holland, Amsterdam, 1993.
[13] Z.G. Zhang, N. Tian, Discrete-time $Geo/G/1$ queue with multiple adaptive vacations, Queueing Syst. 38 (2001) 419–429.
[14] N. Tian, Z.G. Zhang, The discrete-time $GI/Geo/1$ queue with multiple vacations, Queueing Syst. 40 (2002) 283–294.
[15] S.K. Samanta, M.L. Chaudhry, U.C. Gupta, Discrete-time $Geo^X/G^{(a,b)}/1/N$ queues with single and multiple vacations, Math. Comput. Model. 45 (2007) 93–108.
[16] Y. Tang, X. Yun, S. Huang, Discrete-time $Geo^X/G/1$ queue with unreliable server and multiple adaptive delayed vacations, J. Comput. Appl. Math. 220 (2008) 439–455.
[17] A. Burnetas, A. Economou, Equilibrium customer strategies in a single server Markovian queue with setup times, Queueing Syst. 56 (2007) 213–228.
[18] A. Economou, S. Kanta, Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs, Oper. Res. Lett. 36 (2008) 696–699.
[19] W. Sun, P. Guo, N. Tian, Equilibrium threshold strategies in observable queueing systems with setup/closedown times, CEJOR 18 (2010) 241–268.
[20] A. Economou, A. Gómez-Corral, S. Kanta, Optimal balking strategies in single-server queues with general service and vacation times, Performance Evaluat. 68 (2011) 967–982.
[21] W. Liu, Y. Ma, J. Li, Equilibrium threshold strategies in observable queueing systems under single vacation policy, Appl. Math. Model. 36 (2012) 6186–6202.
[22] S.N. Elaydi, An Introduction to Difference Equations, Springer, New York, 1999.