Applied Mathematics and Computation 248 (2014) 195–214

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

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.

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

* Corresponding author.
E-mail addresses: wsun@ysu.edu.cn (W. Sun), shiyongli@ysu.edu.cn (S. Li), liquanlin@tsinghua.edu.cn (Q.-L. Li).

http://dx.doi.org/10.1016/j.amc.2014.09.116
0096-3003/

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 $n \geqslant 2$. 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 equably 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 λ and the server's regular service rate is μ_b. Once the system becomes empty, the server begins a first-stage working vacation V_1, which follows an exponential distribution with parameter θ. During this period, an arriving customer is served at a service rate μ_{v_1}, which is lower than μ_b. 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 μ_{v_1} to μ_{v_2}, which is also lower than μ_b. 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 M/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