Hindawi
Complexity
Volume 2020, Article ID 6015372, 19 pages
https://doi.org/10.1155/2020/6015372

Research Article

Strategic Behavior of Customers and Optimal Control for Batch Service Polling Systems with Priorities

Tao Jiang,[1] **Xingzheng Lu**,[1] **Lu Liu**,[1] **Jun Lv**,[1] **and Xudong Chai**[2]

[1]*College of Economics and Management, Shandong University of Science and Technology, Qingdao, Shandong 266590, China*
[2]*School of Science, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China*

Correspondence should be addressed to Tao Jiang; jtao0728@163.com

Received 16 June 2020; Revised 13 August 2020; Accepted 25 August 2020; Published 7 September 2020

Academic Editor: Qingling Wang

During the past few years, batch service systems have attracted considerable attention due to their wide area of applications. In this present paper, we study a special batch service polling system (the so-called Israeli queue) with priorities. Different from the previous papers which focus on the performance analysis, we aim to investigate the strategic behavior of customers and optimal design for the underlying queueing model. By considering two levels of information (observable and unobservable) provided upon customers' arrival, we, respectively, derive the equilibrium strategies of high-priority and low-priority customers, regarding the joining or balking dilemma. We also present some numerical examples to reveal the impacts of several parameters on the equilibrium strategies, together with some intuitive explanations. Finally, we formulate the revenue function of the service provider and present the Particle Swarm Optimization algorithm to seek the optimal service prices for the high-priority and low-priority customers to maximize the service provider's revenue under the two levels of information.

1. Introduction

Nowadays, queueing models with batch services frequently encountered in practice and studies addressing the batch service queueing systems have been investigated extensively owing to their wide area of applications, such as the transportation systems, tourist services, and telecommunication systems. So far, the literature on this topic can be roughly divided into two categories. The first category is to analyse the batch service models with fixed batch size. For example, at the airports, the airplane departs once all seats are occupied. In tourism services, the tour guide starts his/her work as soon as a fixed number of travelers have been accumulated. Another example is that, in tourist areas, sightseeing bus will start off while the bus is full of travelers. The second category is to deal with the batch service models with infinite batch size. For example, at the bus stations and metro stations, the bus and the train can remove all present passengers. Such batch service models can be also referred to as stochastic clearing models.

Due to the wide range of applications on batch service models, in this paper, we will consider a special polling system with batch service of an unlimited size, i.e., the so-called "Israeli queue," which has been widely used in practice and extensively studied in the literature. As pointed out by Perel and Yechiali [1], Israeli queue stands for a real situation of a physical waiting queue for buying tickets at stations and it is a multiqueue, single-server polling system with unlimited-size batch service; after completion of a visit at a queue, the next queue to be served is the one whose first customer has been waiting for the longest time. For example, when buying ticket at a railway station, a new arriving customer first searches for the acquaintance already standing in the waiting line; if he/she finds one, he/she will ask for the acquaintance to buy the tickets along with him/her. Since the service time of a batch does not depend on its size, the waiting time of customers is not affected by the customers that join the queue in front of them. Actually, this setting can be also seen in real life especially in transportation systems. At a bus station or a metro station, there may be multiple routes through the station. While a bus or a train visits the

station, it can pick up all passengers going to the same destination. Therefore, the passengers may be served prior to the passengers who have already arrived at the station before them, since each arriving bus goes to a specific destination, and passengers waiting to go to different destinations have to wait for their dedicated bus. Therefore, for the Israeli queue, this may not be a real interjection, since the waiting time of customers is not affected by the customers who join the queue in front of them; nonetheless, we could come to an agreement that the queueing phenomenon no longer follows the first-come, first-served (FCFS) discipline and the customers may be served prior to the customers who have already arrived at the system before them.

However, the majority of the papers on batch service systems are devoted to the stationary distribution and the performance evaluation. It seems that only limited attention has been paid to the issue of the batch service systems arising from the customers perspective; i.e., there are few works that consider such batch service systems from the viewpoint of customer behavior (see, e.g., Economou and Manou [2], Manou et al. [3], and Bountali and Economou [4, 5]). Actually, the optimization problems related to the strategic customer behavior in batch service polling systems could not only provide important managerial insights, but also have a wide range of applications in real life. Therefore, it seems important to give a game-theoretic research to the special batch service polling system.

Apart from the game-theoretic research to batch service queueing systems, providing priority access is a widespread method for improving the work efficiency; meanwhile, designing an appropriate priority system is also an efficient scheme for reducing congestion. Therefore, the problems of priority queueing models are frequently encountered in practice and they have been extensively studied due to their wide applications in some health-care systems, production systems, communication systems, etc. For more studies on priority queues, we may refer the interested readers to Drekic and Grassmann [6], Jouini and Roubos [7], Takagi [8], Atencia [9], and the reference therein.

To this end, we incorporate customers' joining/balking decisions into a special batch service polling system with priorities. Different from the classical batch service queueing systems, the queueing phenomenon in consideration no longer follows FCFS discipline. Therefore, we are interested in the following research questions: Q1: How do the strategic customers (high-priority and low-priority customers) make their joining/balking decisions under different levels of information (observable and unobservable cases)? Q2: How do the system parameters affect the customers' joining/balking decisions? Q3: How does the revenue-maximizing decision maker choose the optimal pricing decision to maximize his/her revenue?

To address the research questions, we consider a special batch service polling system with priorities, which can be modelled as a two-class, single-server, preemptive priority queueing model, in which the high-priority customers form an M/M/1 queue, while the low-priority customers form the so-called Israeli queue with at most N different groups. The main contribution of our study can be summarized as follows:

(1) In the existing literature, steady-state analysis of the Israel queue has been well studied; however, only limited attention has been paid to the issue of the queueing service from the viewpoint of customer behavior. Such a treatment (which ignores customer behavior) could be sometimes unrealistic especially when customers' strategic behaviors affect service system performance and system revenue. Based on this situation, we aim to investigate the strategic customers' joining/balking decisions of the underlying Israel queue, and this has not been studied in the open literature.

(2) Different from the classical batch service queueing systems with homogeneous strategic customers, the underlying special batch service polling system adopts classification of service through setting priority due to the heterogeneity of sensitive customers. Therefore, under the two levels of information, we distinguish the high-priority and low-priority customers, respectively, derive their joining/balking strategies, and present the equilibrium rates with respect to various system parameters by considering different scenarios.

(3) In order to achieve optimal pricing strategies and maximize the service provider's revenue by distinguishing the observable and unobservable cases, we establish a new revenue function and use the Particle Swarm Optimization (PSO) algorithm to search for the global numerical solution (feasible combination of the service prices for the high-priority and low-priority customers) and show the difference between the two informational cases.

Some interesting insights can be found from the derived results. First, the joining rates of high-priority and low-priority customers are monotonically decreasing in the arriving rate of high-priority customers. The numerical result shows that if the arriving rate of high-priority customers is high, the low-priority customers are inclined to balk. Conversely, if the arriving rate of high-priority customers is low, both of the high-priority and low-priority customers decide to join the system. Therefore, if the service provider wants to ensure his/her revenue, he/she should effectively control the number of high-priority customers. Second, the low-rate service of high-priority customers could cause the high-priority customers to balk, which could also reduce the waiting time of low-priority customers and lead them to join the Israeli queue with a higher probability. And then, as the service rate of high-priority customers gradually increases, the high-priority customers are more willing to join the system due to the reduction of their waiting time, which indirectly affects the joining strategies of low-priority customers. When the service rate of high-priority customers is large enough, the service time of high-priority customers is quite short, and their expected waiting time tends to be zero; that is, the waiting time of high-priority customers has a smaller effect on the low-priority customers, which exactly promotes the low-priority customers to join the Israeli queue. In short, the joining

probability of low-priority customers is sensitive to the service rate of high-priority customers. Therefore, under a fixed cost, determining a suitable service rate of high-priority customers is vitally important for the service provider.

The rest of the paper is organized as follows. A detailed literature review on related works is given in Section 2. In Section 3, we describe the batch service polling system with priorities and construct the reward-cost structure. Section 4 and Section 5 are devoted to deriving the individual optimal threshold strategies and the mixed equilibrium joining strategies under the two informational cases (observable and unobservable). In Section 6, we construct a revenue function of the service provider and employ the PSO algorithm to solve the optimization problem of service provider's revenue under the observable and unobservable settings. We conclude the paper and summarize the main findings of our work in Section 7.

2. Literature Review

This work is closely related to three streams of literature: the Israeli queues, batch service queueing systems with strategic customers, and queueing models with customer service priorities.

Different from some batch service systems that the server serves customers in batches of fixed size, Israeli queue is a specific polling system with batch service of an unlimited size. Customers arrive at the system and form groups; each group is unrestricted in its size and served in one batch; after one group is served, the next group to be visited is the one whose group leader has been waiting for the longest time. This model was first introduced by Van der Wal and Yechiali [10], who considered a multiqueue, single-server polling system with unlimited-size batch, and customers in the same queue can be served together. Lately, Boxma et al. [11] studied the batch service polling system with unrestricted batch size and finite number of queues. Perel and Yechiali [12] considered a batch service queueing model with infinite number of groups and studied the M/M/c, M/M/1/N-type queues, and a priority queueing model. Perel and Yechiali [1, 13, 14], respectively, investigated the Israeli queue with preemptive priorities, the Israeli queue with retrial, and the Israeli queue with a general group-joining policy. By using the probability generating function and matrix analytic method, the authors provided an extensive probabilistic analysis and derived some key performance measures. Recently, Jiang et al. [15, 16], respectively, investigated the multiserver Israeli queue with dynamic service control and the Israeli queue operating in a multiphase random environment on the basis of Perel and Yechiali's studies. Moreover, in [13], if each retrial customer in the orbit independently repeatedly tries for receiving service rather than only the first customer in the orbit tries for receiving service, an explicit closed-form solution for the steady-state probability distribution may be difficult to derived. Therefore, Jiang [17] again analysed the tail asymptotics property of the Israeli queue with retrials and nonpersistent customers on the basis of Perel and Yechiali's studies [13].

Our model also belongs to the class of queues with strategic customers. The study of queueing systems under a game-theoretic perspective was first introduced by Naor [18] who studied the strategic behavior of customers in M/M/1 queue, where arriving customers know the number of customers in the system and decide whether to join or balk based on their surplus utilities. Edelson and Hildebrand [19] considered the strategic behavior of customers in M/M/1 queue by studying the corresponding unobservable case. Since then, there has been considerable attention paid to the strategic behavior of customers in queueing systems; interested readers may also refer to the books of Hassin and Haviv [20, 21], Hassin [22], recent papers Hassin and Roet-Green [23], Ibrahim [24], and the reference therein. The study of strategic customers in batch service queueing systems is a recent endeavor. For example, Economou and Manou [2] investigated an M/M/1 clearing system in an alternating environment, where all customers are served together at the completion epoch of a service cycle. Manou et al. [3] considered the strategic behavior of customers in a transportation station, where all present passengers can be removed together, once a vehicle reaches the station. Bountali and Economou [4, 5] investigated the strategic behavior of customers in queueing systems with fixed size batch services under the unobservable, observable, and partially observable cases. In [4, 5], the authors discussed the impact of different levels of information and the batch size on the joining strategies of customers and compared the joining strategies of customers under the incomplete information level with the corresponding strategies in the complete information level. Recently, Bountali and Economou [25] studied the strategic customer behavior in a two-stage service system with batch processing. Wang et al. [26] studied passengers' joining strategies in a batch transfer queueing system with gated policy under the full observable case. Adan et al. [27] considered a polling system with two queues and studied the optimal routing problem of customers under different levels of information. Chai et al. [28] studied a batch double-sided service system with impatient servers and boundedly rational customers, where each server could serve the customers with a fixed service capacity N.

The queueing model in consideration is also closely related to queueing models with customer service priorities. In queueing systems with customer service priorities, high-priority customers could skip over low-priority customers; i.e., the high-priority customers are served prior to the low-priority customers, which leads to a longer waiting time for the low-priority customers who may be already in the queue. Therefore, queueing models with customer service priorities can be also regarded as a special queueing system with customer interjection. However, different from the other queueing systems with customer interjection, providing priority access is a widespread method for improving the work efficiency; meanwhile, designing an appropriate priority queueing system is also an efficient method for managing congestion. For more detailed studies on priority queues, readers are referred to Jouini and Roubos [7], Takagi [8], Drekic and Woolford [29], Afèche and Mendelson [30], Gavirneni and Kulkarni [31], Wang et al. [32], and

references therein. Moreover, the study of priority queues from an economic viewpoint is also a recent endeavor. For example, Drekic and Woolford [29] analysed a two-class, single-server, preemptive priority queueing model, where the decision to balk or not of the two types of customers depends on the queue length. Brouns and van der Wal [33] studied the optimal threshold strategies in a preemptive priority queueing system with admission and termination control. Liu and Berry [34] considered a preemptive priority queueing model, where the authors focus their analysis on the price competition and social welfare in an unobservable M/G/1 queue with two types of customers. Later, Xu et al. [35, 36], respectively, obtained the optimal balking strategies of high-priority and low-priority customers in an M/G/1 queue with preemptive priorities. In addition, in the literature of operations management, some authors considered the queueing systems with a pay-for-priority option. For example, Hassin and Haviv [20] considered the equilibrium threshold policies with preemptive priority customers. Afèche and Mendelson [30] investigated the pricing and priority auctions in queueing systems with a generalized delay cost structure. Gavirneni and Kulkarni [31] studied a self-selecting priority queue, where customers have heterogeneous waiting costs , and they decide whether to become priority customers or regular customers or balking customers on the basis of their different waiting costs levels. Wang et al. [32] studied the equilibrium strategies in an M/M/1 priority queue, in which the authors consider the system with a pay-for-priority option, and derive the customers' joint decisions between joining/balking and pay-for-priority. For more details on this topic, the interested readers are referred to Adiri and Yechiali [37], Hassin and Haviv [21], and Hassin [22], in which the authors give comprehensive reviews on the topic of pay-for-priority option.

3. Model Description

We consider a two-class, single-server, preemptive priority queueing model in which the high-priority customers form a classical M/M/1 priority queue, while the low-priority customers form the so-called Israeli queue with at most N different groups. We use "he" to represent a high-priority customer and "she" to represent a low-priority customer for ease of exposition. We also assume that high (low-)priority customers arrive at the system according to a Poisson process with rate $\lambda_1 (\lambda_2)$. Service time of high-priority customers is exponentially distributed with rate μ_1. For the low-priority customers, customers are served in an unlimited-size batch mode; i.e., the low-priority customers in a group are served together at a service period, and the service time of a batch independent of its size is assumed to be exponentially distributed with parameter μ_2. A new arriving low-priority customer who determines to join the system will first search for the leader of each group (the searching sequence follows the age-based discipline; i.e., an arriving low-priority customer first searches for the leader of each group who has waited for the longest time), and if she knows a group leader, she immediately joins the group and receives service in a batch mode along with the group leader. We further assume that a new arriving low-priority customer knows a group leader with probability θ. That is, if $n, 1 \le n \le N-1$, groups exist in the Israeli queue, then an arriving low-priority customer who decides to enter the Israeli queue will join kth group with probability $(1-\theta)^{k-1}\theta$ for $1 \le k \le n$ or will create a new group and become a group leader with probability $(1-\theta)^n$. If N groups are present in the Israeli queue, an arriving low-priority customer who decides to enter the Israeli queue will join N-th group even if she does not know any of the existing group leaders. We further assume that the high-priority customers have preemptive priority over the low-priority customers, which indicates that the service for low-priority customers may be interrupted immediately by high-priority customers.

For the recent study of Israeli queue with priorities in [1], the authors have obtained the stationary distribution and performance measures; however, in the present paper, we are interested in investigating the strategic behavior of customers in the batch service special polling system with priorities, where arriving customers have the option to determine whether or not to join the system according to their expected utilities. We further assume that they make their decisions based on the information of the system including the arrival rate, the service rate, the probability that a new arriving low-priority customer knows a group leader, and the unit waiting cost. Then, they determine whether or not to join the system according to the natural linear reward-cost structure $U \triangleq R - cE[W]$ (service reward $\triangleq$ service value-waiting cost), where U is the expected utility of a customer, R is the reward on completion of a service, c is the average delay cost per unit of time for each customer, and $E[W]$ is the expected sojourn time of a customer who decides to join the system. Then, an arriving customer decides whether or not to join the system based on the value of the expected utility U. If $U \ge 0$, he/she will join the system; otherwise, he/she will balk.

In the following parts, we will investigate the strategic behavior of high-priority and low-priority customers regarding their joining or balking dilemma by distinguishing the observable and unobservable cases, i.e., the arriving customers decide whether or not to join the system according to the information available at their arrival instants.

4. Analysis of the Observable Case

We first start our analysis by investigating the Israeli queue with priorities under the observable case. In the observable case, the customers could receive exact information about the Israeli queue upon their arrival; i.e., they could get informed about the statistical and economic parameters of the system, the number of different groups, and the number of high-priority customers in the system. In order to obtain the individual optimal pure strategy, we follow the example of Perel and Yechiali [1] to calculate the expected sojourn time of customers. According to [1], we have the following conclusions:

(1) Since the high-priority customers form a classical M/M/1 queue, the expected sojourn time of an arbitrary high-priority customer is $1/(\mu_1 - \lambda_1)$.

(2) Define B_m as the time of duration from the first moment when there are m high-priority customers in the system until the first moment that no high-priority customers are present. We have the expected value $E[B_m] = m/\mu_1 - \lambda_1$.

(3) Suppose that there are no high-priority customers in the system and there are $n \geq 1$ low-priority customers groups. Let C_n be the time of duration that these n groups are served and leave the system; then the expected value $E[C_n] = n\mu_1/\mu_2(\mu_1 - \lambda_1)$.

(4) Assume that there are $m \geq 0$ high-priority customers and $n \geq 1$ low-priority customers groups in the system and let $D_{m,n}$ denote the time of duration that the service of all those n low-priority customers groups is completed. We have the expected value $E[D_{m,n}] = E[B_m] + E[C_n] = 1/\mu_1 - \lambda_1(m + n\mu_1/\mu_2)$.

4.1. The Individual Optimal Pure Strategy of High-Priority Customers. Next, based on the above results, we first determine the unique individual optimal pure strategy of high-priority customers. Under the observable case, since the high-priority customers form a classical $M/M/1/n+1$ queue ($n+1$ is the threshold), it is easy to obtain the unique individual optimal pure strategy. For an arriving high-priority customer, when he finds n high-priority customers present in the system, he joins if $U_1 = R_1 - c_1 E[W_n] \geq 0$; otherwise he balks, where $E[W_n]$ represents the expected sojourn time of an arriving high-priority customer who finds n high-priority customers and decides to enter the system.

Lemma 1 (see [18]). *In the observable Israeli queue with priorities, there exists a unique individual optimal pure strategy of high-priority customers which has the following forms:*

Case 1. If $R_1 < c_1/\mu_1$, then an arriving high-priority customer always balks.

Case 2. If $R_1 \geq c_1/\mu_1$, the unique individual optimal strategy is the threshold strategy having the following form: when an arriving high-priority customer observes that there are n high-priority customers present in the system, he joins if $n \leq n_e$; otherwise, he balks, where $n_e (= \lfloor n^* \rfloor)$ with $n^* = R_1\mu_1/c_1 - 1$.

4.2. The Individual Optimal Pure Strategy of Low-Priority Customers. Let $L_1(t)$ and $L_2(t)$, respectively, denote the number of high-priority and low-priority customers in the system at time t. Then, $X(t) = \{(L_1(t), L_2(t)), t \geq 0\}$ is a continuous-time Markov chain with state space $\Omega = \{(m,n),\ m \geq 0, n = 0, 1, \ldots, N\}$. For the low-priority customers, when an arriving low-priority customer finds the system being in state (m,n), she joins the system if $U_2(m,n) = R_2 - c_2 E[S_{m,n}] \geq 0$; otherwise she balks, where $E[S_{m,n}]$ represents the expected sojourn time of an arriving low-priority customer who finds the system being in state (m,n). Next, we determine the unique individual optimal pure strategy of low-priority customers. We first give the following theorem to show the individual optimal pure strategy of low-priority customers under the observable case.

Theorem 1. *In the observable Israeli queue with priorities, there exists a unique individual optimal pure strategy of low-priority customers which has the following forms:*

$$q^e_{m,j} = \begin{cases} 1, m \leq m^j_e, \\ 0, m > m^j_e, \end{cases} \quad 0 \leq j \leq N, \tag{1}$$

where

$$\begin{aligned} m^0_e &= \lfloor m^*_0 \rfloor, m^*_0 = \frac{R_2(\mu_1 - \lambda_1)}{c_2} - \frac{\mu_1}{\mu_2}, \\ m^k_e &= \lfloor m^*_k \rfloor, m^*_k = \frac{R_2(\mu_1 - \lambda_1)}{c_2} - \frac{\mu_1}{\mu_2}\frac{1-(1-\theta)^{k+1}}{\theta}, \\ &\qquad 1 \leq k \leq N-1, \\ m^N_e &= \lfloor m^*_N \rfloor, m^*_N = \frac{R_2(\mu_1 - \lambda_1)}{c_2} - \frac{\mu_1}{\mu_2}\frac{1-(1-\theta)^{N}}{\theta}, \\ &\text{with } m^N_e = m^{N-1}_e \leq m^{N-2}_e \leq \cdots \leq m^0_e. \end{aligned} \tag{2}$$

Proof. We first derive the expected sojourn time of an arriving low-priority customer who finds the system being in state (m,n). Based on the detailed list of results presented in Section 4 (obtained by Perel and Yechiali in [1]), the expected sojourn time of an arriving low-priority customer has the following forms:

$$\begin{aligned} E[S_{m,0}] &= E[D_{m,1}] = \frac{1}{\mu_1 - \lambda_1}\left(m + \frac{\mu_1}{\mu_2}\right), \\ E[S_{m,n}] &= \sum_{k=0}^{n-1}(1-\theta)^k \theta E[D_{m,k+1}] + (1-\theta)^n E[D_{m,n+1}], \\ &\qquad 1 \leq n \leq N-1, \\ E[S_{m,N}] &= \sum_{k=0}^{N-2}(1-\theta)^k \theta E[D_{m,k+1}] + (1-\theta)^{N-1} E[D_{m,N}]. \end{aligned} \tag{3}$$

After some calculations, we have

$$E\left[S_{m,n}\right] = \frac{1}{\mu_1 - \lambda_1}\left(m + \frac{\mu_1}{\mu_2}\frac{1-(1-\theta)^{n+1}}{\theta}\right), \quad 0 \le n \le N-1, \tag{4}$$

$$E\left[S_{m,N}\right] = \frac{1}{\mu_1 - \lambda_1}\left(m + \frac{\mu_1}{\mu_2}\frac{1-(1-\theta)^{N}}{\theta}\right). \tag{5}$$

Differentiating $U_2(m,n)$ with respect to m and n, respectively, we then have $\partial U_2(m,n)/\partial m = -c_2/\mu_1 - \lambda_1 < 0$ and $\partial U_2(m,n)/\partial n = \mu_1 c_2/\mu_2\theta(1-\theta)^{n+1}\ln(1-\theta) < 0$; that is, $U_2(m,n)$ is a decreasing function of m and n. Hence, an arriving low-priority customer balks if $U_2(m,n) < 0$; otherwise, she prefers to join the Israeli queue. In order to obtain the unique individual optimal pure strategy of low-priority customers, we consider the following cases. □

Case 1. The arriving low-priority customer finds no groups in the Israeli queue.

(1) If $U_2(0,0) < 0$, i.e., $R_2 < \mu_1 c_2/\mu_2(\mu_1 - \lambda_1)$, then, an arriving low-priority customer always balks.

(2) If $U_2(0,0) \ge 0$, i.e., $R_2 \ge \mu_1 c_2/\mu_2(\mu_1 - \lambda_1)$, then, there exists a unique root m_0^*, such that $U_2(m_0^*, 0) = 0$. Solving this equation, we have $m_0^* = R_2(\mu_1 - \lambda_1)/c_2 - \mu_1/\mu_2$, since $U_2(m,0)$ is a decreasing function of m; an arriving low-priority customer determines to join the Israeli queue if and only if $m \le m_e^0 (= \lfloor m_0^* \rfloor)$.

Case 2. The arriving low-priority customer finds $k, 1 \le k \le N-1$, groups in the Israeli queue.

(1) If $U_2(0,k) < 0$, i.e., $R_2 < \mu_1 c_2(1-(1-\theta)^{k+1})/\mu_2\theta(\mu_1 - \lambda_1)$, then, an arriving low-priority customer always balks.

(2) If $U_2(0,k) \ge 0$, i.e., $R_2 \ge \mu_1 c_2(1-(1-\theta)^{k+1})/\mu_2\theta(\mu_1 - \lambda_1)$, then, there exists a unique root m_k^*, such that $U_2(m_k^*, k) = 0$. Solving this equation, we have $m_k^* = R_2(\mu_1 - \lambda_1)/c_2 - \mu_1/\mu_2 1-(1-\theta)^{k+1}/\theta$, since $U_2(m,k)$ is a decreasing function about m; an arriving low-priority customer determines to join the Israeli queue if and only if $m \le m_e^k (= \lfloor m_k^* \rfloor)$.

Case 3. The arriving low-priority customer finds N groups in the Israeli queue.

(1) If $U_2(0,N) < 0$, i.e., $R_2 < \mu_1 c_2(1-(1-\theta)^{N})/\mu_2\theta(\mu_1 - \lambda_1)$, then, an arriving low-priority customer always balks.

(2) If $U_2(0,N) \ge 0$, i.e., $R \ge \mu_1 c_2(1-(1-\theta)^{N})/\mu_2\theta(\mu_1 - \lambda_1)$, then, there exists a unique root m_N^*, such that $U_2(m_N^*, N) = 0$. Solving this equation, we have $m_N^* = R_2(\mu_1 - \lambda_1)/c_2 - \mu_1/\mu_2 1-(1-\theta)^{N}/\theta$, since $U_2(m,N)$ is a decreasing function of m; an arriving low-priority customer determines to join the Israeli queue if and only if $m \le m_e^N [= \lfloor m_N^* \rfloor]$.

According to the above analysis, we find that $m_e^N = m_e^{N-1} \le m_e^{N-2} \le \cdots \le m_e^0$.

Remark 1

(1) If the individual optimal threshold strategy of high-priority customers $n_e + 1$ is greater than the individual optimal threshold strategy of low-priority customers m_e^0, i.e., $n_e + 1 > m_e^0$, then, the individual optimal threshold strategy of low-priority customers is the result obtained in Theorem 1.

(2) If there exists k, such that $m_e^k \le n_e + 1 \le m_e^{k-1}$, then the result in Theorem 1 should be

$$q_{m,j}^e = \begin{cases} 1, m \le m_e^j, \\ 0, m > m_e^j, \end{cases} \quad k \le j \le N,$$
$$q_{m,j}^e = \begin{cases} 1, m \le n_e + 1, \\ 0, m > n_e + 1, \end{cases} \quad 0 \le j \le k-1. \tag{6}$$

(3) If $n_e + 1 \le m_e^N$, then the result in Theorem 1 should be

$$q_{m,j}^e = \begin{cases} 1, m \le n_e + 1, \\ 0, m > n_e + 1, \end{cases} \quad 0 \le j \le N. \tag{7}$$

Since $n_e + 1$ is the unique individual optimal pure strategy of high-priority customers, i.e., the number of high-priority customers is no more than $n_e + 1$, we could obtain the results of (6) and (7).

4.3. The Stationary Distribution under the Observable Case. In order to solve the optimization problem of subsequent equilibrium rewards, deriving the steady-state probabilities on the basis of the threshold strategies is quite crucial. According to the results in Lemma 1 and Theorem 1, the equilibrium joining strategies of both types of customers become threshold policies in the observable case. Each customer (either high-priority or low-priority customer) could determine the individual optimal strategy once the state of the system is observable. Meanwhile, it is not difficult to find that when the high-priority customers always balk ($R_1 < c_1/\mu_1$), the model degenerates into a one-dimensional finite Markov chain, and then the underlying queueing model reduces to an $M/M/1/n$ queue with only low-priority customers under a threshold policy; otherwise $\{(L_1(t), L_2(t)), t \ge 0\}$ forms a nonhomogeneous finite Quasi-Birth-and-Death (QBD) process. Considering the variety of situations caused by the threshold policies of high-priority and low-priority customers, taking the situations $m_e^N \ge 0$ and $n_e > m_e^0$, for example, then the transition rate diagram can be shown in Figure 1.

Next, we mainly focus on the general case with the assumption that the high-priority customers do not balk. According to the transition rate diagram given in Figure 1, we refer to Latouche and Ramaswami in [38] to solve the nonhomogeneous finite QBD with level process $L_2(t)$ and phase process $L_1(t)$. The generator Q_0 of the underlying queueing system has the following forms:

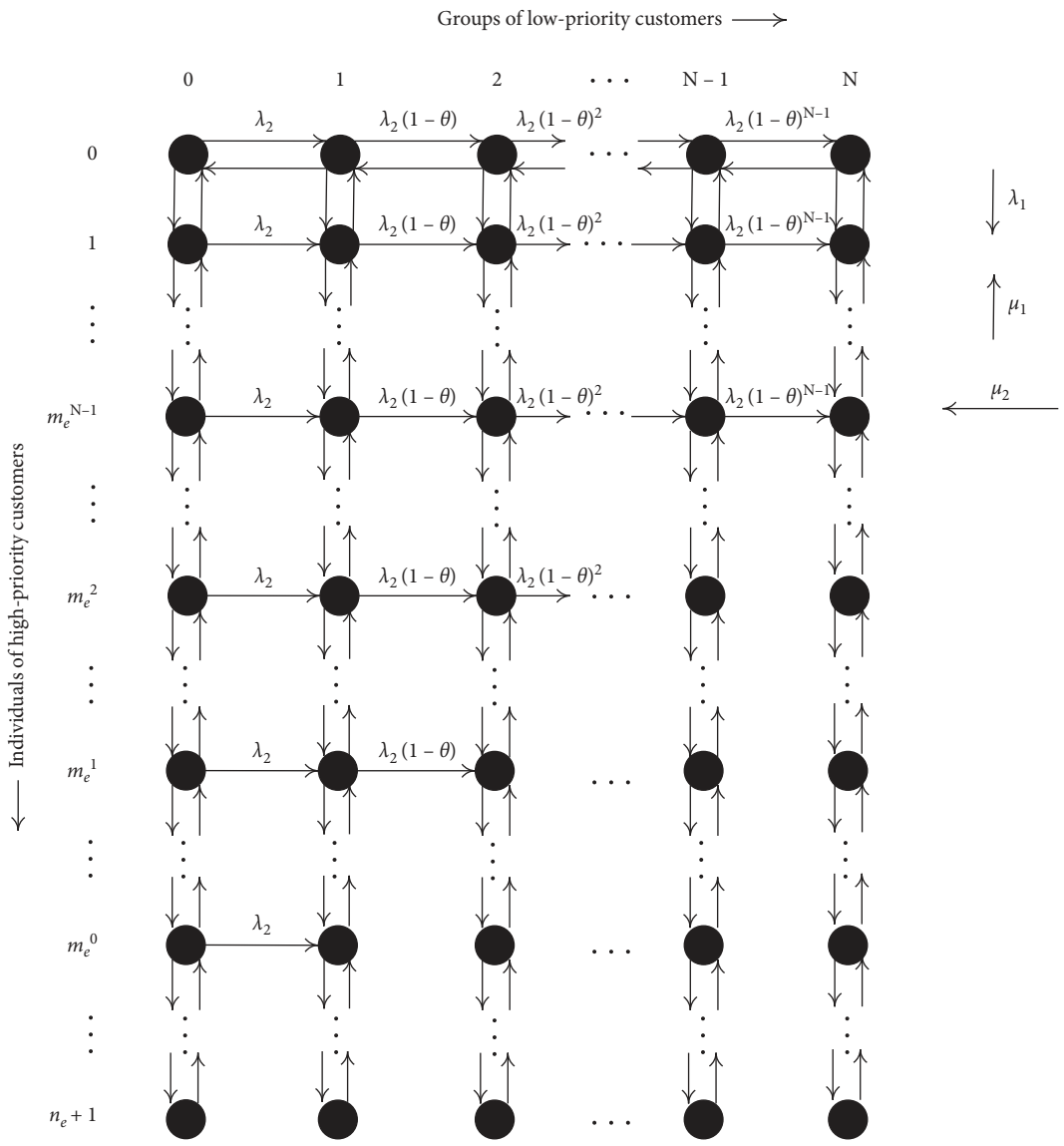

Figure 1: Transition rate diagram of the observable system.

$$Q_0 = \begin{pmatrix} D_0 & B_0 & & & & \\ C_1 & D_1 & B_1 & & & \\ & C_2 & D_2 & B_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & C_{N-1} & D_{N-1} & B_{N-1} \\ & & & & C_N & D_N \end{pmatrix}, \tag{8}$$

where the order of each submatrix is $n_e + 1$ by $n_e + 1$ and all of the block matrices can be easily obtained while the necessary parameters are given. Next, the following theorem indicates how to determine the steady-state probability vector.

Theorem 2. *The stationary probability vector p can be given by*

$$p_j = p_N \Phi_{j+1}, \quad 0 \le j \le N-1, \tag{9}$$

where p_N satisfies

$$\begin{aligned} &p_N \left(\sum_{n=1}^{N} \prod_{i=N}^{n} \phi_i + I \right) e_{n_e+1} = 1, \\ &p_N \phi_N B_{N-1} + p_N D_N = 0, \\ &\Phi_k = \prod_{i=N}^{k} \phi_i, \quad 1 \le k \le N, \\ &\phi_1 = C_1 (-D_0)^{-1}, \\ &\phi_i = C_i \left(-\phi_{i-1} B_{i-2} - D_{i-1}\right)^{-1}, \quad 2 \le i \le N. \end{aligned} \tag{10}$$

Proof. This theorem can be proved by expanding $(p_0, p_1, \ldots, p_N)Q_0 = 0$. We may first obtain p_j in terms of p_N by several iterations. Then, by normalizing the vector p, we could determine p_N. The detailed process is not shown here.

The detailed recursive algorithm for calculating the steady-state probability vectors can be summarized in Algorithm 1. □

5. Analysis of the Unobservable Case

In this section, we will turn to the analysis of the unobservable case. In the unobservable case, the customers could not receive exact information about the Israeli queue except the statistical and economic parameters of the system upon their arrival; i.e., they do not receive any information about the number of different groups and the number of high-priority customers in the system. Before analysing the strategic behavior of customers under this case, we first derive the steady-state probabilities of the system. We assume that an arriving high-priority customer joins the system with probability q_h and an arriving low-priority customer joins the system with probability q_l. Since $\{(L_1(t), L_2(t)), t \geq 0\}$ is a continuous-time Markov chain with state space $\Omega = \{(m, n), m \geq 0, 0 \leq n \leq N\}$, then, by referring to the continuous-time Markov process, we can obtain the state-transition-rate matrix as follows:

$$Q = \begin{pmatrix} C & A_0 & 0 & 0 & \ldots \\ A_2 & A_1 & A_0 & 0 & \ldots \\ 0 & A_2 & A_1 & A_0 & \ldots \\ 0 & 0 & A_2 & A_1 & \ldots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}, \tag{11}$$

where

$$C = \begin{pmatrix} -(\lambda_1 q_h + \lambda_2 q_l) & \lambda_2 q_l & 0 & 0 & \ldots & 0 & 0 \\ \mu_2 & a_1 & \lambda_2 q_l (1-\theta) & 0 & \ldots & 0 & 0 \\ 0 & \mu_2 & a_2 & \lambda_2 q_l (1-\theta)^2 & \ldots & 0 & 0 \\ 0 & 0 & \mu_2 & a_3 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & a_{N-1} & \lambda_2 q_l (1-\theta)^{N-1} \\ 0 & 0 & 0 & 0 & \ldots & \mu_2 & -(\lambda_1 q_h + \mu_2) \end{pmatrix}, \tag{12}$$

with $a_k = -(\lambda_1 q_h + \lambda_2 q_l (1-\theta)^k + \mu_2)$, $A_0 = \lambda_1 q_h I$, $A_2 = \mu_1 I$, and

$$A_1 = \begin{pmatrix} -(\mu_1 + \lambda_1 q_h + \lambda_2 q_l) & \lambda_2 q_l & 0 & 0 & \ldots & 0 & 0 \\ 0 & b_1 & \lambda_2 q_l (1-\theta) & 0 & \ldots & 0 & 0 \\ 0 & 0 & b_2 & \lambda_2 q_l (1-\theta)^2 & \ldots & 0 & 0 \\ 0 & 0 & 0 & b_3 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & b_{N-1} & \lambda_2 q_l (1-\theta)^{N-1} \\ 0 & 0 & 0 & 0 & \ldots & 0 & -(\lambda_1 q_h + \mu_1) \end{pmatrix}, \tag{13}$$

with $b_k = -(\lambda_1 q_h + \lambda_2 q_l (1-\theta)^k + \mu_1)$, and I is the identity matrix of order $N+1$.

In order to analyse the system effectively by matrix analytic method, an important matrix in analysing the process is the rate matrix R, which is the minimal non-negative solution of

$$R^2 A_2 + R A_1 + A_0 = 0. \tag{14}$$

According to the special structure of matrices A_0, A_1, A_2, we know that R is an upper triangular matrix and has a closed-form expression. The explicit calculation of R can be seen in [1]. We denote the steady-state probability vector of the queueing model $\pi = (\pi_0, \pi_1, \ldots)$, where $\pi_n = (\pi_{n,0}, \pi_{n,1}, \ldots, \pi_{n,N})$, $\pi_{n,i} = \lim_{t \longrightarrow \infty} P(L_1(t) = n, L_2(t) = i)$, $n \geq 0, 0 \leq i \leq N$, then, the steady-state probability vector π_0 can be obtained by solving the following linear equations:

Input: $\{\lambda_1, \lambda_2, \mu_1, \mu_2, \theta, N\}$
Output: $\{\pi_1, \pi_2, \ldots, \pi_N\}$
Step 1: set $\phi_1 = C_1(-D_0)^{-1}$
Step 2: for i from 2 to N, set $\phi_i = C_i(-\phi_{i-1}B_{i-2} - D_{i-1})^{-1}$
Step 3: for j from 1 to N, set $\Phi_j = \prod_{i=N}^{j}\phi_i$
Step 4: solve $\pi_N\phi_N B_{N-1} + \pi_N D_N = 0$ and $\pi_N\left[\sum_{n=1}^{N}\prod_{i=N}^{n}\phi_i + I\right]e_{n_e+1} = 1$
Step 5: compute the steady-state probability vectors, $\pi_j = \pi_N\Phi_{j+1}$ for $0 \le j \le N-1$
Step 6: output
where e_{n_e+1} in the above algorithm is a column vector of 1's with the order $n_e + 1$ by 1.

Algorithm 1: A recursive algorithm to calculate the steady-state probability vectors.

$$\begin{aligned}\pi_0(C + RA_2) &= 0,\\ \pi_0(I - R)^{-1}e &= 1.\end{aligned} \tag{15}$$

and the steady-state probability vector $\pi_m, m \ge 1$, can be obtained by $\pi_m = \pi_0 R^m$.

5.1. Equilibrium Strategy of High-Priority Customers. Next, according to the obtained steady-state probability vectors, we investigate the strategic behavior of customers under the unobservable case. It is readily seen that the system is stable if and only if $\lambda_1 q_h < \mu_1$. However, even if $\lambda_1 \ge \mu_1$, the system may be stable. So, we will analyse this problem in two cases: $\lambda_1 < \mu_1$ and $\lambda_1 \ge \mu_1$. We first derive the joining strategies of high-priority customers by the following lemma.

Lemma 2. *In the unobservable Israeli queue with priorities under the condition $\lambda_1 < \mu_1$, there exists a unique equilibrium mixed strategy q_h^e of high-priority customers, which has the following forms:*

$$q_h^e = \begin{cases} 0, R_1 \le \dfrac{c_1}{\mu_1}, \\ \dfrac{\mu_1 - c_1/R_1}{\lambda_1}, \dfrac{c_1}{\mu_1} < R_1 < \dfrac{c_1}{\mu_1 - \lambda_1}, \\ 1, R_1 \ge \dfrac{c_1}{\mu_1 - \lambda_1}. \end{cases} \tag{16}$$

Remark 2. For the case $\lambda_1 \ge \mu_1$, the analog of Lemma 2 is that a unique equilibrium mixed strategy q_h^e of high-priority customers exists, which has the following form:

$$q_h^e = \begin{cases} 0, R_1 \le \dfrac{c_1}{\mu_1}, \\ \dfrac{\mu_1 - c_1/R_1}{\lambda_1}, R_1 > \dfrac{c_1}{\mu_1}. \end{cases} \tag{17}$$

5.2. Equilibrium Strategy of Low-Priority Customers. Next, we aim to obtain the joining strategies of low-priority customers. For the low-priority customers, the expected utility of a low-priority customer while all the other low-priority customers follow the strategy q_l is

$$U_{unl}(q_h, q_l) = R_2 - c_2 E[W_l] = R_2 - c_2 \sum_{m=0}^{\infty}\sum_{n=0}^{N}\pi_{m,n}E\left[S_{m,n}\right], \tag{18}$$

where $E[W_l]$ represents the expected sojourn time of a low-priority customer who decides to join the system, which is a function with regard to q_h and q_l. According to the equilibrium strategies of high-priority customers, we then derive the equilibrium strategies of low-priority customers. Next, we first assume $\lambda_1 < \mu_1$ and consider the following three scenarios to investigate the equilibrium strategies of low-priority customers.

(1) Scenario 1: all of the high-priority customers balk, i.e., $q_h^e = 0$.
(2) Scenario 2: a fraction of the high-priority customers joins the priority queue, i.e., $q_h^e = \mu_1 - c_1/R_1/\lambda_1$.
(3) Scenario 3: all of the high-priority customers join, i.e., $q_h^e = 1$.

Scenario 1. All of the high-priority customers balk, i.e., $q_h^e = 0$. Under this scenario, the system reduces to an M/M/1/N type Israeli queue without priorities; we then have the following theorem.

Theorem 3. *When all of the high-priority customers balk $q_h^e = 0$, $R_1 \le c_1/\mu_1$, then, there exists a unique equilibrium mixed strategy q_l^e of low-priority customers, which has the following forms:*

$$q_l^e = \begin{cases} 0, R_2 \le \dfrac{c_2}{\mu_2}, \\ q_l^{e^*}, \dfrac{c_2}{\mu_2} < R_2 < \dfrac{c_2}{\mu_2\theta} - \left(\dfrac{c_2(1-\theta)(1-\pi_0(1))}{\lambda_2\theta} + \pi_N(1)(1-\theta)^N\right), \\ 1, R_2 \ge \dfrac{c_2}{\mu_2\theta} - \left(\dfrac{c_2(1-\theta)(1-\pi_0(1))}{\lambda_2\theta} + \pi_N(1)(1-\theta)^N\right), \end{cases} \tag{19}$$

where

$$\pi_N(1) = \pi_0\left(\frac{\lambda_2}{\mu_2}\right)^n (1-\theta)^{n(n-1)/2}, \quad 1 \le n \le N,$$
$$\pi_0(1) = \left(\sum_{n=0}^{N}\left(\frac{\lambda_2}{\mu_2}\right)^n (1-\theta)^{n(n-1)/2}\right)^{-1}, \tag{20}$$

and $q_l^{e^*} \in (0,1)$ *is the unique solution of the following equation:*

$$R_2 - \frac{c_2}{\mu_2\theta} + \frac{c_2}{\mu_2\theta}\left(\frac{\mu_2(1-\theta)}{\lambda_2 q_l}\left(1-\pi_0(q_l)\right) + \pi_N(q_l)(1-\theta)^N\right) = 0. \tag{21}$$

Proof. When all of the high-priority customers balk, i.e., $q_h^e = 0$, the system reduces to an M/M/1/N type Israeli queue without priorities.

Under this scenario, the stationary distribution $\pi_n = \lim_{t\longrightarrow\infty} P(L_2(t) = n)$ is easily given by (we can also refer to Perel and Yechiali in [12])

$$\pi_n = \pi_0\left(\frac{\lambda_2 q_l}{\mu_2}\right)^n (1-\theta)^{n(n-1)/2}, \quad 1 \le n \le N,$$
$$\pi_0 = \left(\sum_{n=0}^{N}\left(\frac{\lambda_2 q_l}{\mu_2}\right)^n (1-\theta)^{n(n-1)/2}\right)^{-1}. \tag{22}$$

Then, we have the following expression of $U_{unl}(0, q_l)$:

$$U_{unl}(0,q_l) = R_2 - c_2 E[W_l] = R_2 - c_2 \sum_{n=0}^{N} \pi_n E[W_n], \tag{23}$$

where

$$E[W_0] = \frac{1}{\mu_2},$$
$$E[W_n] = \frac{\theta}{\mu_2} + \cdots + \frac{n\theta(1-\theta)^{n-1}}{\mu_2} + \frac{(n+1)(1-\theta)^n}{\mu_2} = \frac{1-(1-\theta)^{n+1}}{\mu_2\theta}, \quad 1 \le n \le N-1, \tag{24}$$
$$E[W_N] = \frac{\theta}{\mu_2} + \cdots + \frac{(N-1)\theta(1-\theta)^{N-2}}{\mu_2} + \frac{(N)(1-\theta)^{N-1}}{\mu_2} = \frac{1-(1-\theta)^N}{\mu_2\theta}.$$

Next, we will derive the variation trend of the function $U_{unl}(0, q_l)$ on q_l. Define $a_n = (\lambda_2/\mu_2)^n (1-\theta)^{n(n-1)/2}$; then the expression of $U_{unl}(0, q_l)$ can be rewritten as

$$U_{unl}(0,q_l) = R_2 - c_2 \frac{\sum_{n=0}^{N} a_n E(W_n) q_l^n}{\sum_{n=0}^{N} a_n q_l^n}. \tag{25}$$

Differentiating $U_{unl}(0, q_l)$ with respect to q_l, we have

$$\frac{\partial U_{unl}(0,q_l)}{\partial q_l} = -c_2 \frac{\left(\sum_{n=0}^{N} n a_n E(W_n) q_l^{n-1}\right)\left(\sum_{n=0}^{N} a_n q_l^n\right) - \left(\sum_{n=0}^{N} a_n E(W_n) q_l^n\right)\left(\sum_{n=0}^{N} n a_n q_l^{n-1}\right)}{\left(\sum_{n=0}^{N} a_n q_l^n\right)^2}. \tag{26}$$

We next focus on the analysis of

$$A(q_l) = \left(\sum_{n=0}^{N} n a_n E(W_n) q_l^n\right)\left(\sum_{n=0}^{N} a_n q_l^n\right) - \left(\sum_{n=0}^{N} a_n E(W_n) q_l^n\right)\left(\sum_{n=0}^{N} n a_n q_l^n\right), \tag{27}$$

whose sign is exactly contrary to (26) (see Wang and Sun in [39] and Wang et al. in [40]). Expanding the terms

$$\left(\sum_{n=0}^{N} na_n E(W_n) q_l^n\right)\left(\sum_{n=0}^{N} a_n q_l^n\right) \text{ and } \left(\sum_{n=0}^{N} a_n E(W_n) q_l^n\right)\left(\sum_{n=0}^{N} na_n q_l^n\right), \tag{28}$$

into polynomial functions of q_l respectively, we have

$$\begin{aligned}\left(\sum_{n=0}^{N} na_n E(W_n) q_l^n\right)\left(\sum_{n=0}^{N} a_n q_l^n\right) &= \sum_{k=0}^{2N} \sum_{i=\max(0,k-N)}^{\min(k,N)} s_{i,k} q_l^k,\\ \left(\sum_{n=0}^{N} a_n E(W_n) q_l^n\right)\left(\sum_{n=0}^{N} na_n q_l^n\right) &= \sum_{k=0}^{2N} \sum_{i=\max(0,k-N)}^{\min(k,N)} d_{i,k} q_l^k,\end{aligned} \tag{29}$$

where $s_{i,k} = ia_i a_{k-i} E(W_i)$ and $d_{i,k} = (k-i)a_i a_{k-i} E(W_i)$.

Fixing the value of k, for any feasible i, it is not difficult to find that

$$\left(s_{i,k} + s_{k-i,k}\right) - \left(d_{i,k} + d_{k-i,k}\right) = a_i a_{k-i}\left(i - (k-i)\right)\left(E(W_i) - E(W_{k-i})\right) \geq 0, \tag{30}$$

which leads to $A(q_l) \geq 0$; thus we have $\partial U_{unl}(0, q_l)/\partial q_l \leq 0$. Then, the result shows that $U_{unl}(0, q_l)$ is a decreasing function of q_l. Next, we consider the following three cases to obtain the equilibrium strategies of low-priority customers under this scenario. □

Case 1. If $U_{unl}(0,0) \leq 0$, i.e., $R_2 - c_2/\mu_2\theta + c_2(1-\theta)/\mu_2\theta \leq 0$, we have $R_2 \leq c_2/\mu_2$; then $U_{unl}(0, q_l)$ is nonpositive for every q_l; the best response of an arriving low-priority customer is balking and the unique equilibrium point is $q_l^e = 0$.

Case 2. If $U_{unl}(0,0) > 0$ and $U_{unl}(0,1) < 0$, i.e.,

$$\frac{c_2}{\mu_2} < R_2 < \frac{c_2}{\mu_2\theta} - \left(\frac{c_2(1-\theta)\left(1-\pi_0(1)\right)}{\lambda_2\theta} + \pi_N(1)(1-\theta)^N\right), \tag{31}$$

then, there exists a unique solution $q_l^{e^*}$ that lies in $(0,1)$, such that $U_{unl}(0, q_l^{e^*}) = 0$. By solving the following equation:

$$R_2 - \frac{c_2}{\mu_2\theta} + \frac{c_2}{\mu_2\theta}\left(\frac{\mu_2(1-\theta)}{\lambda_2 q_l}\left(1-\pi_0\right) + \pi_N(1-\theta)^N\right) = 0, \tag{32}$$

we could obtain the unique solution q_l^{e*}.

Case 3. If $U_{unl}(0,1) \geq 0$, i.e., $R_2 \geq c_2/\mu_2\theta - (c_2(1-\theta)(1-\pi_0(1))/\lambda_2\theta + \pi_N(1)(1-\theta)^N)$, then $U_{unl}(0, q_l)$ is nonnegative for every q_l and the best response is 1. So, joining the Israeli queue is the unique equilibrium strategy.

Based on the above analysis, we can derive the result of this theorem.

Next, we consider the other two scenarios that high-priority customers join the priority queue; i.e., $q_h^e = \mu_1 - c_1/R_1/\lambda_1$ and $q_h^e = 1$. Under the two scenarios, we first give the expression of $U_{unl}(q_h^e, q_l)$ by referring to the results in Perel and Yechiali in [1]. From [1], we have

$$\sum_{k=0}^{N} \pi_{0,k} = 1 - \frac{\lambda_1 q_h^e}{\mu_1}. \tag{33}$$

According to the two results, the expression $U_{unl}(q_h^e, q_l)$ can be written as

$$U_{unl}(q_h^e, q_l) = R_2 - c_2 E[W_l] = R_2 - c_2 \sum_{m=0}^{\infty} \sum_{n=0}^{N} \pi_{m,n} E\left[S_{m,n}\right], \tag{34}$$

where $E[W_l] = \sum_{m=0}^{\infty}\sum_{n=0}^{N}\pi_{m,n}E[S_{m,n}]$ is the function with regard to q_h^e and q_l, which can be rewritten as follows:

$$\begin{aligned}\sum_{m=0}^{\infty}\sum_{n=0}^{N}\pi_{m,n}E\left[S_{m,n}\right] &= \sum_{m=0}^{\infty}\sum_{n=0}^{N-1}\pi_{m,n}E\left[S_{m,n}\right] + \sum_{m=0}^{\infty}\pi_{m,N}E\left[S_{m,N}\right]\\ &= \sum_{m=0}^{\infty}\sum_{n=0}^{N-1}\pi_{m,n}\frac{1}{\mu_1-\lambda_1 q_h^e}\left(m+\frac{\mu_1}{\mu_2}\frac{1-(1-\theta)^{n+1}}{\theta}\right) + \sum_{m=0}^{\infty}\pi_{m,N}\frac{1}{\mu_1-\lambda_1 q_h^e}\left(m+\frac{\mu_1}{\mu_2}\frac{1-(1-\theta)^{N}}{\theta}\right)\\ &= \frac{1}{\mu_1-\lambda_1 q_h^e}\left(\sum_{m=0}^{\infty}m\pi_{m,\cdot} + \frac{\mu_1}{\mu_2\theta}\sum_{n=0}^{N-1}\pi_{\cdot,n}\left[1-(1-\theta)^{n+1}\right] + \frac{\mu_1}{\mu_2\theta}\pi_{\cdot,N}\left[1-(1-\theta)^{N}\right]\right)\\ &= \frac{1}{\mu_1-\lambda_1 q_h^e}\left(\frac{\lambda_1 q_h^e}{\mu_1-\lambda_1 q_h^e} + \frac{\mu_1}{\mu_2\theta}\sum_{n=0}^{N-1}\pi_{\cdot,n}\left[1-(1-\theta)^{n+1}\right] + \frac{\mu_1}{\mu_2\theta}\pi_{\cdot,N}\left[1-(1-\theta)^{N}\right]\right)\\ &= \frac{1}{\mu_1-\lambda_1 q_h^e}\left(\frac{\lambda_1 q_h^e}{\mu_1-\lambda_1 q_h^e} + \frac{\mu_1}{\mu_2\theta} - \frac{\mu_1}{\lambda_2 q_l\theta}(1-\theta)\left(1-\frac{\lambda_1 q_h^e}{\mu_1}-\pi_{0,0}\right) - \frac{\mu_1}{\mu_2\theta}\pi_{\cdot,N}(1-\theta)^{N}\right).\end{aligned} \tag{35}$$

$\pi_{m,\cdot} = \sum_{n=0}^{N}\pi_{m,n}, m \geq 0$, $\pi_{\cdot,n} = \sum_{m=0}^{\infty}\pi_{m,n}$, $0 \leq n \leq N$, and $\pi_{0,0}$ and $\pi_{\cdot,N}$ are the functions with regard to q_l. Due to the complexity of the expression $E[W_l]$, it is difficult to derive the variation trend of the function $E[W_l]$ on q_l as $q_h^e = \mu_1 - c_1/R_1/\lambda_1$ and $q_h^e = 1$. Due to the lack of analytic results on the monotonicity, we may use the numerical results to show the monotonicity of $E[W_l]$ on q_l. To this end, we use a numerical example to show the monotonicity of the function $E[W_l]$ on q_l. We assume that $\lambda_1 = 2$, $\lambda_2 = 5$, $\mu_1 = 3$, $\mu_2 = 4$, $\theta = 0.2$, $N = 10$, $R_2 = 4$, and $c_2 = 1$.

According to Figure 2, we find that while $q_h^e = \mu_1 - c_1/R_1/\lambda_1$ and $q_h^e = 1$, $E[W_l]$ is an increasing function of q_l. Intuitively, higher arrival rate $\lambda_2 q_l$ leads to more low-priority customers staying in the system, which will necessarily increase the sojourn time of low-priority customers. Since $U_{unl}(q_h^e, q_l) = R_2 - c_2 E[W_l]$, we obtain that $U_{unl}(q_h^e, q_l)$ is a decreasing function of q_l. In order to derive the equilibrium mixed strategy of low-priority customers, we next consider the other two scenarios.

Scenario 2. A fraction of the high-priority customers joins the priority queue; i.e., $q_h^e = \mu_1 - c_1/R_1/\lambda_1$. Under this scenario, we have the following theorem.

Theorem 4. *A fraction of the high-priority customers joins the priority queue $q_h^e = \mu_1 - c_1/R_1/\lambda_1$; if $c_1/\mu_1 < R_1 < c_1/\mu_1 - \lambda_1$, then there exists a unique equilibrium mixed strategy q_l^e of low-priority customers, which has the following forms:*

$$q_l^e = \begin{cases} 0, R_2 \leq c_2 E\left[W_l\right](q_h^e, 0), \\ q_l^{e^*}, c_2 E\left[W_l\right](q_h^e, 0) < R_2 < c_2 E\left[W_l\right](q_h^e, 1), \\ 1, R_2 \geq c_2 E\left[W_l\right](q_h^e, 0), \end{cases} \tag{36}$$

where $E[W_l](q_h^e, 0) = \lim_{q_l \longrightarrow 0} E[W_l](q_h^e, q_l)$ and $q_l^{e^*} \in (0,1)$ is the unique solution of the following equation:

$$\begin{aligned} U_{unl}\left(q_h^e, q_l^{e^*}\right) &= R_2 - c_2 E\left[W_l\right]\left(q_h^e, q_l^{e^*}\right) = 0, \\ E\left[W_l\right](q_h^e, q_l) &= \frac{1}{\mu_1-\lambda_1 q_h^e}\left(\frac{\lambda_1 q_h^e}{\mu_1-\lambda_1 q_h^e} + \frac{\mu_1}{\mu_2\theta} - \frac{\mu_1}{\lambda_2 q_l\theta}(1-\theta)\left(1-\frac{\lambda_1 q_h^e}{\mu_1}-\pi_{0,0}(q_l)\right) - \frac{\mu_1}{\mu_2\theta}\pi_{\cdot,N}(q_l)(1-\theta)^{N}\right). \end{aligned} \tag{37}$$

Proof. In this scenario, we consider the following three cases. □

Case 1. $U_{unl}(q_h^e, 0) \leq 0$; i.e., $R_2 - c_2 E[W_l](q_h^e, 0) \leq 0$; then $U_{unl}(q_h^e, q_l)$ is nonpositive for every q_l; the best response of an arriving low-priority customer is balking and the unique equilibrium point is $q_l^e = 0$.

Case 2. $U_{unl}(q_h^e, 0) > 0$ and $U_{unl}(q_h^e, 1) < 0$; i.e., $c_2 E[W_l](q_h^e, 0) < R_2 < c_2 E[W_l](q_h^e, 1)$; then, there exists a unique solution $q_l^{e^*}$ that lies in $(0,1)$, such that $U_{unl}(q_h^e, q_l^{e^*}) = 0$. By solving the following equation:

$$U_{unl}\left(q_h^e, q_l^{e^*}\right) = R_2 - c_2 E\left[W_l\right]\left(q_h^e, q_l^{e^*}\right) = 0, \tag{38}$$

we could obtain the unique solution q_l^{e*}.

Case 3. $U_{unl}(q_h^e, 1) \geq 0$; i.e., $R_2 \geq c_2 E[W_l](q_h^e, 1)$; then $U_{unl}(q_h^e, q_l)$ is nonnegative for every q_l; i.e., the best response is 1. So, joining the Israeli queue is the unique equilibrium strategy.

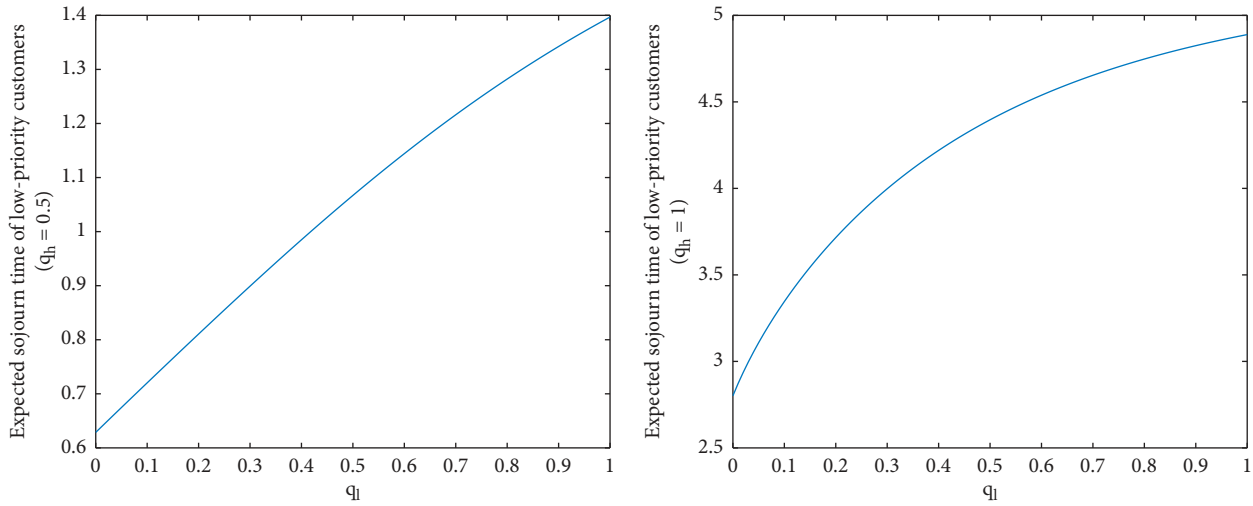

Figure 2: $E[W_l]$ versus q_l of different cases for different values q_h^e. (a) $E[W_l]$ vs. $q_l(q_h^e = 0.5)$. (b) $E[W_l]$ vs. $q_l(q_h^e = 1)$.

Scenario 3. All of the high-priority customers join the priority queue, i.e., $q_h^e = 1$.

The analysis of this scenario is similar to that of Scenario 2, and we will show the results directly in the following theorem.

Theorem 5. *All of the high-priority customers join the priority queue $q_h^e = 1$; if $R_1 \geq c_1/\mu_1 - \lambda_1$, then there exists a unique equilibrium mixed strategy q_l^e of low-priority customers, which has the following forms:*

$$q_l^e = \begin{cases} 0, & R_2 \leq c_2 E[W_l](1,0), \\ q_l^{e^*}, & c_2 E[W_l](1,0) < R_2 < c_2 E[W_l](1,1), \\ 1, & R_2 \geq c_2 E[W_l](1,1), \end{cases} \tag{39}$$

where $E[W_l](1,0) = lim_{q_l \longrightarrow 0} E[W_l](1,q_l)$ and $q_l^{e^*} \in (0,1)$ is the unique solution of the following equation:

$$\begin{aligned} &U_{unl}\left(1, q_l^{e^*}\right) = R_2 - c_2 E[W_l]\left(1, q_l^{e^*}\right) = 0, \\ &E[W_l](1,q_l) = \frac{1}{\mu_1 - \lambda_1}\left(\frac{\lambda_1}{\mu_1 - \lambda_1} + \frac{\mu_1}{\mu_2\theta} - \frac{\mu_1}{\lambda_2 q_l \theta}(1-\theta)\left(1 - \frac{\lambda_1}{\mu_1} - \pi_{0,0}(q_l)\right) - \frac{\mu_1}{\mu_2\theta}\pi_{\bullet,N}(q_l)(1-\theta)^N\right). \end{aligned} \tag{40}$$

Remark 3. For the case $\lambda_1 > \mu_1$, we can obtain the same results as in Theorem 2 and similar results as in Theorem 3; the difference of the results between Theorem 3 and the underlying case is that, under the case $\lambda_1 > \mu_1$, the condition of Theorem 3 "a fraction of high-priority customers joins the priority queue $q_h^e = \mu_1 - c_1/R_1/\lambda_1$ if $c_1/\mu_1 < R_1 < c_1/\mu_1 - \lambda_1$" should be replaced by "a fraction of high-priority customers joins the priority queue $q_h^e = \mu_1 - c_1/R_1/\lambda_1$ if $R_1 > c_1/\mu_1$" without changing the conclusion.

5.3. Sensitivity Analysis. In this subsection, we illustrate some numerical results to study the impact of different parameters on the equilibrium joining probabilities of both types of customers. Without loss of generality, we assume that $\lambda_1 = 3, \lambda_2 = 10, \mu_1 = 5, \mu_2 = 1, R_1 = 10, R_2 = 10, c_1 = 5, c_2 = 1, \theta = 0.2, N = 4$ and conduct the sensitivity analysis by setting a series of values of the target parameters. The exact numerical results are presented in Figure 3.

According to the model assumption, high-priority customers have a preemptive priority than low-priority customers; then, with the change of the parameters related to low-priority customers, there are no impacts on the strategies of high-priority customers, and the exact results are confirmed in Figures 3(b), 3(d)–3(f). In Figure 3(a), we could easily find that both of q_h^e and q_l^e are monotonically decreasing in the parameter λ_1; the difference is that q_l^e decreases faster than q_h^e, which can be attributed to the preemptive priority of high-priority customers. The observations also reveal that, for low values of λ_1, both of the high-priority and low-priority customers have an incentive to join the system because of a shorter waiting time. For high values of λ_1, the system becomes congested; since the high-priority customers have preemptive priority than the low-priority customers, it is plausible that the waiting time of low-priority customers is significantly longer than the high-priority customers, and then the low-priority customers are more willing to balk than the high-priority

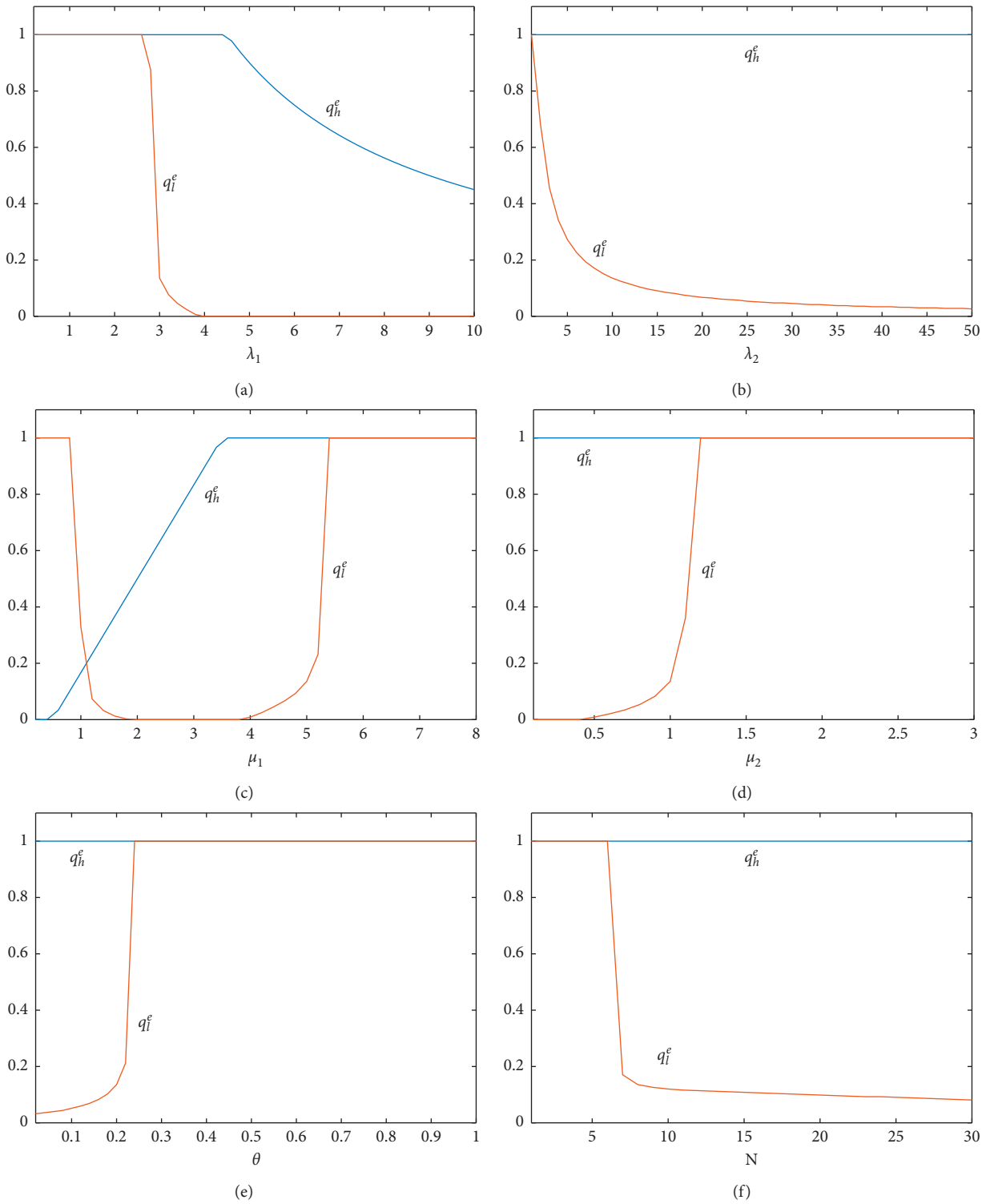

FIGURE 3: q_h^e and q_l^e versus different parameters. (a) q_h^e and q_l^e vs. λ_1. (b) q_h^e and q_l^e vs. λ_2. (c) q_h^e and q_l^e vs. μ_1. (d) q_h^e and q_l^e vs. μ_2. (e) q_h^e and q_l^e vs. θ. (f) q_h^e and q_l^e vs. N.

customers. In addition, Figures 3(b) and 3(d) vividly demonstrate the relationship between the equilibrium joining probability of low-priority customers and their potential arrival rate as well as the corresponding service rate. Figure 3(c) illustrates the following facts: (1) for very low values of μ_1, q_l^e is equal to one, and q_h^e has a low value. (2) For intermediate values of μ_1, q_l^e is equal to zero, and q_h^e gradually increases with the increase of μ_1. (3) For very high values of μ_1, both of q_h^e and q_l^e are equal to one. In short, q_l^e is not a monotonous function of μ_1, and q_h^e is an

increasing function of μ_1. The interesting results are clearly related to the waiting time of joining customers. The low-rate service of high-priority customers causes the high-priority customers to balk, which could reduce the waiting time of low-priority customers and lead the low-priority customers to join the system with a higher probability. And then, as the service rate of high-priority customers gradually increases, the high-priority customers are more willing to join the system due to the reduction of the waiting time, which indirectly affects the joining strategies of low-priority customers. When the service rate of high-priority customers is large enough, the service time of high-priority customers is quite short, and their expected waiting time tends to be zero; that is, the waiting time of high-priority customers has a smaller effect on the low-priority customers, which exactly promotes the low-priority customers to join the system. From Figure 3(e), it is shown that as θ increases, an arriving low-priority customer knows a group leader with a higher probability, and it is more likely to join the group which locates in front of the Israeli queue. So, based on this, there are fewer groups in the system, which also leads to the reduction of the low-priority customers' waiting time and the increase of q_l^e. Figure 3(f) actually indicates that, with the increase of N, the low-priority customers may queue at the back of the Israeli queue, which may increase the expected waiting time of low-priority customers (since more groups exist in the system); at the same time, the low-priority customers are more reluctant to join the system.

6. Optimal Pricing Strategy

In service systems, it is evident that customers are usually heterogeneous in many aspects, such as the evaluation of delay and the waiting cost. Due to the heterogeneity of delay-sensitive customers, the service provider often adopts classification of service through setting priority. However, the classification of service with uniform price will produce the difference of waiting time and service value, which may cause the unfairness psychology for the low-priority customers with longer waiting time. Hence, the service provider needs to adopt the differential pricing so as to eliminate low-priority customers unfairness psychology, where high-priority customers expect to be served as soon as possible and need to pay a high price to save their time in queueing and waiting. Under the specific application background, we will analyse the optimal pricing strategy for both of the high-priority and low-priority customers, so that we could give some management suggestions for the special batch service polling system with priorities. Next, in this section, on the basis of a new assumption that the service provider can set the prices for the two kinds of service, we will deal with the service provider's revenue maximization problem. Hence, the linear utility function of high-priority customers (low-priority customers) $U_1 = R_1 - c_1 E[W_h]$ ($U_2 = R_2 - c_2 E[W_l]$) should be converted to $U_1 = R_1 - p_h - c_1 E[W_h]$ ($U_2 = R_2 - p_l - c_2 E[W_l]$), where p_h and p_l are the setting prices for the high-priority and low-priority service, respectively. That is to say, the symbols R_1 and R_2 in the previous sections should be replaced by $R_1 - p_h$ and $R_2 - p_l$. In order to maximize the service provider's revenue (denoted by Π) in our model, the most common way for the revenue-maximizing service provider is to adjust the prices for both types of services. Owing to the preemptive priority of high-priority customers, setting p_h will affect not only the arrival rate of high-priority customers, but also the arrival rate of low-priority customers, and setting p_l only affects the low-priority customers' joining decisions. So, the decision problem can be formulated as

$$\max_{(p_h,p_l)} \Pi = \lambda_h^e(p_h)p_h + \lambda_l^e(p_l, p_h)p_l, \tag{41}$$

where λ_h^e and λ_l^e are the actual arrival rates. It should be noted that either λ_h^e or λ_l^e has different representations in the observable and unobservable cases. In the observable case, the two types of customers adopt each equilibrium threshold strategy; i.e., λ_h^e and λ_l^e are the effective arrival rates. However, in the unobservable case, both of the high-priority and low-priority customers access the system with each equilibrium joining probability, i.e., $\lambda_h^e = \lambda_1 q_h^e$ and $\lambda_l^e = \lambda_2 q_l^e$. Due to the monotonicity of low-priority customers' utility function, we apply dichotomy method to obtain q_l^e when there exists an intersection between the utility function and x-axis. Meanwhile, q_h^e could be obtained in accordance with Lemma 2 and Remark 2.

However, in most cases, due to the computational complexity, it is quite difficult to obtain the explicit expressions of the effective arrival rates with respect to the corresponding prices. Hence, we aim to search for the numerical optimal solution (p_h^*, p_l^*) of $\max_{(p_h,p_l)} \Pi = \lambda_h^e(p_h)p_h + \lambda_l^e(p_l, p_h)p_l$ through PSO algorithm which will be introduced briefly.

6.1. Particle Swarm Optimization Algorithm. PSO algorithm was firstly presented by Kennedy and Eberhart in [41], which is devoted to solving some continuous nonlinear optimization problems. The algorithm is exempt from the analyticity of the objective function; it does not use the gradient of the optimization problem; that is to say, unlike the classical gradient descent method and quasi-Newton method, PSO algorithm does not require that the optimization problem is differentiable. Due to its simplicity and convenience, this algorithm is frequently applied to search for the global optimal solution. Certainly, PSO is undoubtedly applicable to this two-dimensional optimization problem $\max_{(p_h,p_l)} \Pi = \lambda_h^e(p_h)p_h + \lambda_l^e(p_l, p_h)p_l$. To seek the optimal price combination (p_h^*, p_l^*), the procedure of applying PSO algorithm to this two-dimensional optimization problem is presented as follows: (Algorithm 2)

6.2. Optimal Pricing in the Observable Case. Through Algorithm 1, we can obtain the steady-state probability vectors, and the optimization problem can be formulated as

Input: $\{R_1, \lambda_1, c_1, \mu_1, R_2, \lambda_2, c_2, \mu_2, N\}$
Out: (p_h^*, p_l^*)
Set the number of particles n, the range of (p_h, p_l), and maximum number of iterations.
Generate the number of particles n randomly, denoted by $\{X_1, X_2, \ldots, X_n\}$, where $X_i = (p_h(i), p_l(i))$.
while the number of iterations < the maximum number of iterations **do**
Calculate the fitness value (the profit of service provider) of each particle.
Update the coordinates of the local optimal value {*P*best1, *P*best2, ... *P*best$n$} and the coordinates of the global optimal value *G*best.
Update the coordinates of each particle by the following formulas:
$v_i^k = v_i^{k-1} + a_1 * \text{rand} * (P\text{best}_i^k - X_i^{k-1}) + a_2 * \text{rand} * (G\text{best}_i^k - X_i^{k-1})$.
$X_i^k = X_i^{k-1} + v_i^k$
where k stands for kth iteration, i represents ith particle, a_1 and a_2 are acceleration constants, and \$rand\$ is a random variable in the range of [0, 1].

Algorithm 2: Solve the optimization problem $\max_{(p_h,p_l)} \Pi = \lambda_1^e(p_h)p_h + \lambda_2^e(p_l, p_h)p_l$ by PSO.

$$\max_{(p_h,p_l)} \Pi = \lambda_1(p_h)p_h + \lambda_2(p_l, p_h)p_l, \tag{42}$$

where $\lambda_1(p_h) = \lambda_1(1 - \pi_{n_e+1,\cdot})$ and $\lambda_2(p_l, p_h)$ is an effective arrival rate with a complex expression while high-priority customers do not take the balk strategy.

Similarly, PSO algorithm can be also applied to solve the optimization problem (see, e.g., Wang et al. [42], Zhang et al. [43], and Yang et al. [44]). To keep with the above analysis, we consider a numerical example under the observable case with the assumption that $\lambda_1 = 4, \lambda_2 = 10, R_1 = 10, R_2 = 10, c_1 = 5, c_2 = 1, \theta = 0.2, N = 4$ and illustrate how the parameters μ_1 and μ_2 affect the optimal revenue of the service provider in Table 1, where μ_1 changes from 5 to 10 and μ_2 changes from 1 to 6.

The present results in Table 1 are intuitive and straightforward. As μ_1 or μ_2 is fixed, the optimal value of the service provider's revenue increases with the increase of μ_1 or μ_2. This phenomenon is consistent with the reality; the faster the service is delivered, the more willing the customers are to join the system, which leads to the increasing monotonicity of service provider's revenue.

6.3. Optimal Pricing in the Unobservable Case. We have seen that, for any given price combination (p_h, p_l), there exist the equilibrium joining strategies for both types of customers in the unobservable case. The equilibrium joining probability of high-priority customers can be obtained from the results in Lemma 2 and Remark 2, and the equilibrium joining probability of low-priority customers can only be calculated through numerical algorithm. Hence, we next use PSO algorithm to obtain the numerical results. We first show a numerical example with parameters

$\lambda_1 = 4, \lambda_2 = 10, \quad \mu_1 = 5, \mu_2 = 1, R_1 = 10, R_2 = 10, c_1 = 5, c_2 = 1, \theta = 0.2, N = 4$; then, the revenue function has a maximum value, with the optimal price combination and the corresponding maximum revenue $(p_h^*, p_l^*, \Pi^*) = (8.92, 6.79, 71.39)$. The result reveals that high-priority service price should be greater than low-priority service price; it is reasonable that the difference in waiting time between high-priority and low-priority customers will cause the sense of injustice to the low-priority customers; therefore, the service provider should set a higher price for the high-priority customers to ease the low-priority customers' sense of injustice.

Moreover, under the two levels of information (observable and unobservable), we also find that, for the same parameters, the optimal revenue of the unobservable case is much greater than the one of the observable case; i.e., there is a big gap between the two cases. For this interesting phenomenon, there may be several reasons for explaining this. The setting parameters could lead to a high service intensity of high-priority customers; thus the system is prone to congestion of high-priority customers. In addition, the arrival rate of high-priority customers is greater than the service rate of low-priority customers, which could also increase the possibility of low-priority customers being preempted. So, based on the same parameters $\lambda_1 = 4, \lambda_2 = 10, \mu_1 = 5, \mu_2 = 1, R_1 = 10, \quad R_2 = 10, c_1 = 5, c_2 = 1, \theta = 0.2, N = 4$, for the unobservable case, low-priority customers always join the system with a certain probability, while, for the observable case, the low-priority customers will join the system with pure strategy only when a few of high-priority customers exist in the system. Thus, in an extremely unfavorable situation for low-priority customers, it is more likely to cause the loss of low-priority customers in the observable case. Finally, based on the above intuitive analysis, when there is service intensity of high-priority customers, the potential arrival rate of low-priority and high-priority customers is large, such as the case

$\lambda_1 = 4, \lambda_2 = 10, \mu_1 = 5, \mu_2 = 1, \quad R_1 = 10, R_2 = 10, c_1 = 5, c_2 = 1, \theta = 0.2, N = 4$; the optimal revenue of the unobservable case is greater than the one of the observable case.

In order to verify the accuracy of PSO algorithm, Figure 4 shows the service provider's revenue versus the prices p_h and p_l. Besides, in order to intuitively illustrate the accuracy and rapidity of PSO method, we give the graph of PSO iteration process. From Figure 5, we can observe that, by using PSO algorithm, it just takes 10 iterations to reach an accurate result, which exactly implies the superiority of the PSO algorithm.

Table 1: The optimal revenue of service provider versus μ_1 and μ_2.

μ_2/μ_1	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
1	25.81	27.91	30.16	43.09	52.64	59.92	65.69	70.45	74.02	76.49	79.13
1.5	28.26	41.74	57.52	68.09	74.57	80.89	85.03	88.91	91.35	93.93	95.75
2	39.72	60.17	72.38	80.92	86.82	91.71	95.13	97.81	100.61	101.66	103.98
2.5	50.61	70.67	81.46	88.54	93.65	97.93	101.36	103.62	105.71	106.96	109.05
3	60.45	77.88	87.3	93.75	98.6	102.19	105	106.93	109.24	110.85	112.06
3.5	67.23	82.95	91.43	97.09	101.81	105.4	107.61	109.84	110.66	113.32	114.84
4	73.02	86.89	94.5	99.47	104.27	107.43	109.99	111.58	113.83	115.15	116.29
4.5	76.9	90.21	97.19	101.87	106.12	109.02	111.26	113	115.31	116.22	117.96
5	79.97	92.66	99.03	103.85	107.81	110.19	112.54	114.89	115.98	117.73	119.09
5.5	83.09	94.25	100.28	105.43	109.07	111.77	113.91	115.31	117.32	118.68	119.43
6	84.64	95.98	101.67	106.35	109.75	112.64	114.77	116.8	117.96	119.23	120.51

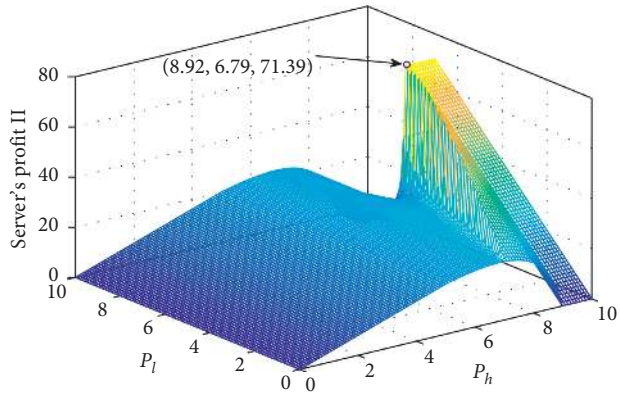

Figure 4: The service provider's revenue versus the prices p_h and p_l.

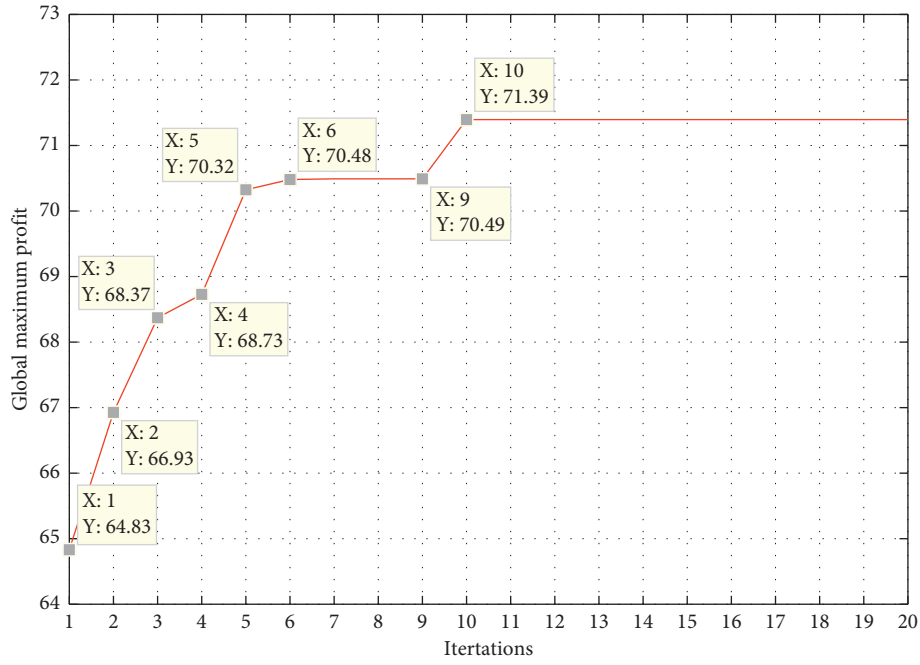

Figure 5: Iteration diagram of global maximum profit.

7. Conclusions and Future Research

In this study, we considered a special batch service polling system with priorities and investigated the strategic behavior of high-priority and low-priority customers, regarding the joining or balking dilemma, under the two levels of information. We first derived the equilibrium joining strategies of both types of customers, and then we illustrated some numerical results to study the impact of different parameters on the equilibrium probabilities of both types of customers. Based on the customer equilibrium analysis, we further analysed the service provider's pricing decisions by employing the PSO algorithm to search for the maximum point of service provider's revenue and showed the optimal price combination for the high-priority and low-priority services under the observable and unobservable cases. We hope that the derived results could bring some management implications to the service provider and expect that the results can be applied to more practical queueing systems.

Apart from the results derived by the present paper, a lot of interesting directions on this topic have not been fully investigated. Therefore, it would be interesting to consider further extensions of the present queueing model. One direction is that the batch service size is a fixed value rather than an unlimited size. Then, we could investigate the influences of the different levels of information and the fixed batch size on the strategic behavior of customers and the service provider's pricing decisions; meanwhile, we could discuss the optimal admission control problem, i.e., how to set the optimal batch size of each group so as to maximize the service provider's revenue.

Another interesting direction for future research is that the strategic customers exhibit bounded rationality rather than full rationality and study the impact of bounded rationality from a profit-maximizing service provider's perspective under different levels of information. We think that it would be an interesting and important direction for future research.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Shandong Provincial Natural Science Foundation (China) (Grant no. ZR2019BG014), National Natural Science Foundation of China (Grant no. 71902105), and the Scientific Research Foundation of Shandong University of Science and Technology for Recruited Talents (Grant no. 2019RCJJ016).

References

[1] N. Perel and U. Yechiali, "The Israeli queue with priorities," *Stochastic Models*, vol. 29, no. 3, pp. 353–379, 2013.

[2] A. Economou and A. Manou, "Equilibrium balking strategies for a clearing queueing system in alternating environment," *Annals of Operations Research*, vol. 208, no. 1, pp. 489–514, 2013.

[3] A. Manou, A. Economou, and F. Karaesmen, "Strategic customers in a transportation station: when is it optimal to wait?" *Operations Research*, vol. 62, no. 4, pp. 910–925, 2014.

[4] O. Bountali and A. Economou, "Equilibrium joining strategies in batch service queueing systems," *European Journal of Operational Research*, vol. 260, no. 3, pp. 1142–1151, 2017.

[5] O. Bountali and A. Economou, "Equilibrium threshold joining strategies in partially observable batch service queueing systems," *Annals of Operations Research*, vol. 277, no. 2, pp. 231–253, 2019.

[6] S. Drekic and W. K. Grassmann, "An eigenvalue approach to analyzing a finite source priority queueing model," *Annals of Operations Research*, vol. 112, no. 1/4, pp. 139–152, 2002.

[7] O. Jouini and A. Roubos, "On multiple priority multi-server queues with impatience," *Journal of the Operational Research Society*, vol. 65, no. 5, pp. 616–632, 2014.

[8] H. Takagi, "Waiting time in the M/M/$$ m $$ m LCFS nonpreemptive priority queue with impatient customers," *Annals of Operations Research*, vol. 247, no. 1, pp. 257–289, 2016.

[9] I. Atencia, "A Geo/G/1 retrial queueing system with priority services," *European Journal of Operational Research*, vol. 256, no. 1, pp. 178–186, 2017.

[10] J. Van der Wal and U. Yechiali, "Dynamic visit-order rules for batch-service polling," *Probability in the Engineering and Informational Sciences*, vol. 17, no. 3, pp. 351–367, 2003.

[11] O. J. Boxma, Y. van der Wal, and U. Yechiali, "Polling with batch service," *Stochastic Models*, vol. 24, no. 4, pp. 604–625, 2008.

[12] N. Perel and U. Yechiali, "The Israeli queue with infinite number of groups," *Probability in the Engineering and Informational Sciences*, vol. 28, no. 1, pp. 1–19, 2014.

[13] N. Perel and U. Yechiali, "The Israeli Queue with retrials," *Queueing Systems*, vol. 78, no. 1, pp. 31–56, 2014.

[14] N. Perel and U. Yechiali, "The Israeli Queue with a general group-joining policy," *Annals of Operations Research*, 2015.

[15] T. Jiang, L. Liu, and L. Liu, "Analysis of a batch service multi-server polling system with dynamic service control," *Journal of Industrial & Management Optimization*, vol. 14, no. 2, pp. 743–757, 2018.

[16] T. Jiang, L. Liu, and Y. Zhu, "Analysis of a batch service polling system in a multi-phase random environment," *Methodology and Computing in Applied Probability*, vol. 20, no. 2, pp. 699–718, 2018.

[17] T. Jiang, "Tail asymptotics for a batch service polling system with retrials and nonpersistent customers," *Journal of Mathematical Analysis and Applications*, vol. 459, no. 2, pp. 893–905, 2018.

[18] P. Naor, "The regulation of queue size by levying tolls," *Econometrica*, vol. 37, no. 1, pp. 15–24, 1969.

[19] N. M. Edelson and D. K. Hilderbrand, "Congestion tolls for Poisson queuing processes," *Econometrica*, vol. 43, no. 1, pp. 81–92, 1975.

[20] R. Hassin and M. Haviv, "Equilibrium threshold strategies: the case of queues with priorities," *Operations Research*, vol. 45, no. 6, pp. 966–973, 1997.

[21] R. Hassin and M. Haviv, *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, Kluwer, Boston, MA, USA, 2003.
[22] R. Hassin, *Rational Queueing*, Chapman & Hall, London, UK, 2016.
[23] R. Hassin and R. Roet-Green, "The impact of inspection cost on equilibrium, revenue, and social welfare in a single-server queue," *Operations Research*, vol. 65, no. 3, pp. 804–820, 2017.
[24] R. Ibrahim, "Sharing delay information in service systems: a literature survey," *Queueing Systems*, vol. 89, no. 1-2, pp. 49–79, 2018.
[25] O. Bountali and A. Economou, "Strategic customer behavior in a two-stage batch processing system," *Queueing Systems*, vol. 93, no. 1-2, pp. 3–29, 2019.
[26] Z. Wang, L. Liu, Y. Shao, X. Chai, and B. Chang, "Equilibrium joining strategy in a batch transfer queuing system with gated policy," *Methodology and Computing in Applied Probability*, vol. 22, no. 1, pp. 75–99, 2018.
[27] I. J. B. F. Adan, V. G. Kulkarni, N. Lee, and E. Lefeber, "Optimal routeing in two-queue polling systems," *Journal of Applied Probability*, vol. 55, no. 3, pp. 944–967, 2018.
[28] X. Chai, L. Liu, B. Chang, T. Jiang, and Z. Wang, "On a batch matching system with impatient servers and boundedly rational customers," *Applied Mathematics and Computation*, vol. 354, pp. 308–328, 2019.
[29] S. Drekic and D. G. Woolford, "A preemptive priority queue with balking," *European Journal of Operational Research*, vol. 164, no. 2, pp. 387–401, 2005.
[30] P. Afèche and H. Mendelson, "Pricing and priority auctions in queueing systems with a generalized delay cost structure," *Management Science*, vol. 50, no. 7, pp. 869–882, 2014.
[31] S. Gavirneni and V. G. Kulkarni, "Self-selecting priority queues with Burr distributed waiting costs," *Production and Operations Management*, vol. 25, no. 6, pp. 979–992, 2016.
[32] J. Wang, S. Cui, and Z. Wang, "Equilibrium strategies inM/M/1 priority queues with balking," *Production and Operations Management*, vol. 28, no. 1, pp. 43–62, 2019.
[33] G. A. J. F. Brouns and J. van der Wal, "Optimal threshold policies in a two-class preemptive priority queue with admission and termination control," *Queueing Systems*, vol. 54, no. 1, pp. 21–33, 2006.
[34] C. Liu and R. Berry, "A priority queue model for competition with shared spectrum," in *Proceedings of the 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 629–636, Monticello, IL, USA, October 2014.
[35] B. Xu, X. Xu, and X. Wang, "Optimal balking strategies for high-priority customers in M/G/1 queues with 2 classes of customers," *Journal of Applied Mathematics and Computing*, vol. 51, no. 1-2, pp. 623–642, 2016.
[36] B. Xu, X. Xu, and Z. Yao, "Equilibrium and optimal balking strategies for low-priority customers in the M/G/1 queue with two classes of customers and preemptive priority," *Journal of Industrial & Management Optimization*, vol. 15, no. 4, pp. 1599–1615, 2019.
[37] I. Adiri and U. Yechiali, "Optimal priority-purchasing and pricing decisions in nonmonopoly and monopoly queues," *Operations Research*, vol. 22, no. 5, pp. 1051–1066, 1974.
[38] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, SIAM, Philadelphia, PA, USA, 1999.
[39] Q. Wang and C. Sun, "Adaptive consensus of multiagent systems with unknown high-frequency gain signs under directed graphs," *IEEE Transactions on Systems Man & Cybernetics Systems*, vol. 50, no. 6, pp. 2181–2186, 2020.
[40] Q. Wang, H. E. Psillakis, and C. Sun, "Cooperative control of multiple agents with unknown high-frequency gain signs under unbalanced and switching topologies," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2495–2501, 2020.
[41] J. Kennedy and R. Eberhart, "Particle swarm optimization. Neural Networks," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Piscataway, NJ, USA, December 1995.
[42] J. Wang, X. Zhang, and P. Huang, "Strategic behavior and social optimization in a constant retrial queue with the N-policy," *European Journal of Operational Research*, vol. 256, no. 3, pp. 841–849, 2017.
[43] X. Zhang, J. Wang, and Q. Ma, "Optimal design for a retrial queueing system with state-dependent service rate," *Journal of Systems Science and Complexity*, vol. 30, no. 4, pp. 883–900, 2017.
[44] D.-Y. Yang, Y.-H. Chen, and C.-H. Wu, "Modelling and optimisation of a two-server queue with multiple vacations and working breakdowns," *International Journal of Production Research*, vol. 58, no. 10, pp. 3036–3048, 2019.