



In-queue priority purchase: a dynamic game approach

Zhongbin Wang¹ · Luyi Yang² · Shiliang Cui³ · Jinting Wang⁴

Received: 29 October 2019 / Revised: 3 February 2021 / Accepted: 8 February 2021 /
Published online: 5 March 2021
© The Author(s) 2021

Abstract

Pay-for-priority is a common practice in congestion-prone service systems. The extant literature on this topic restricts attention to the case where the only epoch for customers to purchase priority is *upon arrival*, and if customers choose not to upgrade when they arrive, they cannot do so later during their wait. A natural alternative is to let customers pay and upgrade to priority at any time during their stay in the queue, even if they choose not to do so initially. This paper builds a queueing-game-theoretic model that explicitly captures self-interested customers' dynamic *in-queue* priority-purchasing behavior. When all customers (who have not upgraded yet) *simultaneously* decide whether to upgrade, we find in our model that pure-strategy equilibria do not exist under some intuitive criteria, contrasting the findings in classical models where customers can only purchase priority upon arrival. However, when customers *sequentially* decide whether to upgrade, threshold-type pure-strategy equilibria may exist. In particular, under sufficiently light traffic, if the number of ordinary customers accumulates to a certain threshold, then it is always the second last customer who upgrades, but in general, it could be a customer from another position, and the queue-length threshold that triggers an upgrade can also vary with the traffic intensity. Finally, we find that in-queue priority purchase subject to the sequential rule yields less revenue than upon-arrival priority purchase in systems with small buffers.

✉ Luyi Yang
luyiyang@berkeley.edu

Zhongbin Wang
zawang@nankai.edu.cn

Shiliang Cui
shiliang.cui@georgetown.edu

Jinting Wang
jtwang@cufe.edu.cn

- ¹ Business School, Nankai University, Tianjin, China
- ² Haas School of Business, The University of California, Berkeley, Berkeley, CA, USA
- ³ McDonough School of Business, Georgetown University, Washington, DC, USA
- ⁴ School of Management Science and Engineering, Central University of Finance and Economics, Beijing, China

Keywords In-queue decision · Dynamic game · Priority upgrade · Self-interested customers

Mathematics Subject Classification 60K30 · 90B22 · 91A25 · 91A80

1 Introduction

Pay-for-priority is a common practice in service systems for customers to cut their waiting time in the queue. For instance, the London Eye sells fast-track tickets to its visitors. Similarly, the U.S. Citizenship and Immigration Services (USCIS) charge a premium in exchange for expedited case processing.¹ One key design variable for such a system is the timing of customers' purchase. The extant literature on priority purchase [1,6,15] has restricted attention to the case where customers can only purchase priority *upon arrival*—if they decide not to pay when they arrive, they will forgo the opportunity to upgrade ever and must remain in the regular line throughout their stay, even though, in practice, they may change their minds as the queue evolves.

A natural alternative is to give customers more flexibility and make the pay-for-priority option always available—even after customers join the system. For instance, the London Eye could in principle allow visitors to purchase the fast-track ticket on their mobile devices at any point in their waiting process. Likewise, USCIS could allow a petitioner to file a request of expedited processing at any time after submission, not just at the moment of submission. If *in-queue* priority purchase is enabled, an immediate follow-up design question is whether to let customers in the regular line upgrade all at once (i.e., *simultaneously*) or one at a time (i.e., *sequentially*). In light of these considerations, this paper examines the following two questions: (1) how will customers behave if they can upgrade to priority at any time during their wait in the queue? (2) how will rules on the specific timing of priority upgrade (i.e., simultaneous vs. sequential) impact customer behavior?

Modeling the behavior of dynamic *in-queue* priority-purchasing is no easy task. When deciding whether to purchase priority, customers must make a delicate trade-off between the priority premium they pay and the waiting time they expect to save, which, in turn, depends on how many other customers are seeking service and competing for priority. Additionally, self-interested customers must take into account all possible future events of the queue and decide not only *whether* to purchase priority (as in the upon-arrival case in the literature) but also *when* to do so (assumed away by the upon-arrival case). These actions, in turn, shape the underlying queueing dynamics. Capturing this feedback loop necessitates (challenging) equilibrium analysis.

In this study, we formulate the problem of in-queue priority purchase as a dynamic game played by homogeneous customers in a queueing system. We first consider a simultaneous upgrade rule whereby customers who have not yet upgraded to priority simultaneously decide whether to upgrade at any time while they are waiting in the queue.

¹ See <https://www.londoneye.com/tickets-and-prices/>; <https://www.uscis.gov/i-907>.

Since the strategy space and the value-function space of our game are massive, we take a two-step approach and demonstrate that pure-strategy equilibria do not exist under certain intuitive criteria on the value functions (of the dynamic game). In Step 1, we establish that if a pure strategy can be sustained in equilibrium subject to the intuitive criteria, it must be of a state-dependent threshold type. That is, customers in the non-priority line will purchase priority if and only if their positions from the head of the line are less than a certain threshold, which varies with the system state characterized by both the number of priority customers and the number of non-priority customers in the system. In Step 2, we show that any threshold strategy cannot be sustained in equilibrium. Further, we investigate a tractable small buffer system that can hold at most two customers (including the customer in service). We analytically confirm, without assuming the intuitive criteria, that such a small buffer system indeed cannot support a pure-strategy equilibrium in which some customers purchase priority.

Our finding of non-equilibria contrasts those in the extant literature that focuses on the setting where customers can purchase priority only at the instant of arrival [1, 6, 15]. In these papers, not only is a pure-strategy equilibrium guaranteed to exist but there often exist multiple pure-strategy equilibria (see Sect. 1.1 for more details about the related literature.) The non-existence of pure-strategy equilibria in our setting also implies that the simultaneous upgrade rule can be troublesome to implement, let alone the practical difficulty that may arise for the system to update state information in real-time without delay.

Next, we consider an alternative sequential upgrade rule whereby each time there is a new customer arrival, the service provider sequentially asks non-priority customers one by one whether to purchase priority, with the first customer in line being called upon first, then the second one, and so on. Today's mobile Internet makes it exceedingly easy to manage such a system efficiently.

We first investigate the cases of sufficiently light traffic and sufficiently heavy traffic. Under sufficiently low traffic, customers can effectively act myopically by ignoring future arrivals. We establish the existence of a pure-strategy equilibrium of the following structure: each time the arrival of a new customer causes the low-priority queue length to tentatively reach a given threshold (given in closed form), the second last customer (and no one else) upgrades; the newcomer will be the next customer to upgrade if the low-priority queue length temporarily reaches the threshold again (due to an arrival) before the newcomer is served. This equilibrium structure implies the low-priority queue length can never exceed the above threshold (and can only stay at the threshold temporarily). This property is preserved under any traffic intensity in equilibrium even though the maximum low-priority queue length would likely decrease with traffic. In particular, under sufficiently heavy traffic, we show that when the priority price is not too high, the equilibrium is such that as soon as the ordinary queue amasses two customers, the first ordinary customer upgrades to the priority queue.

We then again turn to a small buffer system that can hold at most two customers and analytically characterize the pure-strategy equilibrium for any traffic intensity. We also analytically compare in-queue priority purchase subject to the sequential rule with upon-arrival priority purchase in the small buffer system. We find that for a revenue-maximizing service provider, the optimal priority revenue is higher under upon-arrival

priority purchase although it presents customers with fewer upgrade opportunities. This result potentially tells a cautionary tale against in-queue priority purchase (despite its ability to create more selling opportunities) and implies that the service provider may benefit from prohibiting such a practice.

Finally, we numerically solve for the pure-strategy equilibrium of sequential in-queue priority purchase in a system that can hold at most three customers under various traffic intensities and priority prices. We find instances in which as soon as the number of ordinary customers reaches three, it triggers the upgrade of the very first customer but not the second or third customer, defying the equilibrium structure previously identified. Nevertheless, we numerically observe that in such a system, (sequential) in-queue priority purchase still generates less revenue than upon-arrival priority purchase, corroborating the analytical insight gleaned from the small buffer system.

1.1 Literature review

The literature on queueing models with strategic customers dates back to the seminal paper of Naor [20], who studies homogeneous customers' joining and balking decisions upon arrival to an observable M/M/1 queue. Interested readers are referred to Hassin and Haviv [16] and Hassin [14] for comprehensive surveys of this literature. Particularly related to our work are papers on priority-purchasing behavior. While a significant proportion of the priority-purchasing literature focuses on unobservable queues (for example, [2,11–13,19,21,24–26]), a relatively scant stream of the literature examines priority purchasing in an observable-queue setting in which customers make queue-length-dependent decisions, which is most relevant to our work.

Balachandran [9] studies a queueing system in which customers receive a higher priority by choosing a higher price to pay upon arrival from an infinite set of possible payments. Adiri and Yechiali [1] and Hassin and Haviv [15] examine a more practical scenario which serves as the basis for our model. Specifically, Adiri and Yechiali [1] analyze an observable M/M/1 queue when customers make pay-for-priority decisions upon arrival to the system, given a fixed cost of the priority premium. They consider pure threshold equilibrium strategies whereby customers will purchase priority if and only if the total number of customers in the system is above a certain threshold. Because customers are homogeneous, it implies that under any symmetric equilibrium, an arriving customer will also opt into priority if there are other priority customers in the system. Hassin and Haviv [15] build on the work of Adiri and Yechiali [1] and highlight the follow-the-crowd behavior in priority purchasing. Importantly, they show that such follow-the-crowd behavior guarantees the existence of at least one pure-strategy equilibrium and potentially multiple ones. By contrast, we show that this existence result breaks once customers are allowed to defer their priority-purchasing decision (under a simultaneous upgrade rule).

Alperstein [6] further generalizes the framework by Adiri and Yechiali [1] to multiple priority classes and shows that a revenue-maximizing service provider can extract all customer surplus by implementing priorities, effectively achieving a last-in-first-out (LIFO) service discipline. More recently, Wang et al. [22] compare the revenue

of an M/M/1 priority queue with balking, between the observable and unobservable settings.

Similar to this stream of literature, we also study customers' strategic priority-purchasing strategies in an observable M/M/1 queue with a given priority premium charged to customers. However, unlike the extant literature, which assumes that customers can only purchase priority upon arrival, our model allows for priority-purchasing at any time of the queueing process. In other words, the existing models are essentially a one-shot game, whereas ours is a dynamic one. Indeed, customers in our model may defer their priority purchase until the queue gets longer, or find it optimal not to purchase priority upon arrival even if there are other priority customers currently in the system. Such behaviors make intuitive sense but are absent from the existing models.

While we study customers' in-queue priority-purchasing behavior, a few other recent papers consider customers' in-queue renegeing decisions; for example, see Assaf and Haviv [7], Afèche and Sarhangian [3], Maglaras et al. [18], Ata and Peng [8], and Cui et al. [10] for various theoretical models of renegeing. Also see Akşin et al. [4,5] and Webb et al. [23] for empirical investigations. Among these works, the most relevant one to us is Afèche and Sarhangian [3]. They assume two exogenously fixed streams of customers (ordinary and priority) arriving to a priority queue without any priority premium involved, and study how being bumped by the priority customers triggers in-queue renegeing behavior of the ordinary customers (the priority customers may balk upon arrival but do not have any incentive to renege after joining). Importantly, Afèche and Sarhangian [3] assume exogenous priority status to focus on the renegeing behavior, whereas our model complements theirs by allowing customers to endogenously self-select into priorities through monetary payment while assuming away balking and renegeing.

In our investigation of the simultaneous upgrade case, the focus on showing non-equilibria of structurally simple strategies is in the spirit of Kerner et al. [17] who prove that any threshold strategies cannot be an equilibrium in ticket queues. Rather than restricting attention to threshold strategies, we propose certain intuitive criteria that any equilibrium strategy should presumably follow. We then demonstrate that any pure strategies that satisfy these criteria must be of a threshold type, yet they cannot be supported in equilibrium.

2 Model description

Consider an M/M/1 service system. Customers arrive to the system according to a Poisson process with rate λ . The service times are independent and exponentially distributed with rate μ . Customers do not balk or renege. Hence, we focus on the case $\lambda < \mu$ to ensure system stability. Customers are delay-sensitive and their waiting cost per unit time is C . Consistent with the literature, the values of λ , μ , and C are common knowledge.

By default, each customer upon arrival is an *ordinary customer*, i.e., a non-priority customer, and decides whether and when to purchase priority throughout her stay in the system (from the arrival epoch to the departure epoch). The priority price is $P (> 0)$

and is non-refundable. Hence, the decision of upgrading to priority is irrevocable. Once an ordinary customer purchases priority, she becomes a *priority customer* and obtains *preemptive* priority for service over all other ordinary customers. The queue disciplines within the ordinary and priority lines are both FIFO. The state of the system is described by $\{N_o, N_p\}$, where $N_o \in \mathbb{N}_0 := \{0\} \cup \mathbb{N}$ and $N_p \in \mathbb{N}_0$ correspond to the number of ordinary customers and the number of priority customers in the system (including the one in service, if any), respectively. Throughout her time in the system, each customer observes the system state and her position in the queue.

Customers are fully rational in our model in that (i) they act to maximize their own expected utility (or equivalently, minimize the expected cost) at any given time upon and after arriving to the system, and (ii) they take into consideration the actions of other customers, including the current customers in the system as well as any future arrivals. To that end, customers are able to calculate the probabilities for all different sample paths according to which the system evolves. Given that the priority premium is non-refundable, once any customer has purchased priority, they have no more actions to take.

We consider two different rules that can be imposed on the specific timing of priority upgrade: the simultaneous upgrade rule (which we study in detail in Sect. 3) and the sequential-upgrade rule (which we study in detail in Sect. 4).

3 Simultaneous upgrade rule

In this section, we consider a simultaneous upgrade rule specified as follows: Each ordinary customer in the system *continuously* evaluates the options of purchasing and not purchasing priority—until she either upgrades to priority or completes service. The evaluation and priority upgrade are instantaneous. Thus, *at any time point*, all (ordinary) customers *simultaneously* decide whether to upgrade to priority.

When multiple ordinary customers decide to upgrade, it is imperative to specify the order in which these customers join the priority line, i.e., their service order (because the order affects the calculation of customers' expected utilities and hence the equilibrium analysis). We adopt the *first-come-first-upgrade rule*, i.e., customers who upgrade at the same time will join the priority line according to their order of arrival to the system, which is also their order in the ordinary line. First-come-first-upgrade is arguably the fairest and most natural rule for customers.

3.1 Equilibrium definition

We set up the in-queue priority-purchasing problem under the simultaneous rule as a dynamic game. We focus on Markovian priority-purchasing strategies that depend on the system state $\{N_o, N_p\}$ (i.e., the numbers of ordinary and priority customers in the system), and the position of a given customer within the ordinary line (due to the first-come-first-upgrade rule). The information set an ordinary customer acts on can thus be described by a three-dimensional position vector (i, j, k) , where $i \in \mathbb{N}$ indicates the position of the customer in the ordinary line including any ordinary customer at the

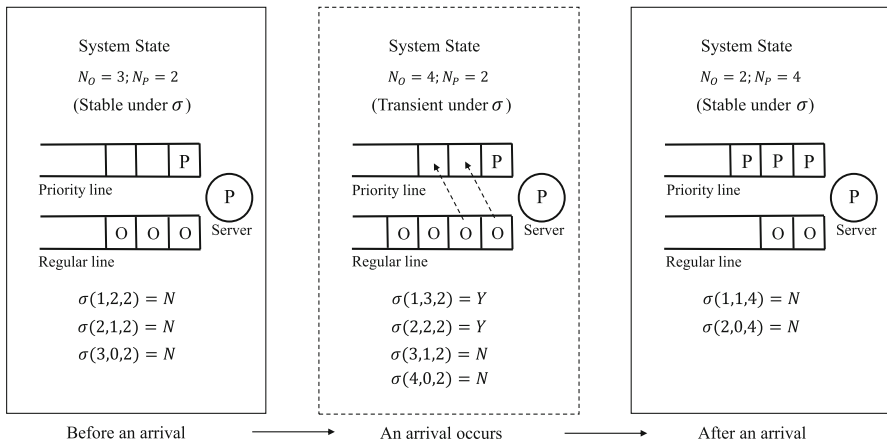


Fig. 1 System dynamics as an arrival occurs, assuming customers follow strategy σ specified in each box. Note: Os and Ps in the figure represent ordinary and priority customers, respectively. As a new arrival occurs (middle panel), the four ordinary customers (including the arrival) act simultaneously according to strategy σ and the first-come-first-upgrade rule. If σ were an equilibrium strategy, no ordinary customer has an incentive to unilaterally deviate from $\sigma(i, j, k)$ given her position (i, j, k) at any stable or transient state under σ

server, $j \in \mathbb{N}_0$ the number of ordinary customers behind her in the ordinary line, and $k \in \mathbb{N}_0$ the number of customers already in the priority queue. For example, if the only customer in the system is an ordinary customer at the server, then her position vector is $(i, j, k) = (1, 0, 0)$. By definition, at any system state $\{N_o, N_p\}$ with $N_o \geq 1$, any ordinary customer’s position vector (i, j, k) must satisfy $i + j = N_o$ and $k = N_p$.

Given the position vector (i, j, k) , each ordinary customer chooses between ‘Yes’ (for purchasing priority) and ‘No’ (for not purchasing priority). A (Markovian) strategy $\sigma : \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}_0 \rightarrow \{Y, N\}$ is a mapping from the position vector (i, j, k) to a Yes or No priority-purchasing action. We use Σ to denote the strategy space. Because customers are homogeneous, we consider symmetric (pure) strategies. When all customers follow strategy σ , we call system state $\{N_o, N_p\}$ a *stable state under σ* if and only if (i) $N_o = 0$, or (ii) $N_o > 0$ and $\sigma(i, N_o - i, N_p) = N$ for all $i \in \{1, 2, \dots, N_o\}$. That is, at a stable state (under σ), the strategy σ specifies that all of the ordinary customers, if any, do not purchase priority (i.e., to stay “stable”). A system state that is not stable under σ will be referred to as a *transient state under σ* . See Fig. 1 for an illustration of the system dynamics when customers follow a specific strategy.

Define the value function $V(i, j, k) : \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}_0 \rightarrow \mathbb{R}$ as the expected utility (or the continuation value) of an ordinary customer at position (i, j, k) . We use \mathcal{V} to denote the value-function space. Note that $V(i, j, k)$ is time-homogeneous because the underlying queueing system evolves according to a time-homogeneous Markov chain.

Definition 1 A symmetric pure-strategy equilibrium under the simultaneous upgrade rule is characterized by any strategy and value-function pair $(\sigma, V) \in \Sigma \times \mathcal{V}$ that satisfies Conditions (1), (3a)–(3b), (4) and (5):

$$\begin{aligned}
 V(i, j, k) &= \begin{cases} -P - \left[k + \sum_{s=1}^{i-1} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}} + 1 \right] \frac{C}{\mu}, & \text{if } \sigma(i, j, k) = Y, \\ V(i', j', k'), & \text{otherwise,} \end{cases} \\
 \forall(i, j, k) &\in \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}_0,
 \end{aligned} \tag{1}$$

where (i', j', k') is as specified by

$$\begin{aligned}
 i' &:= i - \sum_{s=1}^{i-1} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}}, \\
 j' &:= j - \sum_{s=i+1}^{i+j} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}}, \\
 k' &:= k + \sum_{s=1}^{i-1} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}} + \sum_{s=i+1}^{i+j} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}}.
 \end{aligned} \tag{2}$$

In particular, $\forall(i, j, k) \in \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}_0$, if $\sigma(s, i+j-s, k) = N$ for all $s = 1, \dots, i+j$, then

$$\begin{aligned}
 V(i, j, k) &= -\frac{C}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} V(i, j, k - 1) + \frac{\lambda}{\lambda + \mu} V(i, j + 1, k), \\
 \forall(i, j, k) &\in \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N},
 \end{aligned} \tag{3a}$$

$$\begin{aligned}
 V(i, j, 0) &= -\frac{C}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} V(i - 1, j, 0) + \frac{\lambda}{\lambda + \mu} V(i, j + 1, 0), \\
 \forall(i, j) &\in \mathbb{N} \times \mathbb{N}_0,
 \end{aligned} \tag{3b}$$

$$V(0, j, 0) \equiv 0, \quad \forall j \in \mathbb{N}_0, \tag{4}$$

$$\begin{aligned}
 V(i, j, k) &= \max \left\{ -P - \left[k + \sum_{s=1}^{i-1} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}} + 1 \right] \frac{C}{\mu}, V(i', j', k') \right\}, \\
 \forall(i, j, k) &\in \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}_0,
 \end{aligned} \tag{5}$$

where (i', j', k') is as specified by (2).

In Definition 1, Conditions (1), (3a)–(3b) and (4) pin down the value function V (through a system of linear equations) for a given strategy σ . Specifically, Condition (1) gives recursive formulas for the value function when state transitions occur due to priority upgrades without changing the total number of customers in the system. If σ requires that an ordinary customer with position (i, j, k) purchase priority at the cost of premium P , then she will join the priority line with the $\left[k + \sum_{s=1}^{i-1} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}} + 1 \right]^{th}$ position in the priority line according to the first-come-first-upgrade rule, which means she must wait behind the original k priority customers and all upgraders who are originally ahead of her. Her expected waiting cost

until service completion is thus $\left[k + \sum_{s=1}^{i-1} \mathbb{1}_{\{\sigma(s,i+j-s,k)=Y\}} + 1 \right] C/\mu$. If instead she chooses not to purchase priority, her position vector changes from (i, j, k) to (i', j', k') as specified by (2), which indicates that her updated position i' in the ordinary line is her original position i less the number of upgraders who are originally ahead of her; the updated number of customers waiting behind her j' is the original number j less the number of upgraders who are originally behind her; and the updated number of priority customers k' is the original number k plus the total number of upgraders.

Note that Condition (1) is applicable to both transient and stable states, although for a stable state, it would simply give a trivial identity equation, i.e., $V(i, j, k) = V(i, j, k)$. Hence, Conditions (3a)–(3b) add to Condition (1) by specifying the recursive formulas of the value function for stable states (in which everyone in the ordinary line chooses ‘ N ’ according to σ). A stable state evolves if and only if an arrival or departure event occurs, which [unlike transitions in Condition (1)] changes the total number of customers in the system. Specifically, the mean time till the next (arrival or departure) event is $1/(\lambda + \mu)$, during which each ordinary customer incurs a waiting cost of C per unit of time. If an event occurs, it is an arrival with probability $\lambda/(\lambda + \mu)$, in which case the number of customers waiting behind any ordinary customer is incremented by one. On the other hand, with probability $\mu/(\lambda + \mu)$, a departure occurs as the next event, in which case either the number of priority customers is decremented by one if originally there is at least one priority customer (see Condition (3a)), or the position of any ordinary customer moves up by one otherwise (see Condition (3b)). On the other hand, Condition (4) specifies boundary conditions for absorbing states of service completion.

Note that while the value function V is determined by Conditions (1), (3a)–(3b) and (4) for a given strategy σ , there is of guarantee on σ being an equilibrium. Hence, Condition (5) acts as a consistency check that ensures the continuation value $V(i, j, k)$ is indeed the maximum expected utility any customer with position (i, j, k) can obtain (even if they could choose differently than $\sigma(i, j, k)$), provided that all other customers follow σ . That is, when all customers adopt σ , no one can strictly improve their expected utility at any position (i, j, k) by unilaterally deviating from the action specified by $\sigma(i, j, k)$, which implies the best response to σ coincides with σ itself. Taken together, Conditions (1), (3a)–(3b), (4) and (5) close the feedback loop to qualify σ as an *equilibrium strategy*.

Further, we call a stable state under an equilibrium strategy σ an *equilibrium state*. Due to the Markovian property, any equilibrium state of the system will be preserved in equilibrium—that is, all of the ordinary customers will remain in the ordinary line—until either an arrival or a departure event triggers a state transition.

3.2 Analysis

The strategy space and the value-function space of the game are massive. To facilitate equilibrium analysis, we first propose three intuitive criteria on value functions. This enables us to focus on a reasonable subspace of $\Sigma \times \mathcal{V}$ that complies with these criteria.

Intuitive criteria

We propose three intuitive criteria on value functions as follows:

Criterion 1 $V(i + 1, j, k) \leq V(i, j, k)$.

Criterion 2 $V(i, j + 1, k) \leq V(i, j, k)$.

Criterion 3 $V(i - 1, j, k + 1) \leq V(i, j, k)$.

Criterion 1 states that, all else being equal, an ordinary customer does not receive a higher expected utility when there are more customers ahead of her in the ordinary line. Criterion 2 specifies that, all else being equal, an ordinary customer does not receive a higher expected utility when there are more customers behind her in the ordinary line. Criteria 1 and 2 are consistent with the intuition that a non-priority customer cannot be better off with a more congested system (corresponding to more potential competitors for priority). On the other hand, Criterion 3 states that for an ordinary customer, when the number of ordinary customers ahead of her in the ordinary line and the number of priority customers in the priority line add up to a constant, the customer (weakly) prefers the scenario with fewer priority customers. The intuition here is that any customer already in the priority line will be served before the tagged ordinary customer with certainty regardless of whether she upgrades, whereas a customer ahead of her in the ordinary line may not be served before her if she upgrades timely and the other person does not.

Roadmap

The rest of this section is devoted to proving that pure equilibrium strategies do not exist under the intuitive criteria. In particular, we proceed with the following two-step approach which is also illustrated by the corresponding Venn diagrams in Fig. 2:

Step 1 We establish that any pure-strategy equilibrium subject to the intuitive criteria must be of a state-dependent threshold type. In particular, combining the equilibrium conditions from Definition 1 (which characterize a subset of $\Sigma \times \mathcal{V}$, denoted by \mathcal{E}) with the three intuitive criteria (corresponding to set \mathcal{I}), we show that if a pure-strategy equilibrium exists, it must be of a state-dependent threshold type (the set of such threshold strategies is denoted by \mathcal{T}). That is, Step 1 shows that if $(\mathcal{E} \cap \mathcal{I}) \neq \emptyset$, then $(\mathcal{E} \cap \mathcal{I}) \subseteq \mathcal{T}$.

Step 2 We show that any state-dependent threshold strategy cannot be an equilibrium, i.e., $(\mathcal{E} \cap \mathcal{T}) = \emptyset$. The result implies $(\mathcal{E} \cap \mathcal{I}) \not\subseteq \mathcal{T}$, and by the contrapositive argument of Step 1, we can conclude that $(\mathcal{E} \cap \mathcal{I}) = \emptyset$. That is, there does not exist any pure-strategy equilibrium of our game that would satisfy the intuitive criteria. In other words, a pure-strategy equilibrium, even if it ever existed, would be peculiar and thus might be of little practical interest.

3.2.1 Step 1: equilibrium structure

In this subsection, we search for the equilibrium structure subject to Definition 1 and Criteria 1–3. Hence, in this step, we treat the three intuitive criteria as underlying

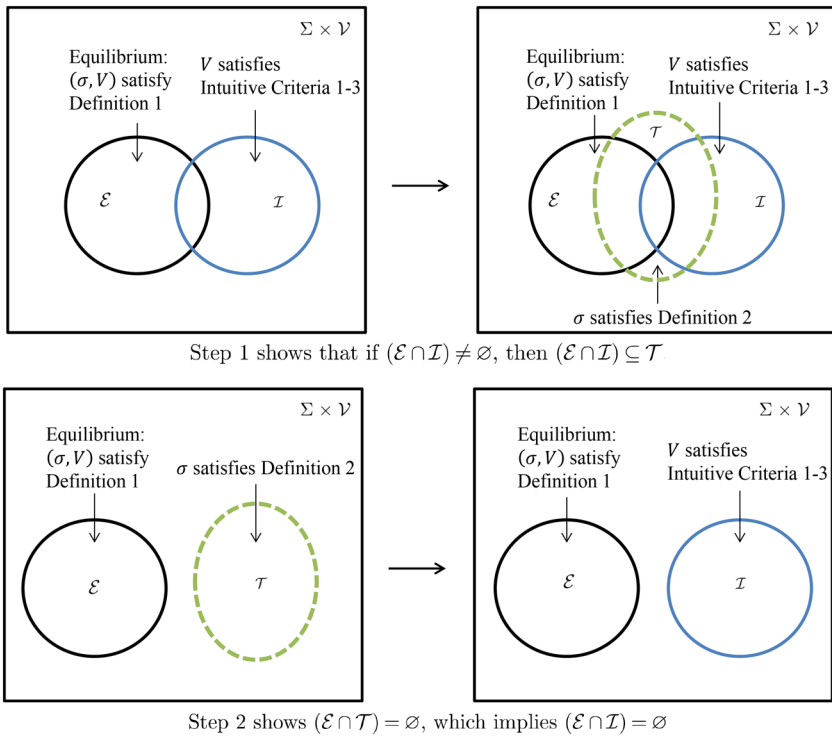


Fig. 2 Illustration of the analytical roadmap

assumptions for all the derivations. For brevity, we do not repeatedly state in each of the formal results in Step 1 their reliance on the three intuitive criteria. We develop some supporting lemmata first.

Lemma 1 *In equilibrium, under Criteria 1-3, for any system state $\{N_o, N_p\} = \{i + j, k\}$, if $\sigma(i, j, k) = N$, we must have $\sigma(i + s, j - s, k) = N$ for all $s \in \{1, 2, \dots, j\}$.*

Lemma 1 implies that under any equilibrium strategy, an ordinary customer does not purchase priority unless all other customers ahead of her in the ordinary line do. This is consistent with the first-come-first-upgrade rule whereby an ordinary customer with a more advanced (resp., backward) position in the ordinary line also receives a more advanced (resp., backward) position in the priority line, when the two of them decide to upgrade to priority at the same time.

Lemma 2 *In equilibrium, $V(i, j, k) - V(i, j, k + 1) \geq C/\mu$.*

Lemma 2 gives a lower bound for the difference between $V(i, j, k)$ and $V(i, j, k + 1)$. Intuitively, when there is one more priority customer in the line (all else being equal), an ordinary customer’s expected utility is reduced by at least the waiting cost of a service period, but could be more because waiting longer in the ordinary line (if she does not upgrade) further increases the likelihood of being overtaken (by other customers).

Lemma 3 *In equilibrium,*

- (i) if $\sigma(1, j, k) = N$, then $\sigma(1, j', k') = N$ for all $j' \in \{0, 1, \dots, j\}$ and $k' \in \{0, 1, \dots, k\}$;
- (ii) if $\sigma(1, j, k) = Y$, then $\sigma(1, j', k') = Y$ for all $j' \in \{j, j + 1, \dots\}$ and $k' \in \{k, k + 1, \dots\}$.

Lemma 3 suggests that the first customer in the ordinary line is more inclined to purchase priority when the lines are longer (either due to more customers waiting behind in the ordinary line or due to more customers waiting ahead in the priority line). Intuitively, a greater number of other customers standing in the ordinary line provides the first customer with a greater incentive to upgrade since their presence spells more competition for priority. More customers being in the priority line also motivates priority-purchasing because a longer waiting time caused by these priority customers implies that the first customer is likely to encounter more future arrivals who pose a threat as prospective competitors for priority.

Based on Definition 1 and Criteria 1–3 and using Lemmata 1–3 as stepping stones, we now establish the structure of pure-strategy equilibrium in the following Theorem 1. Note that it is a necessary condition for any equilibrium strategy under the three intuitive criteria.

Theorem 1 *A pure-strategy equilibrium under the three intuitive criteria must be of a state-dependent threshold type as specified by Definition 2.*

Definition 2 For a weakly-decreasing sequence of non-negative integers $\{n_s\}_{s \in \mathbb{N}_0}$ such that

$$n_0 = \infty \geq n_1 \geq n_2 \geq \dots \geq n_{\underline{m}-1} > n_{\underline{m}} = 0 = n_{(\underline{m}+1)} = n_{(\underline{m}+2)}, \dots, \quad (6)$$

with $\underline{m} \leq \left\lceil \frac{\mu P + C}{C} \right\rceil$, we define the following strategy σ : For any $N_o \in \mathbb{N}$ and $N_p \in \mathbb{N}_0$,

$$\sigma(i, N_o - i, N_p) = \begin{cases} Y & \text{if } i \in \{1, 2, \dots, \bar{n}_{\{N_o, N_p\}}\}, \\ N & \text{if } i \in \{\bar{n}_{\{N_o, N_p\}} + 1, \bar{n}_{\{N_o, N_p\}} + 2, \dots, N_o\}, \end{cases} \quad (7)$$

where

$$\bar{n}_{\{N_o, N_p\}} = \min\{s | N_p + s \leq n_{(N_o-s)} - 1, s \in \mathbb{N}_0\}. \quad (8)$$

Making sense of Theorem 1

Let us unpack Theorem 1. First, it states that given any state $\{N_o, N_p\}$, an equilibrium strategy is a state-dependent threshold strategy which specifies that the first and only the first $\bar{n}_{\{N_o, N_p\}}$ ordinary customers should purchase priority; see Eq. (7). Therefore, $\bar{n}_{\{N_o, N_p\}}$ also corresponds to the number of ordinary customers who purchase priority at state $\{N_o, N_p\}$ under an equilibrium strategy. Note that $\bar{n}_{\{N_o, N_p\}}$ is well-defined by Eq. (8) and $\bar{n}_{\{N_o, N_p\}} \leq N_o$ because N_o is an element that belongs to the set

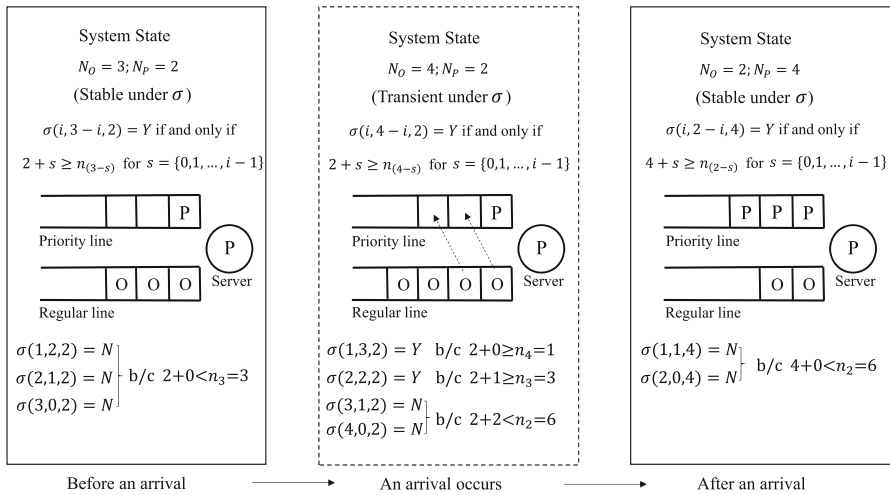


Fig. 3 System dynamics under equilibrium strategy σ specified by $n_0 = \infty, n_1 = \infty, n_2 = 6, n_3 = 3, n_4 = 1$ and $n_5 = n_6 = \dots = 0$. Note: See the discussion following Corollary 1

$\{s | N_p + s \leq n_{(N_o-s)} - 1, s \in \mathbb{N}_0\}$. Because only customers at the front of the ordinary line purchase priority, the FIFO service order is preserved.

Second, Theorem 1 indicates that $\bar{n}_{\{N_o, N_p\}}$ should be derived for any given $\{N_o, N_p\}$ from a weakly decreasing integer sequence $\{n_s\}_{s \in \mathbb{N}_0}$ that only depends on the system parameters but not the system state; see Eq. (8). It is easier to interpret the relationship between an equilibrium strategy and $\{n_s\}_{s \in \mathbb{N}_0}$ through the lens of an equivalent condition presented in the following Corollary 1.

Corollary 1 *In equilibrium, for any given system state $\{N_o, N_p\}$, we have $\sigma(i, N_o - i, N_p) = Y$ if and only if $N_p + s \geq n_{(N_o-s)}$ for $s \in \{0, 1, 2, \dots, i - 1\}$.*

Corollary 1 directly follows from Eqs. (7) and (8) in Theorem 1. To see how Corollary 1 is operationalized, we consider a specific example in Fig. 3. In this example, let us suppose σ is an equilibrium strategy specified by $n_0 = \infty, n_1 = \infty, n_2 = 6, n_3 = 3, n_4 = 1$ and $n_5 = n_6 = \dots = 0$ (which satisfies the descending-order requirement by Theorem 1). At the initial given state $\{3, 2\}$ (left panel of Fig. 3), we know by Corollary 1 that $\sigma(i, 3 - i, 2) = Y$ if and only if $2 + s \geq n_{(3-s)}$ for $s = \{0, 1, \dots, i - 1\}$. This implies $\sigma(1, 2, 2) = \sigma(2, 1, 2) = \sigma(3, 0, 2) = N$ which further implies that the state $\{3, 2\}$ is a stable state under strategy σ . Next, suppose an arrival event occurs which brings the system state to $\{4, 2\}$ (middle panel of Fig. 3). By applying Corollary 1 to the new state $\{4, 2\}$, we know that $\sigma(i, 4 - i, 2) = Y$ if and only if $2 + s \geq n_{(4-s)}$ for $s = \{0, 1, \dots, i - 1\}$. It then follows that $\sigma(1, 3, 2) = \sigma(2, 2, 2) = Y$ and $\sigma(3, 1, 2) = \sigma(4, 0, 2) = N$, meaning the equilibrium strategy σ prescribes that the first two ordinary customers purchase priority at state $\{4, 2\}$. This results in a system transition to state in $\{2, 4\}$ (right panel of Fig. 3), and the system is back to a stable state under σ based on Corollary 1 again. In sum, the first two customers in the ordinary line in this example originally do not purchase priority, but later choose to do so when a new customer arrives.

The second implication of Theorem 1 is the following Corollary 2.

Corollary 2 *In equilibrium, the occurrence of a departure event in a stable state will not trigger any customer to purchase priority.*

Corollary 2 follows directly from the condition in Corollary 1 because by plugging $s = 0$, we know that a system state is stable if and only if $N_p < n_{N_o}$. When a departure event occurs to a stable state, i.e., N_p is reduced by 1 if $N_p \geq 1$ or N_o is reduced by 1 if $N_p = 0$, the condition $N_p < n_{N_o}$ is preserved because n_{N_o} is weakly decreasing in N_o (Theorem 1). Intuitively, upon a departure, all customers expect less cost into the future because they are one step closer to service completion. An ordinary customer in a more backward position of the ordinary line expects a larger cost reduction than one in a more forward position because the one in the back is more vulnerable to being overtaken (because they would stay in the system longer, inviting more overtaking threats from future arrivals). Hence, a departure event attenuates the threat from future arrivals and reduces customers' incentives to purchase priority, particularly for a customer in the back of the ordinary line. Since those standing in the back do not purchase priority upon a departure, those in the front are assured of their positions and thus also have no incentive to purchase. Consequently, all ordinary customers would stay put in the event of a departure.

3.2.2 Step 2: non-existence of pure-strategy equilibria

Thus far, Theorem 1, together with Corollaries 1 and 2, has presented a clear view of the equilibrium structure—if equilibrium strategies exist. To characterize the equilibrium strategies under Criteria 1-3, it is sufficient to focus on the state-dependent threshold strategies in Definition 2. In this step, we now prove that any pure strategy given by Definition 2 cannot be sustained in equilibrium.

Theorem 2 *Any pure strategy defined by Definition 2 cannot be sustained in equilibrium. That is, $(\mathcal{E} \cap \mathcal{T}) = \emptyset$.*

Combining Theorems 1 and 2 yields the following corollary.

Corollary 3 *Under the simultaneous upgrade rule, pure-strategy equilibria do not exist if the intuitive criteria hold. That is, $(\mathcal{E} \cap \mathcal{T}) = \emptyset$.*

We will provide some intuition for why pure-strategy equilibria do not exist through the lens of a tractable small buffer system that has a buffer size $K = 2$, i.e., a system that can hold at most two customers (including the customer in service, if any). A firm understanding of this simplified setting will shed light on our main model that does not have a buffer limit (i.e., $K = \infty$).

3.3 A small buffer system

In this subsection, we analyze a small buffer system with $K = 2$ to sharpen intuition. For notational convenience, we define traffic intensity $\rho \triangleq \lambda/\mu$ and normalized price $v_P \triangleq \mu P/c$. Since a small buffer system is always stable, we relax the assumption $\lambda < \mu$, i.e., traffic intensity ρ can be less than or greater than 1.

Theorem 3 *Consider a small buffer system with $K = 2$ subject to the simultaneous upgrade rule. If $v_P > 1$, then no customers purchasing priority is an equilibrium. Otherwise, i.e., if $v_P \leq 1$, there exist no pure-strategy equilibria.*

Theorem 3 shows that in a small buffer system, no customers purchasing priority is an equilibrium strategy for the customers when the priority price is high ($v_P > 1$). This is intuitive because customers would never purchase priority if the priority price is higher than the maximum expected reduction in delay cost. More importantly, when the priority price is not too high ($v_P \leq 1$), a pure-strategy equilibrium does not exist (note that this result does not rely on the intuitive criteria as an assumption). In this case, customers have conflicting interests: a customer that is ahead in the queue only has an incentive to upgrade if the customer behind her upgrades, whereas the customer that is behind in the queue only gains from upgrading if the customer ahead does not upgrade.

When buffer size K increases, “no customers purchasing priority” will be sustained in equilibrium for any $v_P > K - 1$. On the other hand, if $v_P \leq K - 1$, we conjecture that no pure-strategy equilibria exist. Note that as K gets larger, the condition $v_P > K - 1$ becomes more difficult to satisfy. When $K \uparrow \infty$, we conjecture that there will not be any pure-strategy equilibrium for any $P > 0$ (we have proved this under the caveat of three intuitive criteria in Corollary 3).

The non-existence of pure-strategy equilibria implies that the simultaneous upgrade rule may be troublesome to implement. In addition, the simultaneous rule requires real-time updates of system states without delay (as all the customers move at the same time), which may also pose implementation challenges.

4 Sequential upgrade rule

In this section, we consider an alternative, sequential upgrade rule specified as follows: Each time a new customer arrives, the service provider sequentially asks each ordinary customer one at a time whether they wish to purchase priority, with the first customer in line being called upon first, then the second one, and so on, until all the ordinary customers are asked. Unlike the simultaneous upgrade rule, customers in this case cannot purchase priority until it is their turn. They have a single opportunity to upgrade each time a new customer arrives and if they choose not to upgrade this time, they must wait until the next customer arrival for a new upgrade opportunity. The sequential upgrade rule can be implemented with the aid of today’s mobile Internet, which may efficiently automate the inquiry and upgrade process.

4.1 Equilibrium definition

We set up the in-queue priority-purchasing problem under the sequential rule as a dynamic game. Consider an ordinary customer whose position vector is (i, j, k) , where i is the number of ordinary customers ahead of and including her; j , the number of ordinary customers behind her, and k , the number of priority customers. When it is her turn to upgrade, we define her (Markovian) strategy to be $\sigma(i, j, k) \in \{Y, N\}$, where

“ Y ” indicates “upgrade” (or purchase priority) and “ N ” indicates “no.” Note that in most sequential games, it is necessary to specify a player’s strategy as a function of the possible strategies of all the other players who move before she does (which would be cumbersome). However, in our case, this information is subsumed by a customer’s position vector when she is about to decide on upgrading. In other words, what matters to a customer is not how she reaches her position vector, but what the position vector is at the moment of decision-making. Therefore, for our purpose, it suffices to condition customer strategies only on the position vectors, instead of tracking the history of previous customers’ moves.

Next, we define the value function. Let $V_x(i, j, k)$ denote the expected utility of a customer with position vector (i, j, k) when the x th ordinary customer is about to decide on upgrading, for $x = 1, \dots, i + j + 1$. Specifically, $V_i(i, j, k)$ is the expected utility of customer (i, j, k) when it is her turn to upgrade (right before she acts); $V_{i+j+1}(i, j, k)$ is the expected utility of customer (i, j, k) when the upgrade process is complete (i.e., after all the ordinary customers have been asked).

Definition 3 A symmetric pure-strategy equilibrium under the sequential upgrade rule is characterized by any strategy and value-function pair (σ, V) that satisfies Conditions (9) through (13):

$$V_x(i, j, k) = \begin{cases} V_x(i - 1, j, k + 1), & \text{if } \sigma(x, i + j - x, k) = Y, \\ V_{x+1}(i, j, k), & \text{otherwise,} \end{cases} \quad \forall x = 1, \dots, i - 1, i \geq 2, (j, k) \in \mathbb{N}_0 \times \mathbb{N}_0, \tag{9}$$

$$V_i(i, j, k) = \begin{cases} -\frac{c(k+1)}{\mu} - P, & \text{if } \sigma(i, j, k) = Y, \\ V_{i+1}(i, j, k), & \text{otherwise,} \end{cases} \quad \forall (i, j, k) \in \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}_0, \tag{10}$$

$$V_x(i, j, k) = \begin{cases} V_x(i, j - 1, k + 1), & \text{if } \sigma(x, i + j - x, k) = Y, \\ V_{x+1}(i, j, k), & \text{otherwise,} \end{cases} \quad \forall x = i + 1 \dots, i + j, (i, j, k) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}_0, \tag{11}$$

$$V_{i+j+1}(i, j, k) = -\frac{C}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} V_{i+j+1}(i, j, k - 1) + \frac{\lambda}{\lambda + \mu} V_1(i, j + 1, k), \quad \forall (i, j, k) \in \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}, \tag{12a}$$

$$V_{i+j+1}(i, j, 0) = -\frac{C}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} V_{i+j}(i - 1, j, 0) + \frac{\lambda}{\lambda + \mu} V_1(i, j + 1, 0), \quad \forall (i, j) \in \mathbb{N} \times \mathbb{N}_0, \tag{12b}$$

$$V_{j+1}(0, j, 0) \equiv 0, \quad \forall j \in \mathbb{N}_0, \tag{12c}$$

$$V_i(i, j, k) = \max \left\{ -\frac{c(k + 1)}{\mu} - P, V_{i+1}(i, j, k) \right\}, \quad \forall (i, j, k) \in \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}_0. \tag{13}$$

In Definition 3, Conditions (9) specify the state transition due to the action of the x th ordinary customer who moves before customer (i, j, k) does. If the x th ordinary customer upgrades, then customer (i, j, k) 's position vector becomes $(i - 1, j, k + 1)$ (because now there is one fewer ordinary customer ahead of her but one more customer in the priority line). Note that the next customer who gets to choose whether to upgrade is still the x th customer in the ordinary line as the original x th customer joins the priority line. If the x th ordinary customer does not upgrade, then customer (i, j, k) 's position vector is unchanged before the next customer upgrades, and now the next customer will be the $(x + 1)$ th customer in the ordinary line. Conditions (10) specify the state transition due to customer (i, j, k) 's action. Conditions (11) specify the state transition due to the action of a customer who moves after customer (i, j, k) does. Conditions (12a) through (12c) specify the state transition due to arrivals and departures. Note that a new arrival triggers a new round of priority upgrades starting from the first customer in the ordinary line, and therefore the subscript of the value function is reset to 1. Conditions (13) ensure that customers indeed maximize their expected utility by choosing strategy σ provided that other customers choose σ . While Definition 3 applies to a system without a buffer limit, we can easily modify Conditions (12a) through (12c) by imposing boundary conditions that accommodate any finite buffer K .

Given the complexity of the problem, it is challenging to solve for the equilibrium in general, but we can nevertheless analytically characterize the equilibrium both under sufficiently light traffic and sufficiently heavy traffic, which will shed some light on the equilibrium structure in general.

4.2 Sufficiently light or heavy traffic

We first examine a case of sufficiently light traffic in which arrival rate λ is much smaller than service rate μ (i.e., $\lambda \ll \mu$) such that customers do not need to concern themselves with future arrivals that are not yet present in the system (because $\lambda \ll \mu$ implies that any existing customers in the system will be served long before any future customers arrive). We first define a threshold-type “ X -strategy” below in Definition 4.

Definition 4 (*X-Strategy*) Whenever the number of ordinary customers reaches $X + 1$, the X th customer (or equivalently the second last customer) and only that customer upgrades.

Theorem 4 below shows the X -strategy will arise in equilibrium under sufficiently low traffic.

Theorem 4 When $\lambda \ll \mu$, the X -strategy as defined in Definition 4 with $X = \lceil v_P \rceil$ is a pure-strategy equilibrium, where $v_P = \mu P / C$.

Under sufficiently light traffic ($\lambda \ll \mu$), customers compete only with existing customers and can act myopically without loss of optimality. In essence, the underlying game reduces to a one-shot sequential game. To understand the equilibrium in Theorem 4, consider the example of $v_P = 1.5$. In this case, $\lceil v_P \rceil = 2$, which implies that the maximum stable ordinary-queue length is 2. As soon as a new customer arrives,

the number of customers in the ordinary queue temporarily reaches 3, which triggers the upgrade of the second customer (while the first and third customers stay put), and the newcomer becomes the one to occupy the second position. In this simple case, it is easy to verify the equilibrium. Customer 1 loses C/μ by being overtaken, which is better than paying $P = 1.5C/\mu$ to secure her position. Given that Customer 1 does not upgrade, Customer 2 loses $0.5C/\mu$ by upgrading (paying $1.5C/\mu$ and overtaking one customer), which is better than losing C/μ by being overtaken by Customer 3 (Customer 3 would upgrade if neither Customers 1 nor 2 were to upgrade). Given that Customer 1 stays put and Customer 2 upgrades, Customer 3 is better off not upgrading as she would incur a net cost of $0.5C/\mu$ if he were to upgrade (overtaking Customer 1 and paying $1.5C/\mu$).

Under sufficiently light traffic, the maximum stable queue length of ordinary customers will not exceed $\lceil v_P \rceil$ in equilibrium (it may temporarily reach $\lceil v_P \rceil + 1$, but will immediately return to $\lceil v_P \rceil$ upon an upgrade). When the traffic intensity is higher, the number of customers that can possibly accumulate in the ordinary queue before an upgrade occurs will only (weakly) decrease as customers are under greater pressure to defend their position. We summarize this (immediate) result below.

Corollary 4 *Under any traffic intensity, in equilibrium, the maximum stable queue length of ordinary customers cannot exceed $\lceil v_P \rceil$.*

Intuitively, as the system experiences more traffic, we would expect the maximum stable queue length of ordinary customers to decline. In other words, an upgrade is easier to trigger with more traffic. Under sufficiently heavy traffic, the maximum stable queue length may reduce to 1. Theorem 5 below gives precise conditions on when this equilibrium emerges.

Theorem 5 *If and only if $\rho \geq \sqrt{v_P} - 1$, the X -strategy as defined in Definition 4 with $X = 1$ is a pure-strategy equilibrium, i.e., in this equilibrium, the first and only the first customer upgrades whenever the number of ordinary customers reaches 2.*

Theorem 5 shows that under sufficiently heavy traffic, an equilibrium arises in which as soon as two ordinary customers are present, the first customer upgrades. Note that if the normalized priority price is low, i.e. $v_P \leq 1$, the condition $\rho \geq \sqrt{v_P} - 1$ will in fact be satisfied by any traffic intensity ρ , i.e., the aforementioned equilibrium strategy holds for any traffic intensity. In this case, since $\lceil v_P \rceil = 1$, the equilibrium strategy is the same as the one identified in Theorem 4. Nevertheless, for $v_P \in (1, 4)$, combining Theorems 4 and 5 indicates that the equilibrium strategy under sufficiently heavy traffic (when forward-looking customers must take into account future arrivals) disagrees with the equilibrium strategy under sufficiently light traffic (when customers can act myopically). Although the exact equilibrium strategy differs, the equilibrium structure is still the same, both belonging to the families of X -strategies.

This leaves us with the case of $v_P \geq 4$. In this case, the condition $\rho \geq \sqrt{v_P} - 1$ cannot be satisfied by any traffic intensity $\rho < 1$. Thus, the X -strategy with $X = 1$ is no longer an equilibrium. Intuitively, a high priority price discourages the first customer from purchasing priority when the second customer arrives. While the actual equilibrium eludes us, we show in Theorem 6 below that under sufficiently heavy traffic, the equilibrium structure will be different from that of X -strategies.

Theorem 6 *If $v_P \geq 4$ and $\rho > v_P/(v_P + 1)$, any X -strategy as defined in Definition 4 is not an equilibrium.*

When both the traffic intensity and the priority price are high, we have already explained an X -strategy with $X = 1$ cannot be an equilibrium; any X -strategy with $X \geq 2$ cannot be an equilibrium either because before the second last customer upgrades, the customers ahead of her would already have an incentive to purchase priority in order to preempt competition from future arrivals. Hence, the equilibrium structure is likely to be more complex. Note that, for any given v_P , we can always satisfy the condition $\rho > v_P/(v_P + 1)$ in Theorem 6 with sufficiently high traffic intensity ρ while still guaranteeing stability ($\rho < 1$).

4.3 Small buffer systems

In this subsection, we investigate customers' equilibrium strategies in small buffer systems to sharpen intuition. Specifically, we will first analytically study a queueing system that can hold at most two customers (i.e., $K = 2$), similar to what we did for the simultaneous upgrade rule (see Sect. 3.3). We will then numerically study a queueing system that can hold at most three customers (i.e., $K = 3$).

4.3.1 $K = 2$

If a single ordinary customer is present in the queue, it is straightforward that the customer will not upgrade. Thus, since $K = 2$, specifying customer strategies when there are two ordinary customers will pin down the equilibrium. Theorem 7 below characterizes the pure-strategy equilibrium through a tuple in which the first (resp., second) element indicates the first (resp., second) ordinary customer's strategy.

Theorem 7 *Consider a small buffer system with $K = 2$ subject to the sequential upgrade rule. The unique pure-strategy equilibrium is (Y, N) if $v_P \leq 1$ and (N, N) otherwise.*

When the priority price is high ($v_P > 1$), it is intuitive that no customers purchase priority. However, unlike the simultaneous case (as shown in Theorem 3), when the priority price is not too high ($v_P \leq 1$), a pure-strategy equilibrium exists, and in equilibrium, the first customer upgrades while the second customer does not. Under the sequential upgrade rule, the first customer gets to decide first. Should she upgrade, she would lose P ; should she not upgrade, the second customer would upgrade, causing the first customer to lose C/μ . Since $P \leq C/\mu$ (or equivalently $v_P \leq 1$), the first customer is better off not upgrading, which, in turn, removes the incentive for the second customer to upgrade.

Note that while the second customer chooses to pass on priority purchasing at the moment of her arrival, she may eventually upgrade if she moves to the first spot and a new customer arrives before she completes service. Interestingly, the equilibrium strategy here does not vary with ρ , suggesting that customers act as if they were myopic, ignoring future arrivals. We caution that this equivalence between myopic

and forward-looking behaviors is driven by the assumption $K = 2$, and does not hold in general (in fact, it will break when $K = 3$, as we will see in Sect. 4.3.2). Based on the equilibrium strategy given in Theorem 7, we further derive the priority revenue in Corollary 5 below.

Corollary 5 Consider a small buffer system with $K = 2$ subject to the sequential upgrade rule. The priority revenue per unit time as a function of v_P is given by

$$\Pi_{K=2}^{SQ}(v_P) = \begin{cases} \frac{C\rho^2 v_P}{1+\rho+\rho^2}, & \text{if } v_P \leq 1, \\ 0, & \text{if } v_P > 1. \end{cases}$$

Next, we compare the priority revenue generated from in-queue priority purchase subject to the sequential upgrade rule with that from the classical upon-arrival priority purchase under $K = 2$. To be clear, in the upon-arrival priority purchase scheme, customers are presented with the option to purchase priority only upon arrival and if they choose not to purchase the moment they arrive, they cannot do so later. The literature (for example, [1, 15]) has shown that customers in this case follow a threshold strategy in equilibrium: they purchase if and only if the queue length they see upon arrival reaches a certain threshold. Hassin and Haviv [15] have also shown the possibility of multiple equilibria. Whenever multiple equilibria arise, we select the *Pareto-dominant* equilibrium in which all customers have higher expected utility than they would in other equilibria. Note that the Pareto-dominant equilibrium effectively corresponds to the one with the highest threshold. Proposition 1 below characterizes the equilibrium and the corresponding priority revenue when customers can purchase priority only upon arrival.

Proposition 1 Consider a small buffer system with $K = 2$ where customers can purchase priority only upon arrival. The unique Pareto-dominant pure-strategy equilibrium strategy is given below.

- If $v_P \leq \min\{\rho, 1\}$, then all customers purchase priority.
- If $\rho < v_P < 1$, then a customer purchases priority if and only if she sees one (ordinary) customer in the system upon arrival.
- If $v_P \geq 1$, then no customers purchase priority.

The priority revenue per unit time as a function of v_P is given by

$$\Pi_{K=2}^{ARR}(v_P) = \begin{cases} \frac{C\rho(1+\rho)v_P}{1+\rho+\rho^2}, & \text{if } v_P \leq \min\{\rho, 1\}, \\ \frac{C\rho^2 v_P}{1+\rho+\rho^2}, & \text{if } \rho < v_P < 1, \\ 0, & \text{if } v_P \geq 1. \end{cases}$$

Next, we compare the maximum revenue (by optimizing over the priority price) of upon-arrival priority purchase, $\Pi_{K=2}^{ARR}$, and that of in-queue priority purchase subject to the sequential rule, $\Pi_{K=2}^{SQ}$, where

$$\Pi_{K=2}^{ARR} = \max_{v_P} \Pi_{K=2}^{ARR}(v_P), \quad \Pi_{K=2}^{SQ} = \max_{v_P} \Pi_{K=2}^{SQ}(v_P).$$

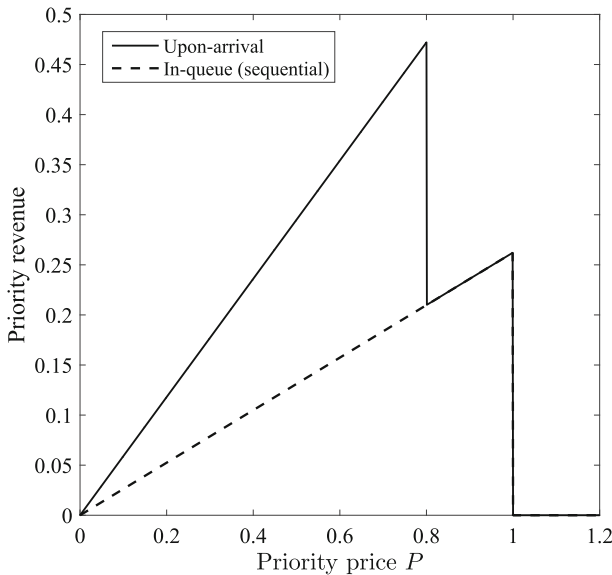


Fig. 4 Comparison of priority revenues in a small buffer system with $K = 2$. Note. $\mu = 1$, $C = 1$ and $\lambda = 0.8$

Theorem 8 Consider a small buffer system with $K = 2$. The optimal revenue of upon-arrival priority purchase is higher than that of in-queue priority purchase subject to the sequential rule, i.e., $\Pi_{K=2}^{ARR} > \Pi_{K=2}^{SQ}$.

Theorem 8 shows that in a small buffer system with $K = 2$, the upon-arrival priority purchase scheme yields higher revenue than in-queue priority purchase, although the latter presents customers with more upgrade opportunities. See Fig. 4 for an illustration. Recall that in-queue priority purchase lessens customers’ fear of being overtaken if they do not purchase priority upon arrival by allowing them to defer their purchase decision. As a result, customers have less of an incentive to purchase than they would in the upon-arrival case, causing the in-queue purchase scheme to fall short of the upon-arrival purchase scheme in terms of priority revenue.

More specifically, in the small buffer system, to maximize the priority revenue in the upon-arrival purchase scheme, the service provider should always charge a low enough price to induce all customers to purchase priority upon arrival. In particular, even those who arrive at an empty system will purchase immediately in order to secure their current position. By contrast, in the sequential in-queue purchase scheme, customers who arrive at an empty system will never have a motivation to purchase (as they can always wait until a new customer arrives), which explains the revenue gap between the upon-arrival and in-queue purchase schemes. Note that in the former, the Pareto-dominant equilibrium we select in the event of multiple equilibria represents the most advantageous outcome to customers but the least advantageous as far as priority revenue is concerned. It implies that the priority revenue of the upon-arrival purchase scheme we use for comparison is the most conservative prediction of what

Table 1 The equilibrium strategies of sequential in-queue priority purchase under various prices for $K = 3$

Priority price	$P = 1.2$	$P = 1.5$	$P = 1.8$	$P = 2.1$
$\sigma(1, 1, 0)$	Y	N	N	N
$\sigma(2, 0, 0)$	Y	N	N	N
$\sigma(1, 2, 0)$	Y	Y	N	N
$\sigma(2, 1, 0)$	Y	Y	Y	N
$\sigma(3, 0, 0)$	Y	Y	Y	N
$\sigma(1, 1, 1)$	Y	N	N	N
$\sigma(2, 0, 1)$	Y	N	N	N
Type	(1)	(2)	(3)	(4)

$\mu = 1, C = 1, \lambda = 0.8$

may arise in equilibrium, but even this conservative prediction surpasses the revenue of the in-queue purchase scheme.

While in-queue priority purchase creates more selling opportunities for the service provider, Theorem 8 tells a cautionary tale against this practice. It implies that the service provider may instead benefit from giving customers a *buy-it-or-lose-it* ultimatum when they arrive.

4.3.2 $K = 3$

As a proof of concept, we numerically solve for the pure-strategy equilibrium of sequential in-queue priority purchase in a system that can hold at most three customers (i.e., $K = 3$). In each numerical instance, we enumerate all the possible strategies σ ; for each given σ , we solve a modified version of Conditions (9) through (12c) that accommodates the finite-buffer system of $K = 3$ (a system of linear equations) for the corresponding the value function V ; we then check if V satisfies (13); if so, the (σ, V) pair is an equilibrium. In all the numerical instances tested, we have consistently found a unique equilibrium.

Tables 1 reports the equilibrium strategies under various priority prices for a fixed arrival rate; Table 2 reports the equilibrium strategies under various arrival rates for a fixed priority price. For brevity, we only report equilibrium strategies when two or three ordinary customers are present (because, as argued earlier, if only one ordinary customer is present, the customer will trivially not upgrade).

We observe four types of possible equilibria:

- (1) Whenever the ordinary queue length reaches 2, the first customer upgrades.
- (2) Whenever the ordinary queue length reaches 3, the first customer upgrades.
- (3) Whenever the ordinary queue length reaches 3, the second customer upgrades.
- (4) No customers upgrade.

We make the following observations from Table 1. First, as the priority price increases, customers are less prone to upgrade (in the sense that we see fewer Y s and more N s), which is intuitive. Second, many of the strategies are defined on position vectors off the equilibrium path but are nevertheless critical to governing equilibrium behavior. For instance, for $P = 1.2$, the equilibrium is of type (1), which implies that in

Table 2 The equilibrium strategies of sequential in-queue priority purchase under various arrival rates for $K = 3$

Arrival rate	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$
$\sigma(1, 1, 0)$	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>Y</i>
$\sigma(2, 0, 0)$	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>Y</i>
$\sigma(1, 2, 0)$	<i>N</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
$\sigma(2, 1, 0)$	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
$\sigma(3, 0, 0)$	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
$\sigma(1, 1, 1)$	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>Y</i>
$\sigma(2, 0, 1)$	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>Y</i>
Type	(3)	(2)	(2)	(2)	(1)

$\mu = 1, C = 1, P = 1.5$

equilibrium the system will never reach a state where the first ordinary customer does not upgrade, i.e., a customer cannot be at position $(2, 0, 0)$ when making a decision, but should that occur, $\sigma(2, 0, 0) = Y$ implies the second customer would upgrade. Third, for both $P = 1.5$ and $P = 1.8$, the equilibrium structure is that the arrival of a third ordinary customer triggers a priority purchase, but the two cases differ in who is the purchasing customer. When $P = 1.5$, the priority price is low enough to make the first customer purchase in order to avoid being overtaken. When $P = 1.8$, the priority price is high enough to convince the first customer that not purchasing and being overtaken is a better choice; seeing the first customer not upgrade, the second customer upgrades and jumps ahead of the first one.

We make the following observations from Table 2. First, as the arrival rate (and thus traffic intensity) increases, customers are more prone to upgrade (in the sense that we see more *Y*s and fewer *N*s). This contrasts the case of $K = 2$, in which the equilibrium strategy does not vary with traffic (see Theorem 7), but makes intuitive sense because a higher traffic intensity implies future arrivals pose a greater threat, which warrants more preemptive upgrades. Second, when $\lambda = 0.5$ (a case of light traffic), the equilibrium behavior is of type (3) and indeed agrees with the one identified in Theorem 4 for sufficiently light traffic. Third, when the traffic intensity is higher ($\lambda = 0.6, 0.7, 0.8, 0.9$), the equilibrium behavior departs from what occurs under light traffic. In particular, when $\lambda = 0.6, 0.7, 0.8$, the equilibrium is of type (2) and it is not the second (to the last) customer but the first customer who upgrades. When $\lambda = 0.9$, the equilibrium is of type (1) and the queue length only needs to reach 2, instead of 3, to trigger an upgrade. Fourth, the number of priority customers (0 or 1) does not affect the equilibrium strategy, i.e., $\sigma(1, 1, 0) = \sigma(1, 1, 1)$ and $\sigma(2, 0, 0) = \sigma(2, 0, 1)$. In fact, we can analytically prove this result, as shown in Theorem 9.

Theorem 9 Consider a small buffer system with $K = 3$ subject to the sequential upgrade rule. In a pure-strategy equilibrium, $\sigma(1, 1, 0) = \sigma(1, 1, 1)$ and $\sigma(2, 0, 0) = \sigma(2, 0, 1)$.

We caution that the result of Theorem 9 (the equilibrium strategy being independent of the number of priority customers) is likely driven by the assumption $K = 3$ and is

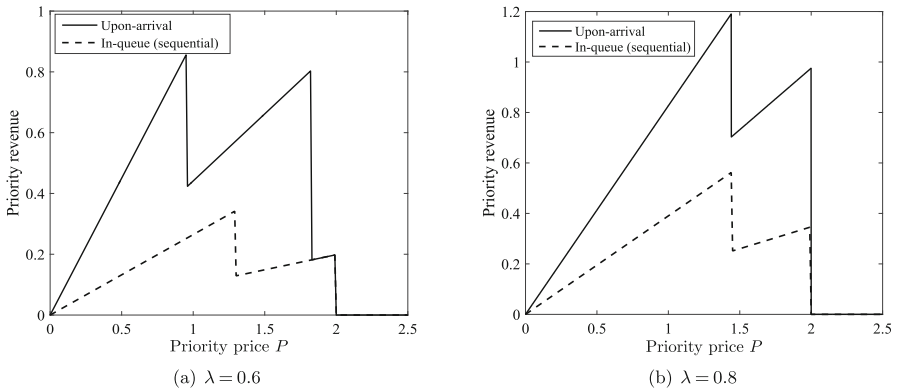


Fig. 5 Comparison of priority revenues for $K = 3$. Note: $\mu = 1$, $C = 1$

not meant to be interpreted as a general result that applies to systems with any buffer size. What is special about $K = 3$ is that if there is one priority customer and two ordinary customers in the system, then the system is already full. As a result, the two ordinary customers would not be much concerned about future arrivals (because they know the next event can only be a departure). Hence, their strategy in equilibrium is no different than if the priority customer did not exist. However, this reasoning would not generalize to a system with a larger buffer and it is reasonable to conjecture that ordinary customers would act differently should the number of priority customers in the system differ.

We also numerically compare the priority revenue generated in sequential in-queue priority purchase with that in the upon-arrival purchasing model. The results are reported in Fig. 5. We observe that in the system of $K = 3$, sequential in-queue priority purchase still generates less revenue than upon-arrival priority purchase (for a given priority price and under the optimal priority price), which parallels the analytical insight gleaned from the small buffer system with $K = 2$ (see Theorem 8).

5 Concluding remarks

The extant priority-purchasing literature has restricted attention to the case where customers who would like to purchase priority must do so upon arrival to the service system. A natural alternative is to allow customers to upgrade to priority at any time after they have joined the (non-priority) line. Under such circumstances, little, if any, is known about how self-interested customers would behave. Our paper seeks to fill this gap by formulating a dynamic game that models customers' in-queue priority-purchasing behavior.

When the simultaneous upgrade rule is imposed, we find that pure-strategy equilibria do not exist under certain intuitive criteria, contrasting the extant priority-purchasing literature, which instead shows the existence and sometimes multiplicity of pure-strategy equilibria when customers can only purchase priority upon arrival.

However, when the sequential upgrade rule is implemented, pure-strategy equilibria may exist. The upgrading behavior can be complex. Under sufficiently light traffic, if the number of ordinary customers accumulates to a certain threshold, then it is always the second last customer who upgrades, but in general, it could be a customer from another position, and the queue-length threshold that triggers an upgrade can also vary with the traffic intensity. Under sufficiently heavy traffic and a not-too-high priority price, as soon as two ordinary customers gather, the first customer will upgrade. Our analytical and numerical results on systems with relatively small buffers consistently show that in-queue priority purchase does not yield as much revenue as upon-arrival priority purchase.

To sum up, in priority purchasing, the fundamental conflict among customers is that those in the front of the queue would like to upgrade only because they fear that those in the back would upgrade, but those in the back would only gain from upgrading if customers in the front do not upgrade. The conflict is difficult to resolve when customers try to upgrade at the same time, which leads to the potential non-existence of pure-strategy equilibria under the simultaneous upgrade rule. However, if customers make upgrade decisions one at a time, then those who decide later can adjust their decisions to what earlier customers do, which facilitates the resolution of the aforementioned conflict. Thus, under the sequential upgrade rule, it is more promising for customers to reach pure-strategy equilibria. Moreover, the tension among customers escalates if a sufficient number of customers congregate and if the threat from future arrivals intensifies, which explains why a threshold-type equilibrium strategy might emerge (i.e., an upgrade is triggered only if the number of ordinary customers reaches a certain threshold), and why this threshold can decrease with the traffic intensity.

Finally, the non-equilibria in the simultaneous case is also partially attributed to the assumption of homogeneous customers. If customers are heterogeneous in their delay sensitivity, then the aforementioned conflict can be somewhat alleviated as those who are more sensitive to waiting are more eager to purchase priority. If delay sensitivity differs vastly, then customers' priority purchasing decisions will be mostly driven by the differences in their delay sensitivity rather than the differences in their relative positions in the queue.

Acknowledgements The authors thank the guest editors and two anonymous reviewers for their constructive comments that have significantly improved the paper. Zhongbin Wang is supported in part by the National Natural Science Foundation of China [Grant 72001118] and the China Postdoctoral Science Foundation [Grant 2020M670635]. Jinting Wang is supported in part by the National Natural Science Foundation of China [Grant 71871008].

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

For the proofs, define V_σ as the value function determined by Condition (1) of Definition 1 for a given strategy σ . To prove Lemmas 1–3, we first provide the following two lemmas.

Lemma A.1 $-P - (i + k)C/\mu \leq V(i, j, k) \leq -kC/\mu, \forall (i, j, k) \in \mathbb{N} \times \mathbb{N}_0 \times \mathbb{N}_0$.

In the result above, the upper bound $-kC/\mu$ corresponds to the minimum expected waiting cost of the tagged customer. On the other hand, an ordinary customer at position (i, j, k) always has the option to upgrade to priority by paying premium P . Therefore, $-P - (i + k)C/\mu$ serves as a lower bound for $V(i, j, k)$ because this customer's expected waiting cost with the priority upgrade is at most $(i + k)C/\mu$, which is evident from Eq. (5).

Lemma A.2 *If strategy σ is an equilibrium strategy, and $\{i + j, k\}$ is a transient state under σ , then the new state $\{i + j - \sum_{s=1}^{i+j} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}}, k + \sum_{s=1}^{i+j} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}}\}$ to which the system transitions under σ , is stable under σ , which is then by definition an equilibrium state.*

Proof Suppose the new system state

$\{i + j - \sum_{s=1}^{i+j} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}}, k + \sum_{s=1}^{i+j} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}}\}$ is not a stable state under σ , then without loss of generality (WLOG) there exists some customer with position (i', j', k') at the new state whose equilibrium strategy specifies that $\sigma(i', j', k') = Y$, where $i' + j' = i + j - \sum_{s=1}^{i+j} \mathbb{1}_{\{\sigma(s, i+j-s, k)=Y\}}$. Therefore, her maximum expected utility

$$\begin{aligned} V_\sigma(i', j', k') &= -P - \left[k' + \sum_{s=1}^{i'-1} \mathbb{1}_{\{\sigma(s, i'+j'-s, k)=Y\}} + 1 \right] C/\mu \\ &\leq -P - (k' + 1)C/\mu. \end{aligned} \quad (\text{A.1})$$

WLOG, suppose the customer's position was (i, j, k) at the previous transient system state $\{i + j, k\}$. It is clear that $\sigma(i, j, k) = N$ because otherwise the customer will not be in the ordinary line at the new system state. We can infer that there were $i - i' \geq 0$ and $j - j' \geq 0$ ordinary customers ahead of and behind her, respectively, at the old state $\{i + j, k\}$, who have purchased priority according to the strategy σ . Because $\sigma(i, j, k) = N$, we have $V_\sigma(i', j', k') = V_\sigma(i, j, k) > -P - [k + (i - i') + 1]C/\mu$ [from (5)] $\geq -P - [k + (i - i') + (j - j') + 1]C/\mu = -P - (k' + 1)C/\mu$ which contradicts inequality (A.1). Therefore, the new system state must be stable under σ . Furthermore, the new state is also an equilibrium state because σ is given as an equilibrium strategy. \square

Lemma A.2 implies that when an arrival or departure event occurs to an equilibrium state (under an equilibrium strategy σ) and if it creates a transient state, then the execution of strategy σ at the transient state will bring the system back to another equilibrium state in one "iteration" of reshuffling. That is, it cannot occur in equilibrium

Table 3 Payoff matrix for the i th and $(i + 1)$ th ordinary customers

	$\sigma(i + 1, j - 1, k) = Y$	$\sigma(i + 1, j - 1, k) = N$
$\sigma(i, j, k) = Y$	(U_1, U_2)	(U_1, U_3)
$\sigma(i, j, k) = N$	(U_3, U_1)	(U_4, U_5)

that a rational ordinary customer originally decides not to upgrade, but later changes her mind even though there is no change in the total queue length (i.e., no arrivals or departures). Hence, while rational ordinary customers, in principle, must continuously evaluate the priority-purchasing option, it suffices, in actuality, that they evaluate it and possibly upgrade only at discrete time points of arrivals or departures. Moreover, this result also implies that in Definition 1, we can substitute Eqs. (3a)–(3b) and (4) for $V(i', j', k')$ in both Conditions (1) and (5) [i.e., the “if” part of Conditions (3a)–(3b) will be guaranteed by Condition (5)]. Hence, we are effectively left with two sets of equilibrium conditions: Condition (1) maps strategies to value functions, whereas Condition (5) corresponds value functions to their necessitated strategies.

Proof of Lemma 1 For part (i), it suffices to prove that $\sigma(i, j, k) = N$ implies $\sigma(i + 1, j - 1, k) = N$. Given system state $\{i + j, k\}$, let $n_b := \sum_{s=1}^{i-1} \mathbb{I}_{\{\sigma(s, i+j-s, k)=Y\}}$ and $n_a := \sum_{s=i+2}^{i+j} \mathbb{I}_{\{\sigma(s, i+j-s, k)=Y\}}$. That is, n_b and n_a denote the number of ordinary customers in the ordinary line, before and after both the i^{th} and the $(i + 1)^{th}$ ordinary customers, who will purchase priority according to strategy σ at state $\{i + j, k\}$. We can generate the expected utilities for the i^{th} and the $(i + 1)^{th}$ ordinary customers when their priority-purchasing decisions are Y or N . The 2×2 payoff matrix is presented in Table 3, where $U_1 = -P - (k + 1 + n_b)C/\mu$, $U_2 = -P - (k + 2 + n_b)C/\mu$, $U_3 = V_\sigma(i - n_b, j - 1 - n_a, k + n_a + n_b + 1)$, $U_4 = V_\sigma(i - n_b, j - n_a, k + n_a + n_b)$ and $U_5 = V_\sigma(i + 1 - n_b, j - 1 - n_a, k + n_a + n_b)$.

Suppose $\sigma(i, j, k) = N$ and $\sigma(i + 1, j - 1, k) = Y$ form an equilibrium between the two players. Then we must have $U_3 > U_1$ and $U_1 \geq U_5$ (recall that a customer prefers purchasing priority if both the priority and the non-priority options generate the same expected utility). It follows that $U_3 > U_5$ which implies $V_\sigma(i - n_b, j - 1 - n_a, k + n_a + n_b + 1) > V_\sigma(i - n_b + 1, j - n_a - 1, k + n_a + n_b)$, but this is a contradiction to Criterion 3. Therefore, $\sigma(i, j, k) = N$ and $\sigma(i + 1, j - 1, k) = Y$ cannot be an equilibrium between the i^{th} and the $(i + 1)^{th}$ ordinary customers, which implies that if $\sigma(i, j, k) = N$, then $\sigma(i + 1, j - 1, k) = N$. By recursive induction, it is easy to see that $\sigma(i + s, j - s, k) = N$ for all $s \in \{1, 2, \dots, j\}$. On the other hand, part (i) implies part (ii). □

Proof of Lemma 2 Case (1) Suppose $\{i + j, k + 1\}$ is a stable state under σ (which is an equilibrium state because σ is an equilibrium strategy), then one of the two transactions can happen: i) With probability $\frac{\lambda}{\lambda + \mu}$, a new arrival occurs (before a departure does); and ii) with probability $\frac{\mu}{\lambda + \mu}$, a departure occurs (first). Therefore, the expected utility of the customer with position $(i, j, k + 1)$ is equal to

$$V_\sigma(i, j, k + 1)$$

$$\begin{aligned}
 &= -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} V_\sigma(i, j + 1, k + 1) + \frac{\mu}{\lambda + \mu} V_\sigma(i, j, k) \\
 &\leq -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} V_\sigma(i, j, k + 1) + \frac{\mu}{\lambda + \mu} V_\sigma(i, j, k) \text{ by Criterion 2,}
 \end{aligned}$$

which implies $V_\sigma(i, j, k) - V_\sigma(i, j, k + 1) \geq C/\mu$ after quick algebra manipulations.

Case (2) Suppose $\{i + j, k + 1\}$ is a transient state under σ . We shall consider two subcases.

Case (2)-a: If $\sigma(i, j, k + 1) = Y$, by Lemma 1(i) we have $\sigma(s, i + j - s, k + 1) = Y$ for all $s \in \{1, 2, \dots, i - 1\}$. By the first-come-first-upgrade rule, we have $V_\sigma(i, j, k + 1) = -P - (i + k + 1)C/\mu$. It follows from Lemma 1 that $V_\sigma(i, j, k) \geq -P - (i + k)C/\mu = V_\sigma(i, j, k + 1) + C/\mu$, which implies $V_\sigma(i, j, k) - V_\sigma(i, j, k + 1) \geq C/\mu$.

Case (2)-b: If $\sigma(i, j, k + 1) = N$, we must have $\sigma(s, i + j - s, k + 1) = N$ for $s \in \{i + 1, i + 2, \dots, i + j\}$ from Lemma 1(ii). There must exist a largest $n \in \{1, 2, \dots, i - 1\}$ such that $\sigma(n, i + j - n, k + 1) = Y$ because otherwise $\{i + j, k + 1\}$ would be a stable state. By Lemma 1(i), we have $\sigma(s, i + j - s, k + 1) = Y$ for all $s \in \{1, 2, \dots, n\}$. Therefore, n corresponds to the number of ordinary customers who will purchase priority at state $\{i + j, k + 1\}$ according to σ . By Lemma 2, $\{i + j, k + 1\}$ will evolve into an equilibrium state $\{i + j - n, k + n + 1\}$ according to σ . The position of the customer who had position $(i, j, k + 1)$ at state $\{i + j, k + 1\}$ becomes $(i - n, j, k + n + 1)$ at $\{i + j - n, k + n + 1\}$ and her expected utility is given according to (5) by

$$V_\sigma(i, j, k + 1) = V_\sigma(i - n, j, k + n + 1). \tag{A.2}$$

Because the new state is stable under σ , we know by Case (1) that $V_\sigma(i - n, j, k + n) - V_\sigma(i - n, j, k + n + 1) \geq C/\mu$, which implies

$$V_\sigma(i - n, j, k + n) \geq V_\sigma(i - n, j, k + n + 1) + C/\mu, \tag{A.3}$$

Finally, $V_\sigma(i, j, k) \geq V_\sigma(i - n, j, k + n)$ [by Criterion 3] $\geq V_\sigma(i - n, j, k + n + 1) + C/\mu$ [by (A.3)] $= V_\sigma(i, j, k + 1) + C/\mu$ [by (A.2)], which implies that $V_\sigma(i, j, k) - V_\sigma(i, j, k + 1) \geq C/\mu$. □

Proof of Lemma 3 For part (i), because $\sigma(1, j, k) = N$, we know by (5) that

$$V_\sigma(1, j, k) > -P - (k + 1)C/\mu. \tag{A.4}$$

It follows that $V_\sigma(1, j - 1, k) \geq V_\sigma(1, j, k)$ [by Criterion 1] $> -P - (k + 1)C/\mu$ [by (A.4)]. We must have $\sigma(1, j - 1, k) = N$ by (5) when σ is an equilibrium strategy. On the other hand, we know $V_\sigma(1, j, k - 1) \geq V_\sigma(1, j, k) + C/\mu$ [by Lemma 2] $> -P - kC/\mu$ [by (A.4)], which implies $\sigma(1, j, k - 1) = N$. By recursive induction, we can deduce that $\sigma(1, j', k') = N$ for all $j' \in \{0, 1, \dots, j\}$ and $k' \in \{0, 1, \dots, k\}$.

For part (ii), because $\sigma(1, j, k) = Y$, we know by (5) that

$$V_\sigma(1, j, k) \leq -P - (k + 1)C/\mu. \tag{A.5}$$

It follows that $V_\sigma(1, j + 1, k) \leq V_\sigma(1, j, k)$ [by Criterion 1] $\leq -P - (k + 1)C/\mu$ [by (A.5)]. We must have $\sigma(1, j + 1, k) = Y$ by (5) when σ is an equilibrium strategy. On the other hand, we know $V_\sigma(1, j, k + 1) \leq V_\sigma(1, j, k) - C/\mu$ [by Lemma 2] $\leq -P - (k + 2)C/\mu$ [by (A.5)], which implies that $\sigma(1, j, k + 1) = Y$. By recursive induction, we can deduce that $\sigma(1, j', k') = Y$ for all $j' \in \{j, j + 1, \dots\}$ and $k' \in \{k, k + 1, \dots\}$. □

Proof of Theorem 1 Suppose σ is an equilibrium strategy. We will prove Theorem 1 in four steps.

Step (i): We should show that for any $i \in \{1, 2, \dots\}$, there exists a unique threshold $n_i \in \mathbb{N}_0$ such that $\sigma(1, i - 1, k) = Y$ if and only if $k \geq n_i$. By Lemma 2, we know that the expected utility of an ordinary customer with position $(1, i - 1, k)$ is decreasing in k and that $V_\sigma(1, i - 1, k) - V_\sigma(1, i - 1, k + 1) \geq C/\mu$. We discuss two cases. Case (1): If $V_\sigma(1, i - 1, 0) \leq -P - C/\mu$, then clearly $\sigma(1, i - 1, 0) = Y$. It follows from Lemma 3 that $\sigma(1, i - 1, k) = Y$ for all $k \in \{0, 1, \dots\}$, so n_i is equal to zero in this case. Case (2): If $V_\sigma(1, i - 1, 0) > -P - C/\mu$, i.e., $V_\sigma(1, i - 1, 0) - [-P - C/\mu] > 0$, we define $g(k) := V_\sigma(1, i - 1, k) - [-P - (k + 1)C/\mu] > 0$ for $k \in \mathbb{N}_0$ and, by construction, $g(0) > 0$. Moreover, because $V_\sigma(1, i - 1, k)$ is reduced by C/μ or more each time k increases by 1 (by Lemma 2), it is clear that $g(k)$ weakly decreases in k . Therefore, either $g(k)$ stays positive for all $k \in \mathbb{N}_0$ or it crosses zero at least once. Either way, there exists a unique threshold $n_i \geq 1$, which can possibly be infinity, such that $g(k) > 0$ for $k \in \{0, 1, 2, \dots, n_i - 1\}$ and $g(k) \leq 0$ for $k \in \{n_i, n_i + 1, n_i + 2, \dots\}$. Equivalently, we have $V_\sigma(1, i - 1, k) > -P - (k + 1)C/\mu$ for $k \in \{0, 1, 2, \dots, n_i - 1\}$ and $V_\sigma(1, i - 1, k) \leq -P - (k + 1)C/\mu$ for $k \in \{n_i, n_i + 1, n_i + 2, \dots\}$, which implies that $\sigma(1, i - 1, k) = Y$ if and only if $k \geq n_i$.

Step (ii): We should show that $\sigma(2, i - 2, k) = Y$ if and only if $k \geq n_i$ and $k + 1 \geq n_{(i-1)}$. We discuss two cases. Case (1): If $k \leq n_i - 1$, then $\sigma(1, i - 1, k) = N$ by the definition of n_i in Step (i) which implies $\sigma(2, i - 2, k) = N$ by Lemma 1. Therefore, $\sigma(2, i - 2, k) \neq Y$ when $k \leq n_i - 1$. Case (2): Consider $k \geq n_i$, i.e., $\sigma(1, i, k) = Y$. There are two subcases. Case (2)-a: Suppose $\sigma(2, i - 2, k) = N$, we must have $\sigma(s, i - s, k) = N$ for $s \in \{2, 3, \dots, i\}$ by Lemma 1. That is, only the ordinary customer at the head of the ordinary line purchases priority at system state $\{i, k\}$, according to σ . The resulting state $\{i - 1, k + 1\}$ is a stable state under σ by Lemma 2. The position of the tagged customer changes from $(2, i - 2, k)$ (at system state $\{i, k\}$) to $(1, i - 2, k + 1)$ (at system state $\{i - 1, k + 1\}$). Because $\sigma(1, i - 2, k + 1) = N$ due to the definition of a stable state, we have $k + 1 \leq n_{(i-1)} - 1$ again by the definition of n_i in Step (i). Therefore, when $k \geq n_i$, we have shown $\sigma(2, i - 2, k) = N \Rightarrow k + 1 \leq n_{(i-1)} - 1$. By the contrapositive argument, this means

$$k + 1 \geq n_{(i-1)} \Rightarrow \sigma(2, i - 2, k) = Y \text{ when } k \geq n_i. \tag{A.6}$$

Case (2)-b: Now suppose $\sigma(2, i - 2, k) = Y$. Because $\sigma(1, i - 1, k) = Y$ from the assumption of Case (2), we have $V_\sigma(2, i - 2, k) = -P - (k + 2)C/\mu$. We should prove by contradiction that $k + 1 \geq n_{(i-1)}$. If not, i.e., if $k + 1 \leq n_{(i-1)} - 1$, it means $\sigma(1, i - 2, k + 1) = N$ by the definition of n_i in Step (i). It follows that the expected

utility of a customer with position $(1, i - 2, k + 1)$ satisfies $V_\sigma(1, i - 2, k + 1) > -P - (k + 2)C/\mu = V_\sigma(2, i - 2, k)$, which contradicts Criterion 3. Therefore, we must have

$$\sigma(2, i - 2, k) = Y \Rightarrow k + 1 \geq n_{(i-1)} \text{ when } k \geq n_i. \tag{A.7}$$

Based on (A.6) and (A.7), we can conclude that $\sigma(2, i - 2, k) = Y$ if and only if $k \geq n_i$ and $k + 1 \geq n_{(i-1)}$. By recursive induction, we can deduce that, for any customer with position $(s, i - s, k)$,

$$\sigma(s, i - s, k) = Y \text{ if and only if } k + t \geq n_{(i-t)} \text{ for } t = \{0, 1, \dots, s - 1\}. \tag{A.8}$$

Furthermore, at the system state $\{N_o, N_p\} = \{i, k\}$, $\sigma(s, i - s, k) = Y$ implies $\sigma(t, i - t, k) = Y$ for $t \in \{1, 2, \dots, s\}$ by Lemma 1, which means that there exists $\max\{s | \sigma(s, i - s, k) = Y, s \in \mathbb{N}_0\}$. Therefore, $\bar{n}_{\{i,k\}} = \sum_{s=1}^i \mathbb{1}_{\{\sigma(s,i-s,k)=Y\}} = \max\{s | \sigma(s, i - s, k) = Y, s \in \mathbb{N}_0\} = \min\{s | \sigma(s, i - s, k) = N, s \in \mathbb{N}_0\} - 1 = \{s | k + t \geq n_{(i-t)} \text{ for } t \in \{0, 1, \dots, s - 1\} \text{ and } k + s \leq n_{(i-s)} - 1\}$ [by (A.8)] $= \min\{s | k + s \leq n_{(i-s)} - 1, s \in \mathbb{N}_0\}$. Thus, $\bar{n}_{\{N_o, N_p\}} = \min\{s | N_p + s \leq n_{(N_o-s)} - 1, s \in \mathbb{N}_0\}$.

Step (iii): We shall now prove by contradiction that $n_{(i-1)} \geq n_i$ for all $i \in \{2, 3, \dots\}$. Suppose there were some $i \in \{2, 3, \dots\}$ such that $n_{(i-1)} < n_i$, then $n_i - n_{(i-1)} \geq 1$. Let $k_i = n_i - n_{(i-1)} - 1 \geq 0$, so by construction $n_i - 1 = n_{(i-1)} + k_i$. It follows that $V_\sigma(1, i - 1, n_i - 1) \leq V_\sigma(1, i - 2, n_i - 1)$ [by Criterion 1] $= V_\sigma(1, i - 2, n_{(i-1)} + k_i) \leq V_\sigma(1, i - 2, n_{(i-1)}) - k_i C/\mu$ [by Lemma 2] $= -P - (n_{(i-1)} + k_i + 1)C/\mu$ [by Step (i)] $= -P - n_i C/\mu$. The fact that $V_\sigma(1, i - 1, n_i - 1) \leq -P - n_i C/\mu$ implies that the best response for the customer with position $(1, i - 1, n_i - 1)$ is to purchase priority under σ , which is a contradiction to the definition of n_i in Step (i). Therefore, we must have $n_{(i-1)} \geq n_i$ for all $i \in \{2, 3, \dots\}$.

Step (iv): We shall ‘ $\underline{m} \in \mathbb{N}$ such that $n(\underline{m}) = 0$. Otherwise, $n_i \geq 1$ for all $i \in \{1, 2, \dots\}$. It follows by the definition of n_i in Step (i) that $\sigma(1, i - 1, 0) = N$ for all $i \in \{1, 2, \dots\}$. As σ is an equilibrium strategy, by Lemma 1, we have $\sigma(s, i - s, 0) = N$ for all $s \in \{1, 2, \dots, i\}$, i.e., no customer purchases priority when the priority line is empty. It follows that $-iC/\mu \geq V_\sigma(i, 0, 0) > -P - C/\mu$. However, $-iC/\mu$ cannot be greater than $-P - \frac{C}{\mu}$ for all $i \in \mathbb{N}$ because $\lim_{i \rightarrow \infty} -iC/\mu = -\infty$. Therefore, by contradiction, there exists a minimum $\underline{m} \in \mathbb{N}$ such that $n_{\underline{m}} = 0$ and, by Step (ii), $n_i > 0$ for $i \in \{1, 2, \dots, \underline{m} - 1\}$ and $n_i = 0$ for $i \in \{\underline{m} + 1, \underline{m} + 2, \dots\}$. Finally, let $\bar{m} := \left\lceil \frac{\mu P + C}{C} \right\rceil$ for which it is easy to verify that

$$-\bar{m}C/\mu \leq -P - C/\mu, \tag{A.9}$$

and in what follows we will establish that $\underline{m} \leq \bar{m}$. Suppose $\underline{m} > \bar{m}$, then $n_{\bar{m}} > 0$ which implies that $\sigma(1, \bar{m} - 1, 0) = N$ by the definition of n_i in Step (i). Because σ is an equilibrium strategy, we have $\sigma(s, \bar{m} - s, 0) = N$ for $s \in \{1, 2, \dots, \bar{m}\}$ by Lemma 1, and in particular $\sigma(\bar{m}, 0, 0) = N$. It follows from (5) that

$$V_\sigma(\bar{m}, 0, 0) > -P - C/\mu. \tag{A.10}$$

Let Ω_1 and Ω_2 be sample spaces that contain sample paths for a customer with position $(\bar{m}, 0, 0)$ to and not to purchase priority, respectively, according to σ , before she completes her service. If $\omega \in \Omega_1$, it is clear that $V_\sigma(\bar{m}, 0, 0|\omega \in \Omega_1) \leq -P - C/\mu$. If $\omega \in \Omega_2$, then the $(\bar{m} - 1)$ ordinary customers in the ordinary line before the tagged customer with position $(\bar{m}, 0, 0)$ will always get served before her so $V_\sigma(\bar{m}, 0, 0|\omega \in \Omega_2) \leq -\bar{m}C/\mu \leq -P - C/\mu$ [by (A.9)]. Therefore, $V_\sigma(\bar{m}, 0, 0) = Pr(\omega \in \Omega_1) \cdot V_\sigma(\bar{m}, 0, 0|\omega \in \Omega_1) + Pr(\omega \in \Omega_2) \cdot V_\sigma(\bar{m}, 0, 0|\omega \in \Omega_2) \leq -P - C/\mu$ which contradicts (A.10). Therefore, we must have $\underline{m} \leq \bar{m} = \left\lceil \frac{\mu P + C}{C} \right\rceil$. \square

Proof of Corollary 1 By (7) and (8), we know that $\sigma(i, N_o - i, N_p) = Y$ if and only if $i \in \{1, 2, \dots, \bar{n}_{\{N_o, N_p\}}\}$, where $\bar{n}_{\{N_o, N_p\}} = \min\{s|N_p + s \leq n_{(N_o-s)} - 1, s \in \mathbb{N}_0\}$. By the definition of $\bar{n}_{\{N_o, N_p\}}$, it is clear that $N_p + s \geq n_{(N_o-s)}$ for $1 \leq s < \bar{n}_{\{N_o, N_p\}}$ and $s \in \mathbb{N}$. Therefore, if $\sigma(i, N_o - i, N_p) = Y$, we must have $N_p + s \geq n_{(N_o-s)}$ for $s \in \{1, 2, \dots, i - 1\}$ because $i \leq \bar{n}_{\{N_o, N_p\}}$. On the other hand, if $N_p + s \geq n_{(N_o-s)}$ for $s \in \{1, 2, \dots, i - 1\}$, it implies that $i \leq \bar{n}_{\{N_o, N_p\}}$ so $\sigma(i, N_o - i, N_p) = Y$. Hence, $\sigma(i, N_o - i, N_p) = Y$ if and only if $N_p + s \geq n_{(N_o-s)}$ for $s \in \{0, 1, 2, \dots, i - 1\}$. \square

Corollary 6 In a candidate pure strategy, $\bar{n}_{\{N_o, N_p\}}$ is weakly increasing in N_o and in N_p .

Proof of Corollary 6 Consider two system states $\{N_o, N_p\}$ and $\{N_o + 1, N_p\}$. Suppose $0 \leq \bar{n}_{\{N_o+1, N_p\}} < \bar{n}_{\{N_o, N_p\}}$. Then $\bar{n}_{\{N_o, N_p\}} \geq 1$ which implies that $\sigma(1, N_o - 1, N_p) = Y$. Then by the definition of $\bar{n}_{\{N_o+1, N_p\}}$ from (8), we have $N_p + \bar{n}_{\{N_o+1, N_p\}} \leq n_{(N_o+1-\bar{n}_{\{N_o+1, N_p\}})} - 1$, or equivalently, $N_p + [\bar{n}_{\{N_o+1, N_p\}} - 1] \leq n_{(N_o-[\bar{n}_{\{N_o+1, N_p\}}-1])} - 2$. If $\bar{n}_{\{N_o+1, N_p\}} \geq 1$, then $[\bar{n}_{\{N_o+1, N_p\}} - 1] \in \{N_p + s \leq n_{(N_o-s)} - 1, s \in \mathbb{N}_0\}$ which implies $\bar{n}_{\{N_o, N_p\}} \leq \bar{n}_{\{N_o+1, N_p\}}$ and it brings about a contradiction to $\bar{n}_{\{N_o+1, N_p\}} < \bar{n}_{\{N_o, N_p\}}$. Otherwise, if $\bar{n}_{\{N_o+1, N_p\}} = 0$, then $\sigma(1, N_o, N_p) = N$, which implies $\sigma(1, N_o - 1, N_p) = N$ by Lemma 3 and it is a contradiction to $\sigma(1, N_o - 1, N_p) = Y$. Therefore, we must have $\bar{n}_{\{N_o+1, N_p\}} \geq \bar{n}_{\{N_o, N_p\}}$.

On the other hand, we can consider two system states $\{N_o, N_p\}$ and $\{N_o, N_p + 1\}$. Suppose $0 \leq \bar{n}_{\{N_o, N_p+1\}} < \bar{n}_{\{N_o, N_p\}}$. Then $\bar{n}_{\{N_o, N_p\}} > 1$, which implies that $\sigma(1, N_o - 1, N_p) = Y$. Then, by the definition of $\bar{n}_{\{N_o, N_p+1\}}$ from (8), we have $N_p + 1 + \bar{n}_{\{N_o, N_p+1\}} \leq n_{(N_o-\bar{n}_{\{N_o, N_p+1\}})} - 1$, or equivalently, $N_p + \bar{n}_{\{N_o, N_p+1\}} \leq n_{(N_o-\bar{n}_{\{N_o, N_p+1\}})} - 2$. It follows that $\bar{n}_{\{N_o, N_p+1\}} \in \{N_p + s \leq n_{(N_o-s)} - 1, s \in \mathbb{N}_0\}$, which implies $\bar{n}_{\{N_o, N_p\}} \leq \bar{n}_{\{N_o, N_p+1\}}$ and is a contradiction to $\bar{n}_{\{N_o, N_p+1\}} < \bar{n}_{\{N_o, N_p\}}$. Therefore, we must have $\bar{n}_{\{N_o, N_p+1\}} \geq \bar{n}_{\{N_o, N_p\}}$.

By recursive induction, we can conclude that $\bar{n}_{\{N_o, N_p\}}$ weakly increases in N_o and in N_p . \square

Proof of Corollary 2 Given an equilibrium strategy σ , it is sufficient to show that the occurrence of a departure event to a stable state always leads to another stable state. Let (N_o, N_p) be an arbitrary stable state. Then we must have $\sigma(i, N_o - i, N_p) = N$ for $i \in \{1, 2, \dots, N_o\}$, i.e., $\bar{n}_{\{N_o, N_p\}} = \min\{s|N_p + s \leq n_{(N_o-s)} - 1, s \in \mathbb{N}_0\} = 0$. We shall show that the occurrence of a departure event will transform (N_o, N_p) into another stable state.

Case (i): Suppose $N_p \geq 1$. Then a departure event transforms the state (N_o, N_p) into $(N_o, N_p - 1)$. By Corollary 6 (whose proof does not require Corollary 2), we

know that $\bar{n}_{\{N_o, N_p-1\}} \leq \bar{n}_{\{N_o, N_p\}} = 0$, which implies that $\sigma(i, N_o - i, N_p - 1) = N$ for $i \in \{1, 2, \dots, N_o\}$, so $(N_o, N_p - 1)$ is a stable state.

Case (ii): Suppose $N_p = 0$. Then a departure event transforms the state $(N_o, 0)$ into $(N_o - 1, 0)$. By Corollary 6, we know that $\bar{n}_{\{N_o-1, 0\}} \leq \bar{n}_{\{N_o, 0\}} = 0$, which implies that $\sigma(i, N_o - 1 - i, 0) = N$ for $i \in \{1, 2, \dots, N_o - 1\}$, so $(N_o - 1, 0)$ is a stable state. □

Proof of Theorem 2 Consider a strategy σ that is defined according to Definition 2. If all customers follow σ , the system is FIFO. If σ is an equilibrium, we must have $\sigma(1, 0, k) = N$ for all $k \in \mathbb{N}_0$, because for a customer with position $(1, 0, k)$, her expected utility for purchasing priority is strictly dominated by not purchasing. For example, the customer can consider strategy σ' which specifies that she should purchase priority as long as a new arrival occurs before her service completion. Then we have

$$\begin{aligned}
 V_{\sigma, \sigma'}(1, 0, k) &= \left(\frac{\mu}{\lambda + \mu}\right)^{k+1} \left(-\frac{(k+1)C}{\mu}\right) + \left(1 - \left(\frac{\mu}{\lambda + \mu}\right)^{k+1}\right) \left(-P - \frac{(k+1)C}{\mu}\right) \\
 &> -P - \frac{(k+1)C}{\mu},
 \end{aligned}
 \tag{A.11}$$

where $\left(\frac{\mu}{\lambda + \mu}\right)^{k+1}$ is the probability that there are no new arrivals before she completes her service. Therefore, $\bar{n}_{\{N_o, N_p\}} = 0$ when $N_o = 1$ for all $N_p \in \mathbb{N}_0$. It follows from the definition of $\bar{n}_{\{N_o, N_p\}}$ (see Definition 2) that $N_p \leq n_1 - 1$ for all $N_p \in \mathbb{N}_0$, implying that $n_1 = \infty$.

Then we must have $\underline{m} \geq 2$. Consider a customer with position $(2, 0, k)$ for any k . Under strategy σ , this tagged customer will purchase priority only if the customer before her in the regular queue also purchases. That is, if $\sigma(1, 1, k) = N$, we must have $\sigma(2, 0, k) = N$. If $\sigma(1, 1, k) = Y$, this tagged customer is still better off by not upgrading to priority because $V_{\sigma}(2, 0, k) = V_{\sigma}(1, 0, k + 1) > -P - \frac{(k+2)C}{\mu}$ (A.11). Therefore, under equilibrium, we have $\sigma(2, 0, k) = N$ for all $k \in \mathbb{N}_0$. Therefore, if σ is an equilibrium, the customer with position $(1, 1, k)$ will not be overtaken by the one after her. Then it is not necessary for the customer with position $(1, 1, k)$ to purchase priority. In other words, we have $\sigma(1, 1, k) = N$ which implies that $n_2 = \infty$.

Now let us assume that $n_i = \infty$ for $i = 1, 2, \dots, \underline{m} - 1$, where $\underline{m} - 1 \geq 2$. Consider the customer with position $(\underline{m}, 0, k)$ for any k . Denote by n_b the number of customers before her who choose to purchase priority under σ . If σ is an equilibrium, we must have

- (1) If $n_b < \underline{m} - 1$, then $\sigma(\underline{m}, 0, k) = N$ by the definition of σ .
- (2) If $n_b = \underline{m} - 1$, then all customers before the tagged customer would purchase priority, we consider two cases: (i) if $\sigma(\underline{m}, 0, k) = Y$, then it gives the customer a utility of $V_{\sigma}(\underline{m}, 0, k) = -P - \frac{(\underline{m}+k)C}{\mu}$. (ii) if $\sigma(\underline{m}, 0, k) = N$, then $V_{\sigma}(\underline{m}, 0, k) = V_{\sigma}(1, 0, k + \underline{m} - 1) > -P - \frac{(\underline{m}+k)C}{\mu}$ because $n_1 = \infty$. That is, $\sigma(\underline{m}, 0, k) = N$ is the best response for the tagged customer.

Therefore, if σ is an equilibrium, we have $\sigma(\underline{m}, 0, k) = N$ for any $k \in \mathbb{N}_0$. Then we must have $\sigma(\underline{m} - 1, 1, k) = N$, too. Otherwise, if $\sigma(\underline{m} - 1, 1, k) = Y$, by the definition of σ we must have $\sigma(i, \underline{m} - i - 1, k) = Y$. In this case, consider strategy σ' where the tagged customer with position $(\underline{m} - 1, 1, k)$ purchases priority if and

only if there is a new arrival before she completes service. Then $V_\sigma(\underline{m} - 1, 1, k) = -P - \frac{(\underline{m}+k-1)C}{\mu} \geq V_\sigma(1, 1, \underline{m} + k - 2) > V_{\sigma, \sigma'}(1, 1, \underline{m} + k - 2) > -P - \frac{(\underline{m}+k-1)C}{\mu}$, which brings up a contradiction to $\sigma(\underline{m} - 1, 1, k) = N$. Therefore, we can derive that $\sigma(1, \underline{m} - 1, k) = N$ for all $k \in \mathbb{N}_0$, which implies that $n_{\underline{m}} > k$ for all $k \in \mathbb{N}_0$, i.e., $n_{\underline{m}} = \infty$.

This contradicts our result in Theorem 1 that $n_{\underline{m}} = 0$. That is, we can conclude that any strategy defined by Definition 2 cannot be an equilibrium, i.e., $(\mathcal{E} \cap \mathcal{T}) = \emptyset$. \square

Proof of Corollary 3 By Theorem 2, we have $(\mathcal{E} \cap \mathcal{T}) = \emptyset$ which implies $(\mathcal{E} \cap \mathcal{I}) \not\subseteq \mathcal{T}$. Recall we have established in Step 1 that if $(\mathcal{E} \cap \mathcal{I}) \neq \emptyset$ then $(\mathcal{E} \cap \mathcal{I}) \subseteq \mathcal{T}$. By the contrapositive argument, we have $(\mathcal{E} \cap \mathcal{I}) = \emptyset$. \square

Proof of Theorem 3 Let $\hat{\sigma}$ denote a customer’s best response to strategy σ adopted by all other customers, and σ' the strategy that the customer will purchase priority as long as there is a new arrival to the system. In the best response $\hat{\sigma}$, a customer with position $(1, 0, 0)$ or $(1, 0, 1)$ chooses to stay in the regular line because

$$\begin{aligned} V_{\hat{\sigma}, \sigma}(1, 0, 0) &\geq V_{\sigma', \sigma}(1, 0, 0) \\ &= -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \cdot \left(-P - \frac{C}{\mu}\right) + \frac{\mu}{\lambda + \mu} \cdot 0 \\ &> -P - \frac{C}{\mu}, \\ V_{\hat{\sigma}, \sigma}(1, 0, 1) &\geq V_{\sigma', \sigma}(1, 0, 1) = -\frac{C}{\mu} + V_{\sigma', \sigma}(1, 0, 0) > -P - \frac{2C}{\mu}. \end{aligned}$$

Therefore, an equilibrium strategy σ must have $\sigma(1, 0, 0) = \sigma(1, 0, 1) = N$.

Next, we examine the system state with two regular customers in the queue, the first customer holding the position $(1, 1, 0)$ and the second customer holding the position $(2, 0, 0)$. We investigate the strategies of customers at positions $(1, 1, 0)$ and $(2, 0, 0)$ by specifying the payoffs under different strategy combinations:

- (1) If $\sigma(1, 1, 0) = N$ and $\sigma(2, 0, 0) = N$, we have $V_\sigma(1, 1, 0) = -\frac{C}{\mu}$ and $V_\sigma(2, 0, 0) = -\frac{2C}{\mu}$.
- (2) If $\sigma(1, 1, 0) = N$ and $\sigma(2, 0, 0) = Y$, then $V_\sigma(2, 0, 0) = -P - \frac{C}{\mu}$. And we have $V_\sigma(1, 1, 0) = V_\sigma(1, 0, 1) = -\frac{C}{\mu} + V_\sigma(1, 0, 0)$. Notice that under a strategy σ , $V_\sigma(1, 0, 0) = -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} V_\sigma(1, 1, 0) + \frac{\mu}{\lambda + \mu} V_\sigma \cdot 0$. Combining the two equations above, we can obtain that $V_\sigma(1, 1, 0) = -\frac{(2+\rho)C}{\mu}$.
- (3) If $\sigma(1, 1, 0) = Y$ and $\sigma(2, 0, 0) = N$, then $V_\sigma(1, 1, 0) = -P - \frac{C}{\mu}$. And we have $V_\sigma(2, 0, 0) = -\frac{C}{\mu} + V_\sigma(1, 0, 0)$. Notice that under a strategy σ , $V_\sigma(1, 0, 0) = -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu}(-P - \frac{C}{\mu}) + \frac{\mu}{\lambda + \mu} V_\sigma \cdot 0$. Combining the two equations above, we can obtain that $V_\sigma(1, 1, 0) = -\frac{2C}{\mu} - \frac{\rho}{1+\rho} P$.
- (4) If $\sigma(1, 1, 0) = Y$, $\sigma(2, 0, 0) = Y$, we have $V_\sigma(1, 1, 0) = -P - \frac{C}{\mu}$ and $V_\sigma(2, 0, 0) = -P - \frac{2C}{\mu}$.

Table 4 Payoff matrix for customers with positions (1, 1, 0) and (2, 0, 0)

	$\sigma(2, 0, 0) = Y$	$\sigma(2, 0, 0) = N$
$\sigma(1, 1, 0) = Y$	$(-P - \frac{C}{\mu}, -P - \frac{2C}{\mu})$	$(-P - \frac{C}{\mu}, -\frac{\rho}{1+\rho}P - \frac{2C}{\mu})$
$\sigma(1, 1, 0) = N$	$(-\frac{(2+\rho)C}{\mu}, -P - \frac{C}{\mu})$	$(-\frac{C}{\mu}, -\frac{2C}{\mu})$

Combing cases (1)–(4) above yields a 2×2 payoff matrix between the customers with positions (1, 1, 0) and (2, 0, 0), as shown in Table 4.

From Table 4, given $\sigma(1, 1, 0) = Y$, then the best response of the customer with position (2, 0, 0) must be N because $-P - \frac{2C}{\mu} < -\frac{\rho}{1+\rho} \cdot P - \frac{2C}{\mu}$. Therefore, (Y, Y) cannot be an equilibrium. However, (Y, N) cannot be an equilibrium, either; this is because given $\sigma(2, 0, 0) = N$, the best response of the customer with position (1, 1, 0) must be N because $-P - \frac{C}{\mu} < -\frac{C}{\mu}$. Furthermore, when $P > \frac{C}{\mu}$, we have $-P - \frac{C}{\mu} < -\frac{2C}{\mu}$ and $-P - \frac{C}{\mu} < -\frac{C}{\mu}$; thus (N, N) is an equilibrium. When $0 < P \leq \frac{C}{\mu}$, we have $-P - \frac{C}{\mu} \geq -\frac{2C}{\mu} > -\frac{(2+\rho)C}{\mu}$, so neither (N, N) nor (N, Y) is an equilibrium. This completes the proof. \square

Proof of Proposition 1 We define the probability to purchase priority to be $pr(i, j) \in [0, 1]$ for an arriving customer when there are i ordinary and j priority customers in the system, where $(i, j) \in \Omega_s = \{(0, 0), (0, 1), (1, 0)\}$. For any given $pr(i, j)$, $(i, j) \in \Omega_s$, similarly to the proof of Theorem 3, we can obtain the expected utilities of the tagged customer below:

- (1) Upon seeing state (1, 0), if the tagged customer chooses to purchase, i.e., $pr'(1, 0) = 1$, her expected utility is $V_\sigma(0, 0) = -\frac{C}{\mu} - P$. Otherwise, if $pr'(1, 0) = 0$, we have $V_\sigma(0, 0) = -\frac{C}{\mu} \cdot \frac{2(1+\rho)-pr(1,0)\rho}{1+\rho-pr(1,0)\rho}$. Therefore, (i) $pr(1, 0) = 1$ is an equilibrium if and only if $-\frac{C}{\mu} - P \geq -\frac{C}{\mu} \cdot \frac{2(1+\rho)-\rho}{1+\rho-\rho} \Leftrightarrow v_P \leq 1 + \rho$; (ii) $pr(1, 0) = 0$ is an equilibrium if and only if $-\frac{C}{\mu} - P \leq -\frac{C}{\mu} \cdot \frac{2(1+\rho)}{1+\rho} \Leftrightarrow v_P \geq 1$; (iii) Any $pr(1, 0) \in (0, 1)$ is an equilibrium if and only if $-\frac{C}{\mu} - P = -\frac{C}{\mu} \cdot \frac{1+\rho}{1+\rho-pr(0,0)\rho}$, i.e., $pr(1, 0) = \frac{v_P-1}{v_P} \cdot \frac{1+\rho}{\rho} \in [0, 1]$. Notice that $pr(1, 0) = 0$ is the unique Pareto-dominant equilibrium strategy in the presence of multiple equilibria. Thus, the unique Pareto-dominant equilibrium strategy is

$$pr^e(1, 0) = \begin{cases} 1, & \text{if } v_P < 1, \\ 0, & \text{if } v_P \geq 1. \end{cases}$$

- (2) Upon seeing state (0, 0), if the tagged customer chooses to purchase, i.e., $pr'(0, 0) = 1$, her expected utility is $V_\sigma(0, 0) = -\frac{C}{\mu} - P$. Otherwise, if $pr'(0, 0) = 0$, we have $V_\sigma(0, 0) = -\frac{C}{\mu+\lambda} + \frac{\lambda}{\mu+\lambda}[pr(1, 0)V_\sigma(0, 0) - C/\mu]$, which gives $V_\sigma(0, 0) = -\frac{C}{\mu} \cdot \frac{1+\rho}{1+\rho-pr(1,0)\rho}$. Thus (i) $pr(0, 0) = 1$ is an equilibrium if and only if $-\frac{C}{\mu} - P \geq -\frac{C}{\mu} \cdot \frac{1+\rho}{1+\rho-pr(1,0)\rho} \Leftrightarrow \frac{v_P(1+\rho)}{(1+v_P)\rho} \leq pr(1, 0) \leq 1$; (ii) $pr(0, 0) = 0$ is an equilibrium if and only if $-\frac{C}{\mu} - P \leq -\frac{C}{\mu} \cdot \frac{1+\rho}{1+\rho-pr(1,0)\rho} \Leftrightarrow$

$pr(1, 0) \leq \min \left\{ 1, \frac{v_P(1+\rho)}{(1+v_P)\rho} \right\}$; (iii) Any $pr(0, 0) \in (0, 1)$ is an equilibrium if and only if $-\frac{C}{\mu} - P = -\frac{C}{\mu} \cdot \frac{1+\rho}{1+\rho-pr(0,0)\rho}$, i.e., $pr(1, 0) = \frac{v_P(1+\rho)}{(1+v_P)\rho} \leq 1$.

- (3) Upon seeing state $(0, 1)$, if the tagged customer chooses to purchase, i.e., $pr'(0, 1) = 1$, her expected utility is $V_\sigma(0, 1) = -\frac{2C}{\mu} - P$. Otherwise, if $pr'(0, 1) = 0$, we have $V_\sigma(0, 1) = -\frac{C}{\mu} - \frac{C}{\mu} \cdot \frac{1+\rho}{1+\rho-pr(1,0)\rho} = -\frac{C}{\mu} \cdot \frac{2(1+\rho)-pr(1,0)\rho}{1+\rho-pr(1,0)\rho}$. Thus, (i) $pr(0, 1) = 1$ is an equilibrium if and only if $-\frac{2C}{\mu} - P \geq -\frac{C}{\mu} \cdot \frac{2(1+\rho)-pr(1,0)\rho}{1+\rho-pr(1,0)\rho} \Leftrightarrow \frac{v_P(1+\rho)}{(1+v_P)\rho} \leq pr(1, 0) \leq 1$; (ii) $pr(0, 0) = 0$ is an equilibrium if and only if $-\frac{2C}{\mu} - P \leq -\frac{C}{\mu} \cdot \frac{2(1+\rho)-pr(1,0)\rho}{1+\rho-pr(1,0)\rho} \Leftrightarrow pr(1, 0) \leq \min \left\{ 1, \frac{v_P(1+\rho)}{(1+v_P)\rho} \right\}$; (iii) Any $pr(0, 1) \in (0, 1)$ is an equilibrium if and only if $pr(1, 0) = \frac{v_P(1+\rho)}{(1+v_P)\rho} \leq 1$.

Combing cases (1)–(3) above gives

- If $v_P < 1$, (i) when $v_P \leq \rho$, we have $pr^e(0, 0) = pr^e(0, 1) = 1$; (ii) when $\rho < v_P < 1$, $pr^e(0, 0) = pr^e(0, 1) = 0$.
- If $v_P \geq 1$, then $pr^e(0, 0) = pr^e(0, 1) = pr^e(1, 0) = 0$.

Combining the results above completes the proof of the equilibrium strategies. Next, we characterize the priority revenue. We consider three cases below:

- If $v_P \leq \rho$, notice that in equilibrium, the arriving customers will always purchase priority (if admissible), then $\pi_{1,0} = 0$, and we can get that $\Pi_{K=2}^{ARR}(v_P) = \lambda P(\pi_{0,0} + \pi_{0,1}) = \frac{\lambda P(1+\rho)}{1+\rho+\rho^2}$.
- If $\rho < v_P < 1$, the arriving customers will purchase priority when the system state is found to be $(1, 0)$, then we have $\Pi_{K=2}^{ARR}(v_P) = \lambda P\pi_{1,0} = \frac{\lambda P\rho}{1+\rho+\rho^2}$.
- If $v_P \geq 1$, we have $\Pi_{K=2}^{ARR}(v_P) = 0$.

□

Proof of Theorem 4 Since $\lambda \ll \mu$, future arrivals and the number of existing priority customers do not affect customer strategies. Thus, without loss of generality, we let the number of existing priority customers be zero for simplicity. We proceed by backward induction:

- (i) The very last customer with position $\lceil v_P \rceil + 1$: Given the strategies of all the previous customers, if she upgrades, then her expected utility is $-P - \frac{2C}{\mu}$; if she does not, then her expected utility is $-\frac{(\lceil v_P \rceil + 1)C}{\mu}$. Note that $-P - \frac{2C}{\mu} < -\frac{(\lceil v_P \rceil + 1)C}{\mu}$ because $\lceil v_P \rceil = \lceil \mu P / C \rceil$. Therefore, the very last customer indeed will not upgrade given the strategies of all the previous customers.
- (ii) The second last customer with position $\lceil v_P \rceil$: Given the strategies of all the previous customers, if she upgrades, then her expected utility is $-P - \frac{C}{\mu}$. If she were not to upgrade, then the last customer would upgrade because, for the last customer, $-P - \frac{C}{\mu} \geq -\frac{(\lceil v_P \rceil + 1)C}{\mu}$. Thus, for the second last customer, the expected utility if she does not upgrade is $-\frac{(\lceil v_P \rceil + 1)C}{\mu}$. Since $-P - \frac{C}{\mu} \geq$

- $\frac{(\lceil v_P \rceil + 1)C}{\mu}$, the second last customer will indeed upgrade given the strategies of all the previous customers.
- (iii) Customer with position $\lceil v_P \rceil - 1$: Given the strategies of all the previous customers, if she upgrades, her expected utility is $-P - C/\mu$; if she does not upgrade, then the second last customer and only that customer will upgrade, which implies that the expected utility of the customer with position $\lceil v_P \rceil - 1$ is $-\lceil v_P \rceil C/\mu$. Since $-P - C/\mu < -\lceil v_P \rceil C/\mu$, the customer will indeed not upgrade.
 - (iv) We complete the proof by induction. Suppose customer i does not upgrade given that all the previous customers do not, $i \leq \lceil v_P \rceil - 1$. We show that customer $i - 1$ also does not upgrade. If customer $i - 1$ upgrades, her expected utility is $-P - C/\mu$; if she does not, then her expected utility is $-iC/\mu$. Since $i \leq \lceil v_P \rceil - 1$, $-P - C/\mu < -iC/\mu$, and customer $i - 1$ will indeed not upgrade.

□

Proof of Theorem 5 First, we show that under the proposed strategy, if the ordinary queue has only one customer, the customer will indeed not upgrade. The expected utility of not upgrading for the customer is $V_2(1, 0, k)$:

$$\begin{aligned}
 &V_2(1, 0, k) \\
 &= -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \left[-P - \frac{(k + 1)C}{\mu} \right] + \frac{\mu}{\lambda + \mu} V_2(1, 0, k - 1), \quad \forall k \geq 1; \\
 &V_2(1, 0, 0) \\
 &= -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \left[-P - \frac{C}{\mu} \right].
 \end{aligned}$$

Thus,

$$V_2(1, 0, k) = -\frac{(k + 1)C}{\mu} - \left[1 - \left(\frac{\mu}{\lambda + \mu} \right)^{k+1} \right] P. \tag{A.12}$$

Therefore, $V_2(1, 0, k) > -\frac{(k+1)C}{\mu} - P$, which implies that the customer will indeed not upgrade, i.e., $\sigma(1, 0, k) = \frac{\mu}{N}, \forall k$.

Next, we verify that if $\rho \geq \sqrt{v_P} - 1$, under the proposed strategy, customers will indeed choose $\sigma(i, j, k) = Y$ for $i + j \geq 2$.

For customer (i, j, k) who is about to decide, the expected utility of upgrading is $-\frac{(k+1)C}{\mu} - P$. The expected utility of not upgrading under the proposed strategy is $V_{i+1}(i, j, k) = V_{i+1}(i, 0, k + j)$ (because all the customers behind would upgrade):

$$\begin{aligned}
 &V_{i+1}(i, 0, k + j) \\
 &= -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \left[-P - \frac{(k + i + j)C}{\mu} \right] + \frac{\mu}{\lambda + \mu} V_{i+1}(i, 0, k + j - 1), \\
 &\quad \text{if } k + j \geq 1; \\
 &V_{i+1}(i, 0, k + j)
 \end{aligned} \tag{A.13}$$

$$= -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \left[-P - \frac{(k + i + j)C}{\mu} \right] + \frac{\mu}{\lambda + \mu} V_i(i - 1, 0, 0),$$

if $k = j = 0$. (A.14)

Thus, upgrading is incentive compatible if and only if

$$-\frac{(k + 1)C}{\mu} - P \geq V_{i+1}(i, 0, k + j), \quad \forall (i, j, k) : i + j \geq 2. \quad (A.15)$$

Under the proposed strategy, we can verify that $V_{i+1}(i, 0, k + j) = V_{i+1}(i, 0, 0) - (k + j)C/\mu, \forall k \geq 0$, which is decreasing in i and j by (A.13)–(A.13). Thus, Conditions (A.15) are satisfied if and only if

$$-\frac{(k + 1)C}{\mu} - P \geq V_2(1, 0, k + 1), \quad \forall k \geq 0 \quad \text{and} \quad (A.16)$$

$$-\frac{(k + 1)C}{\mu} - P \geq V_3(2, 0, k), \quad \forall k \geq 0. \quad (A.17)$$

Plugging (A.12) into Conditions (A.16) and collecting terms yields

$$\frac{C}{\mu} \geq \left(\frac{\mu}{\lambda + \mu} \right)^{k+2} P, \quad \forall k \geq 0 \Leftrightarrow \frac{C}{\mu} \geq \left(\frac{\mu}{\lambda + \mu} \right)^2 P \Leftrightarrow \rho \geq \sqrt{v_P} - 1.$$

Since $V_3(2, 0, k) - V_3(2, 0, k + 1) = C/\mu, \forall k$, (A.17) is equivalent to $-\frac{C}{\mu} - P \geq V_3(2, 0, 0)$. Plugging (A.14) into the inequality gives

$$-\frac{C}{\mu} - P \geq -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \left[-P - \frac{2C}{\mu} \right] + \frac{\mu}{\lambda + \mu} V_2(1, 0, 0) \Leftrightarrow$$

$$-\frac{C}{\mu} - P \geq -\frac{C}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \left[-P - \frac{2C}{\mu} \right] + \frac{\mu}{\lambda + \mu} \left[-\frac{C}{\mu} - \frac{\lambda}{\lambda + \mu} P \right].$$

Collecting terms also gives $\rho \geq \sqrt{v_P} - 1$.

Therefore, if $\rho \geq \sqrt{v_P} - 1$, the proposed strategy σ meets all the equilibrium conditions, which implies that σ is an equilibrium strategy. This completes this proof. □

Proof of Theorem 6 Suppose all customers adopt a threshold strategy $n \geq 2$, i.e., the arrival of the $(n + 1)$ th ordinary customer triggers the purchase of the n th ordinary customer only. For customer at position $(1, n - 1, 0)$, her expected waiting time is one busy period $1/(\mu - \lambda)$. Thus, the expected utility by following threshold strategy n is $-C/(\mu - \lambda)$. If she upgrades, her expected utility is $-P - C/\mu$. Therefore, if $-\frac{C}{\mu - \lambda} < -P - C/\mu \Leftrightarrow \lambda > \frac{\mu v_P}{v_P + 1}$, it is optimal for her to deviate and purchase priority. That is to say, the threshold strategy $n \geq 2$ is not an equilibrium. □

Proof of Theorem 7 Let $U^*(1, 0)$ denote the maximum expected utility of customer in the system when the system state is $(1, 0)$. Obviously, we have $U^*(1, 0) \leq -C/\mu$.

On the other hand, if the customer in the position $(1, 0, 0)$ purchases the priority as long as an arrival happens before her service completion, her expected utility is given by $-\frac{C}{\lambda+\mu} + \frac{C\lambda(-\nu_P-1)}{\mu(\lambda+\mu)}$. Then we have $U^*(1, 0) \geq \frac{C}{\mu} \left(-1 - \frac{\rho\nu_P}{1+\rho}\right)$.

When the system has two ordinary customers and the strategy of the two customers is (Y, Y) , their expected utility is given by $\left(-\frac{C}{\mu}(1 + \nu_P), -\frac{C}{\mu}(2 + \nu_P)\right)$, accordingly. If their strategy is (Y, N) , the corresponding expected utility is $\left(-\frac{C}{\mu}(1 + \nu_P), -\frac{C}{\mu}[1 - U^*(1, 0)]\right)$. If their strategy is (N, Y) , the corresponding expected utility is $\left(-\frac{C}{\mu}[1 - U^*(1, 0)], -\frac{C}{\mu}(1 + \nu_P)\right)$. If their strategy is (Y, N) , the corresponding expected utility is $\left(-1, -\frac{C}{\mu}[1 - U^*(1, 0)]\right)$. We consider the following two cases:

- (1) When $\nu_P > 1$, and all customers adopt strategy N no matter where they are, then it is not difficult to verify that the best response for each customer is N . Thus (N, N) is an equilibrium.
- (2) When $\nu_P \leq 1$, as $U^*(1, 0) \leq -\frac{C}{\mu}$, the best response of customer $(2, 0, 0)$ to $\sigma(1, 1, 0) = N$ is Y . Then the expected utility of customer $(1, 1, 0)$ is $-\frac{C}{\mu}[1 - U^*(1, 0)]$ if she chooses N . Since $-\frac{C}{\mu}(1 + \nu_P) \geq -\frac{C}{\mu}[1 - U^*(1, 0)]$, the best response of customer $(1, 1, 0)$ is Y . Thus (Y, N) is the unique equilibrium by noticing that $-\frac{C}{\mu}[1 - U^*(1, 0)] > -\frac{C}{\mu}(2 + \nu_P)$.

□

Proof of Theorem 8 (1) In upon-arrival priority purchase, it is not difficult to verify that the optimal priority price is $P_{ARR} = C \min\{\rho, 1\}/\mu$. Then $\Pi_{K=2}^{ARR} = \frac{C\rho \min\{\rho, 1\}(1+\rho)}{1+\rho+\rho^2}$.

- (2) In sequential in-queue priority purchase, the optimal price is $P_{SQ} = C/\mu$, and the optimal revenue is $\Pi_{K=2}^{SQ} = \frac{C\rho^2}{1+\rho+\rho^2}$. And it is not difficult to verify that $\Pi_{K=2}^{ARR} > \Pi_{K=2}^{SQ}$, which completes the proof.

□

Proof of Theorem 9 Notice that $\sigma(2, 0, 1) = Y \Leftrightarrow -2C/\mu - P \geq -C/\mu + V_3(2, 0, 0) \Leftrightarrow -C/\mu - P \geq V_3(2, 0, 0)$ and $\sigma(2, 0, 0) = Y \Leftrightarrow -C/\mu - P \geq V_3(2, 0, 0)$. It follows that $\sigma(2, 0, 1) = Y \Leftrightarrow \sigma(2, 0, 0) = Y$. On the other hand, we have $\sigma(1, 1, 1) = Y \Leftrightarrow -2C/\mu - P \geq V_2(1, 1, 1)$ and $\sigma(1, 1, 0) = Y \Leftrightarrow -C/\mu - P \geq V_2(1, 1, 0)$. We consider two cases below:

- (i) If $\sigma(2, 0, 1) = \sigma(2, 0, 0) = Y$, it implies that $V_2(1, 1, 1) = V_2(1, 0, 2) = -C/\mu + V_2(1, 0, 1)$ and $V_2(1, 1, 0) = V_2(1, 0, 1)$. Thus, $\sigma(1, 1, 1) = Y \Leftrightarrow -C/\mu - P \geq V_2(1, 0, 1) \Leftrightarrow \sigma(1, 1, 0) = Y$.
- (ii) If $\sigma(2, 0, 1) = \sigma(2, 0, 0) = N$, it implies that $V_2(1, 1, 1) = V_3(1, 1, 1) = -C/\mu + V_3(1, 1, 0)$ and $V_2(1, 1, 0) = V_3(1, 1, 0)$. Thus, we can still have $\sigma(1, 1, 1) = Y \Leftrightarrow V_2(1, 0, 1) \Leftrightarrow \sigma(1, 1, 0) = Y$.

Combining the two cases completes the proof.

□

References

1. Adiri, I., Yechiali, U.: Optimal priority-purchasing and pricing decisions in nonmonopoly and monopoly queues. *Oper. Res.* **22**(5), 1051–1066 (1974)
2. Afèche, P.: Incentive-compatible revenue management in queueing systems: optimal strategic delay. *Manuf. Serv. Oper. Manag.* **15**(3), 423–443 (2013)
3. Afèche, P., Sarhangian, V.: Rational abandonment from priority queues: equilibrium strategy and pricing implications. Working paper, University of Toronto (2015)
4. Akşin, Z., Ata, B., Emadi, S.M., Su, C.L.: Structural estimation of callers' delay sensitivity in call centers. *Manag. Sci.* **59**(12), 2727–2746 (2013)
5. Akşin, Z., Ata, B., Emadi, S.M., Su, C.L.: Impact of delay announcements in call centers: an empirical approach. *Oper. Res.* **65**(1), 242–265 (2016)
6. Alperstein, H.: Note-optimal pricing policy for the service facility offering a set of priority prices. *Manag. Sci.* **34**(5), 666–671 (1988)
7. Assaf, D., Haviv, M.: Reneging from processor sharing systems and random queues. *Math. Oper. Res.* **15**(1), 129–138 (1990)
8. Ata, B., Peng, X.: An equilibrium analysis of a multiclass queue with endogenous abandonments in heavy traffic. *Oper. Res.* **66**(1), 163–183 (2018)
9. Balachandran, K.: Purchasing priorities in queues. *Manag. Sci.* **18**(5–part–1), 319–326 (1972)
10. Cui, S., Veeraraghavan, S., Wang, J., Zhang, Y.: In-queue observation and abandonment. Working paper, Georgetown University (2018)
11. Cui, S., Wang, Z., Yang, L.: The economics of line-sitting. *Manag. Sci.* **66**(1), 227–242 (2020)
12. Gavirneni, S., Kulkarni, V.G.: Self-selecting priority queues with burr distributed waiting costs. *Product. Oper. Manag.* **25**(6), 979–992 (2016)
13. Gurvich, I., Lariviere, M.A., Ozkan, C.: Coverage, coarseness, and classification: determinants of social efficiency in priority queues. *Manag. Sci.* **65**(3), 1061–1075 (2019)
14. Hassin, R.: *Rational Queueing*. Chapman and Hall/CRC, Boca Raton (2016)
15. Hassin, R., Haviv, M.: Equilibrium threshold strategies: the case of queues with priorities. *Oper. Res.* **45**(6), 966–973 (1997)
16. Hassin, R., Haviv, M.: *To Queue Or Not To Queue: Equilibrium Behavior in Queueing Systems*, vol. 59. Springer, New York (2003)
17. Kerner, Y., Sherzer, E., Yanco, M.A.: On non-equilibria threshold strategies in ticket queues. *Queueing Syst.* **86**(3–4), 419–431 (2017)
18. Maglaras, C., Yao, J., Zeevi, A.: Observational learning and abandonment in a congested queue. Working paper, Columbia University (2017)
19. Mendelson, H., Whang, S.: Optimal incentive-compatible priority pricing for the M/M/1 queue. *Oper. Res.* **38**(5), 870–883 (1990)
20. Naor, P.: The regulation of queue size by levying tolls. *Econ. J. Econ. Soc.* **37**, 15–24 (1969)
21. Nazerzadeh, H., Randhawa, R.S.: Near-optimality of coarse service grades for customer differentiation in queueing systems. *Prod. Oper. Manag.* **27**(3), 578–595 (2018)
22. Wang, J., Cui, S., Wang, Z.: Equilibrium strategies in M/M/1 priority queues with balking. *Prod. Oper. Manag.* **28**(1), 43–62 (2019)
23. Webb, E.M., Yu, Q., Bretthauer, K.M.: Linking delay announcements, abandonment, and service time. Working paper, Indiana University (2017)
24. Yang, L., Debo, L.: Referral priority program: leveraging social ties via operational incentives. *Manag. Sci.* **65**(5), 2231–2248 (2019)
25. Yang, L., Debo, L., Gupta, V.: Trading time in a congested environment. *Manag. Sci.* **63**(7), 2377–2395 (2017)
26. Yang, L., Wang, Z., Cui, S.: A model of queue-scalping. *Manag. Sci.* <https://doi.org/10.1287/mnsc.2020.3865> (2021)