Equilibrium Arrival Times to Queues: The Case of Last-Come First-Serve Preemptive-Resume

by

Jesper Breinbjerg and Lars Peter Østerdal

Equilibrium Arrival Times to Queues: The Case of Last-Come First-Serve Preemptive-Resume[*]

Jesper Breinbjerg[†]   Lars Peter Østerdal[‡]

March 17, 2017

Abstract

We consider a non-cooperative queueing environment where a finite number of customers independently choose when to arrive at a queueing system that opens at a given point in time and serves customers on a last-come first-serve preemptive-resume (LCFS-PR) basis. Each customer has a service time requirement which is identically and independently distributed according to some general probability distribution, and they want to complete service as early as possible while minimizing the time spent in the queue. In this setting, we establish the existence of an arrival time strategy that constitutes a symmetric (mixed) Nash equilibrium, and show that there is at most one symmetric equilibrium. We provide a numerical method to compute this equilibrium and demonstrate by a numerical example that the social efficiency can be lower than the efficiency induced by a similar queueing system that serves customers on a first-come first-serve (FCFS) basis.

Keywords   Queueing · Strategic arrival times to a queue · Non-cooperative games

JEL Classification   C72 · D62 · R41

[†]Department of Business and Economics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M. Email: jnb@sam.sdu.dk

[‡]Department of Economics, Copenhagen Business School, Porcelænshaven 16A, DK-2000 Frederiksberg. Email: lpo.eco@cbs.dk

1   Introduction

Customers routinely experience waiting in line whenever the demand for some service exceeds the capacity to provide it. Examples of such a situation include purchasing popular concert tickets, calling a telephone hotline, and conducting online bank transactions. To cope with the excess demand, the allocation of service to customers is often managed with a queueing system. Since waiting in line is not just a source of emotional frustration for the customer, but also exerts substantial societal costs, the study of how to optimally manage such systems constitutes an important mechanism design problem.

This paper considers a non-cooperative environment where a finite number of customers independently need to choose when to arrive at a queueing system that provides some desired service at the end. The customers are served by a single server that only operates during a certain interval of time, and each customer wishes to complete service as early as possible, while minimizing the expected wait in the queue. The order in which waiting customers are served is determined by the service discipline. While the most frequently used discipline is the *First-Come First-Serve* (FCFS), our goal here is to study the strategic arrival behavior of self-interested customers in a system that employs the *Last-Come First-Servce Preemptive-Resume* (LCFS-PR) service discipline. The LCFS-PR discipline admits any newly arrived customer into service immediately, possibly preempting the service progress of another customer. The preempted customer then joins the queue where later arrivals are prioritized over earlier arrivals. When a preempted customer re-enters service again, her service is resumed from the point of interruption.

The strategic choices of arrival time to queues have been studied for almost half a century (see Hassin, 2016, for an extensive survey). The problem was first approached by considering a fluid model for congestion dynamics that studied the equilibrium arrival behavior of a continuum of customers (Vickrey, 1969). In this model, each customer must choose their arrival time to a continuously open bottleneck, and each customer has a preferred time for passing the bottleneck and will incur a cost from being early or late. Similar fluid models have further been studied and extended in various directions, e.g. to treat heterogeneous customers (Arnott et al., 1989), elastic customer demand (Arnott et al., 1993), and hypercongestion (Verhoef, 2003).

The study of strategic arrivals in queueing systems where the server has a limited service period (i.e. the server admits an opening and/or closing time) was first formulated for a Poisson-distributed number of identical customers with exponential service requirements that arrive at a server with a known opening and closing time, and wish to minimize their own waiting time (Glazer and Hassin, 1983). This work showed that, in a symmetric equilibrium, the customers arrive according to a continuous probability distribution that extends over a finite interval before and after the opening time. Several variations of this

2

model have since been considered, e.g. to treat bulk service (Glazer and Hassin, 1987), no arrivals prior to opening (Hassin and Kleiner, 2011) and discrete arrival times and deterministic service times (Rapoport et al., 2004; Seale et al., 2005; Stein et al., 2007). While the aforementioned studies assumed that customers only want to minimize their wait in the queue, another body of literature studies environments where customers also care about being served at an early time. This additional type of disutility has been modelled as a tardiness cost that increases the later one is admitted into service. The equilibrium behavior induced by such customer preferences has been studied for several variants of assumptions. Specifically, the symmetric equilibrium have been studied for a Poisson-distributed number of identical customers with multilinear cost of waiting and tardiness in time and exponential service time requirements with and without early arrivals, along with their fluid analogues (Jain et al., 2011; Haviv, 2013). A complete analysis of the existence and uniqueness of the equilibrium for a general population size with multilinear costs and exponential service times showed that there always exists an equilibrium and that it is in fact symmetric (Juneja and Shimkin, 2013). Lastly, the uniqueness and existence of the symmetric equilibrium was established for more general classes of utility functions and service time distributions (Breinbjerg, 2017).

The above-mentioned studies all considered queueing environments where the FCFS service discipline was employed. While the FCFS discipline is intuitively fair and acceptable to most people, it is not the only way, and sometimes not even the optimal way, of settling a queue (Hassin, 1985). In particular, a recent study shows that the FCFS discipline in fact provides the lowest level of social efficiency (measured as the sum of aggregated customer utilities) relative to any other service discipline in queueing environments where the server opens at a given point in time and a continuum of customers choose their arrival time when they associate costs of queueing and being served late (Platz and Østerdal, 2017). Conversely, the same study also shows that the *Last-Come First-Serve* (LCFS) discipline provides the highest level of social efficiency. Thus, these two disciplines provide an upper and a lower bound for efficiency under any general stochastic service disciplines.[1] Further support for such equilibrium utility bound has been established in queueing environments for a finite (but small) population size where each customer chooses their arrival time from a discrete set of time slots (Breinbjerg et al., 2016).

In the present paper, we consider a queueing environment where a finite number of customers with identical preferences, composed of waiting and tardiness costs, choose when to arrive at a single-server facility that opens at a commonly known point in time and serves customers on a LCFS-PR basis. We shall restrict attention to facilities with

---

[1] For the fluid model, it has been show for varying degrees of random sorting, ranging from FCFS to a completely random service order, that the choice of service discipline does not play a role for the properties of social efficiency (de Palma and Fosgerau, 2013).

no closing time and do not allow customers to leave the queue once they have arrived. Our main contributions are the following: First, we develop a constructive procedure that establishes the existence of a symmetric (mixed) Nash equilibrium and show that there is at most one symmetric equilibrium. Second, we show by a numerical example that there exists a symmetric Nash equilibrium under the LCFS-PR discipline where the social efficiency (measured via the price of anarchy) is lower than the efficiency induced by an identical queueing system that instead serves customers on FCFS basis. Thus, the equilibrium utility bound found by Platz and Østerdal (2017) does not hold generally for the queueing environment considered within this paper.

The paper is organized as follows: Section 2 formalizes the queueing environment and model assumptions. Section 3 defines the relevant notion of the equilibrium solution, presents the equilibrium properties of the queueing model, and provides the proof hereof. Section 4 presents a numerical method to compute the symmetric equilibrium and then compares the resulting social efficiency with that obtained in a corresponding queueing system that employs the FCFS service discipline. We conclude the paper in Section 5 with a brief summary and future research directions. Proofs that require technical notation for the stochastic queueing dynamics are relegated to the Appendix.

2 The LCFS-PR Queueing Game[2]

A finite set of customers $\mathcal{N} = \{1, 2, ..., \eta\}$, $\eta \geq 2$ must obtain service by a single-server facility. The facility processes customers for service within the bounded interval of time $[0, \infty)$, i.e. the facility starts service at time 0 and does not close before all customers have been served. The facility serves one customer at a time according to a work-conserving LCFS-PR regime. We assume that customers may arrive and queue up at the facility both before and after the opening time, and moreover, that a customer cannot leave the queue once arrived. The time required for the facility to complete the service of each customer is assumed to be independent and identically distributed according to the cumulative distribution function S which is absolutely continuous, strictly increasing and has finite moments.

Strategy of Arrival. Each customer $i \in \mathcal{N}$ randomly chooses her time of arrival according to the (*symmetric*) *strategy* F which represents a cumulative probability distribution (cdf) that assigns to each point in time t on the real line $\mathbb{R}$ the probability $F(t)$ that customer i has arrived by time t. We assume that F is piecewise differentiable and let

_____

[2]The queueing game of this paper closely resembles that presented in a companion paper by Breinbjerg (2017) with the exception of the considered service discipline and the customers' set of possible arrival times.

$f_+(t)$ denote the right-derivative of F at point t. Let $\mathcal{S}(F)$ denote the *support* of strategy F which is the smallest closed set of probability 1, namely $\int_{\mathcal{S}(F)} dF(t) = 1$.

Time of Departure. Given a strategy F, we consider the probabilities associated with the time for which a customer has completed her service and departs the system. Let D_i denote the ex-ante *cumulative departure time distribution* for any customer i such that $D_i(d \mid t, F)$ is the probability that customer i has departed the system by time $d \in \mathbb{R}$ given that she arrived at time t and the $\eta - 1$ other customers arrive according to F. Note that $\lim_{d \to \infty} D_i(d \mid t, F) = 1$ for all t since the customer population is finite, the service time distribution S has finite moments, and LCFS-PR is work-conserving. Note also that $D_i(d \mid t, F) = 0$ for all $d \leq t$ and all $d \leq 0$.

Utility Function. We assume that the customers have identical preferences. Each customer wants to receive service as early as possible and spend a minimum of time in the queue. To capture such preferences, let $V(t, d)$ be the utility of a customer who arrives at time t and departs the system at time d after waiting in the queue for $d - t$ time units. We assume that V is well-defined and continuous for all $d \geq t$, and strictly decreasing in both the departure time d and the waiting time $d - t$. Moreover, V is bounded from above and $\lim_{t \to \infty} V(t, t) = -\infty$.

We assume that every customer is (von Neumann-Morgenstern) rational and aims to maximize her expected utility with respect to her time of arrival. For a given strategy F, we denote by U_i customer i's expected utility by arriving at time t with certainty when the $\eta - 1$ other customers arrive according to F, so

$$U_i(t, F) = \int_t^\infty V(t, d) \; dD_i(d \mid t, F) \tag{1}$$

where $\int$ is the Lebesgue integral over the departure time distribution D_i. A LCFS-PR queueing game is thus represented by a tuple $\mathcal{G} = \langle \, \eta, V, S \, \rangle$.

3 Equilibrium Analysis

This section first defines the notion of a *symmetric equilibrium* in Section 3.1, which is used as the solution concept to study outcomes of queueing game $\mathcal{G}$. Subsequently in Section 3.2, we present our main results in Theorem 1. The theorem establishes the existence and uniqueness of a symmetric Nash equilibrium as well as the general properties of such equilibrium. The proof of Theorem 1 is presented in Section 3.3.

3.1 Symmetric Equilibrium

To study the strategic arrivals of customers in a queueing game $\mathcal{G}$, we adopt the Nash equilibrium concept and restrict the analysis only to consider symmetric solutions. This

restriction is customary when analyzing non-cooperative queueing games (see Breinbjerg, 2017, for further justification). Formally, we define the symmetric equilibrium as follows,

Definition 1   *For any queueing game $\mathcal{G}$, we say that strategy F constitutes a symmetric (Nash) equilibrium if, and only if, it holds for every customer $i \in \mathcal{N}$ that*

   (i) $U_i(t, F) \geq U_i(s, F)$ *for all* $t \in \mathcal{S}(F)$ *and* $s \in \mathbb{R}$

   (ii) $U_i(t, F) = U_i(s, F)$ *for all* $t, s \in \mathcal{S}(F)$.

Definition 1 prescribes a probability distribution that assigns to every point in time, the probability of each customer arriving at the facility, such that no customer can change her strategy unilaterally and obtain strictly higher expected utility. We occasionally refer to a strategy that constitutes a symmetric equilibrium as an *equilibrium strategy*.

3.2   Results

The main results are summarized in the following theorem which presents some general properties of the equilibrium strategy under a queueing game $\mathcal{G}$.

Theorem 1   *For any queueing game $\mathcal{G}$, there exists one, and only one, strategy F that constitutes a symmetric equilibrium. Moreover, the following properties hold for F:*

   (i) $F(t)$ is continuous at all $t \in \mathbb{R}$ and has $F(s) = 0$ for all $s \leq 0$.

   (ii) The support $\mathcal{S}(F)$ of F is a compact and connected set.

Figure 1 depicts a graphical representation of an equilibrium strategy F for a queueing game $\mathcal{G}$. Intuitively speaking, Theorem 1 suggests that, in equilibrium, the customers will not arrive at any point before (or at) the opening time of the system, while they arrive according to a continuous and strictly increasing probability distribution that extends over a finite interval of time starting immediately after the opening time.

3.3   Proof of Theorem 1

This section is devoted to the proof of Theorem 1 which proceeds through several lemmas. We start by establishing a directional result between the expected utility U_i and the cumulative departure time distribution D_i for any customer i.

Lemma 1 *For any queueing game $\mathcal{G}$, let F and $\tilde{F}$ be two distinct strategies. For any customer $i \in \mathcal{N}$, then $U_i(t, F) \geq U_i(t, \tilde{F})$ for any $t \in \mathbb{R}$ if $D_i(d \mid t, F) \geq D_i(d \mid t, \tilde{F})$ for all $d \in \mathbb{R}$. Furthermore, if strict inequality holds for some d, then $U_i(t, F) > U_i(t, \tilde{F})$.*

This claim follows immediately once we note that the utility function V is monotone decreasing in the waiting and departure time.

   The next result addresses the continuity of an equilibrium strategy.
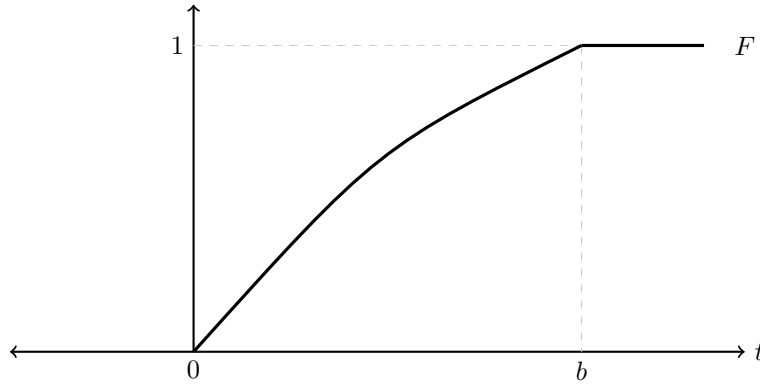
Fig. 1: *Example of an equilibrium strategy:* Let strategy F be an equilibrium strategy under a queueing game $\mathcal{G}$. Then the support $\mathcal{S}(F) = [0, b]$ where $b < \infty$ such that F is strictly increasing over the interval $[0, b]$. Moreover, $F(s) = 0$ for all $s \leq 0$.

Lemma 2 *For any queueing game $\mathcal{G}$, let F be an equilibrium strategy. Then $F(t) = \lim_{s \uparrow t} F(s)$ for all $t \in \mathbb{R}$.*

Proof. We prove by contradiction. Let F be an equilibrium strategy and suppose for some point in time $t \in \mathbb{R}$ that F has a point of upwards discontinuity such that $F(t) > \lim_{s \uparrow t} F(s)$. We then note the following: For any $t \geq 0$, a customer can arrive immediately after time t and start service instantaneously since any newly arrived customer preempts the service progress of a customer already residing in the queue. For any $t < 0$, a customer can arrive immediately after time t and be prioritized for service ahead of the customers already residing in the queue. Formally, this means that for any $t \in \mathbb{R}$ and any $i \in \mathcal{N}$, $D_i(d \mid t, F) \leq \lim_{s \downarrow t} D_i(d \mid s, F)$ for all $d \in \mathbb{R}$ with strict inequality at some d, hence $U_i(t, F) < \lim_{s \downarrow t} U_i(s, F)$ by Lemma 1. This contradicts the equilibrium definition and proves that $F(t) = \lim_{s \uparrow t} F(s)$ for all $t \in \mathbb{R}$. $\qquad \square$

Since F is right-continuous, by definition, it follows by Lemma 2 that any equilibrium strategy F is continuous at all $t \in \mathbb{R}$. By a similar argument, it can also be shown that any equilibrium strategy F must have $F(s) = 0$ for all $s \leq 0$, since a customer can arrive immediately after the opening time and start service instantaneously. Hence, Theorem 1, part (i), follows immediately.

The next result establishes that all customers, in equilibrium, arrive at the facility within some bounded interval of time.

Lemma 3 *For any queueing game $\mathcal{G}$, let F be an equilibrium strategy. Then $\mathcal{S}(F)$ is a compact set.*

Proof. Let F be an equilibrium strategy. We note that the support $\mathcal{S}(F)$ of F is bounded from below at 0 since $F(s) = 0$ for all $s \leq 0$. Moreover, $\mathcal{S}(F)$ is also bounded from above

7

since the customer population is finite, the service time distribution S has finite moments, the LCFS-PR discipline is work-conserving, and $\lim_{t\to\infty} V(t,t) = -\infty$ (a formal proof of this claim can be seen in Breinbjerg, 2017, Lemma 4). The support $\mathcal{S}(F)$ is thus bounded. Since $\mathcal{S}(F)$ is closed, by definition, it follows immediately from the Heine–Borel theorem that $\mathcal{S}(F)$ is compact. $\qquad\square$

The next result addresses the monotonicity of an equilibrium strategy.

Lemma 4 *For any queueing game $\mathcal{G}$, let F be an equilibrium strategy. Then $\mathcal{S}(F)$ is a connected set, i.e. $\mathcal{S}(F)$ cannot be divided into two disjoint and nonempty sets.*

Proof. Let F be an equilibrium strategy. It then follows that F has a bounded support $\mathcal{S}(F)$ with supremum $0 < b < \infty$ by Lemma 3, F is everywhere continuous and have $F(s) = 0$ for all $s \le 0$ by Lemma 2.

Now, suppose that $\mathcal{S}(F)$ is not a connected set, in the sense that $\mathcal{S}(F)$ can be divided into two disjoint nonempty sets. This implies that there exists an interval $0 \le t_1 < t_2 \le b$ such that $F(t_1) = F(t_2)$. However this leads to a contradiction of the equilibrium definition since $U_i(t_1, F) > U_i(t_2, F)$ for any $i \in \mathcal{N}$. To see this, note that any customer that arrives at time t_1 will start service instantaneously according to the LCFS-PR service discipline. Given that V is strictly decreasing in departure time, then any customer has $U_i(t_1, F) > U_i(t_2, F)$ since they can capitalize on the time interval $[t_1, t_2]$ where no customers arrive. Hence, any strategy F with an interior hole in $\mathcal{S}(F)$ cannot be an equilibrium strategy. $\qquad\square$

The claim of Theorem 1, part (ii), thus follows immediately from Lemma 3 and 4. We next address the existence of an equilibrium strategy for a queueing game $\mathcal{G}$.

Lemma 5 *There exists a strategy F that constitutes a symmetric equilibrium for a queueing game $\mathcal{G}$.*

Proof. We constructively prove this claim by defining a family of functions $\{X_b\}_{0<b<\infty}$ where, for each constant b, X_b is the limit of a convergent and recursive sequence $\{X_{b,h} \mid 0 < b < \infty\}_{h\in\mathbb{N}}$ indexed by the non-negative integer h. We then show that a member of the family $\{X_b\}$ represents an equilibrium strategy.

We start by providing some useful notation. For a given $0 < b < \infty$ and $h \in \mathbb{N}$, let $X_{b,h} : \mathbb{R} \to [0,1]$ be a function where $X_{b,h}(t)$ is an image of $X_{b,h}$ at t. For any point in time t where $X_{b,h}$ is non-decreasing and right-continuous over $[t,b]$ and $X_{b,h}(s) = 1$ for all $s \ge b$, then the expected utility $U_i(t, X_{b,h})$ is well-defined for any customer $i \in \mathcal{N}$. Lastly, let $A_h(b) = \sup\{t \in \mathbb{R} \mid X_{b,h}(t) = 0\}$ denote the maximal value of t where $X_{b,h}(t)$ equals 0.

Fix b to be a constant, $0 < b < \infty$, and let $X_{b,h}(t) = 1$ for all $t \geq b$. Intuitively, one may think of $X_{b,h}(b)$ as the earliest point in time where $\eta - 1$ customers have arrived at the facility with certainty. In this case, if the other $\eta - 1$ customers arrive according to $X_{b,h}$, then customer i can arrive at time b and start service instantaneously without being preempted, thus obtaining an expected utility of $U_i(b, X_{b,h})$.

We now define a sequence of recursive functions $X_{b,0}, X_{b,1}, X_{b,2} \ldots$ where $X_{b,0}$ is the designated starting term. We start by characterizing the properties of $X_{b,0}$. For each $t \leq b$, let

$$X_{b,0}(t) = \begin{cases} 1 & \text{if } x_{b,0}^t < 0 \\ 1 - x_{b,0}^t & \text{if } x_{b,0}^t \in [0,1] \\ 0 & \text{if } x_{b,0}^t > 1 \end{cases} \tag{2}$$

where

$$x_{b,0}^t = \sup \left\{ x \in [0, \infty) \mid U_i(b, X_{b,0}) \leq \lim_{s \uparrow t} U_i(s, X_{b,0}^x) \right\} \tag{3}$$

and $X_{b,0}^x$ is a function defined as

$$X_{b,0}^x(s) = \begin{cases} 1 - x & \text{for } s < t \\ 1 & \text{for } s \geq t \end{cases}. \tag{4}$$

for any $s \in \mathbb{R}$ given b, x, and t. Otherwise, $X_{b,0}(t) = 1$ for all $t > b$. Intuitively speaking, $x_{b,0}^t$ represents for a given customer i the maximal expected share of the other $\eta - 1$ customers that can arrive simultaneously at time t such that customer i yields at least the same expected utility by arriving immediately before time t compared to that of arriving at time b, assuming that the remaining expected share of customers $1 - x_{b,0}^t$ arrived before the arrival of customer i. Note that $x_{b,0}^t$ is uniquely determined for each t since $\lim_{s \uparrow t} U_i(s, X_{b,0}^x)$ is strictly decreasing in x. This claim follows immediately by the preemption property of the LCFS-PR discipline. Therefore, the higher expected share of customers arriving exactly at time t, the longer customer i is expected to wait in line before service completion. Note also that $A_0(b)$ exists and that $X_{b,0}$, by construction, is strictly increasing over the interval $[A_0(b), b]$ since the utility function V is strictly decreasing over both departure and waiting time. Figure 2 graphically illustrates an example of $X_{b,0}$.

We next characterize the recursive statement of $X_{b,h}$ for each $h > 0$. Suppose that $X_{b,h-1}$ has been defined for $h > 0$. For each $t \leq b$, let

$$X_{b,h}(t) = \begin{cases} X_{b,h-1}(t) & \text{if } x_{b,h}^t < 0 \\ X_{b,h-1}(t) - x_{b,h}^t & \text{if } x_{b,h}^t \in [0, X_{b,h-1}(t)] \\ 0 & \text{if } x_{b,h}^t > X_{b,h-1}(t) \end{cases} \tag{5}$$
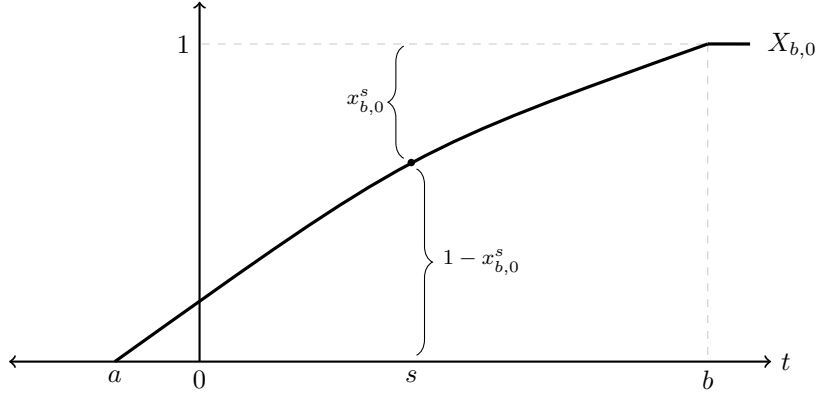
Fig. 2: *Example of $X_{b,0}$:* The constant b is an arbitrarily fixed point in time for which $X_{b,0}(t) = 1$ for all $t \geq b$. Note that $X_{b,0}$ is continuous and strictly increasing over the time interval $[a, b]$ where $a = A_0(b)$.

where

$$x_{b,h}^t = \sup \left\{ x \in [0, \infty) \mid U_i(b, X_{b,h}) \leq \lim_{s \uparrow t} U_i(s, X_{b,h}^x) \right\} \tag{6}$$

and $X_{b,h}^x$ is a function defined as

$$X_{b,h}^x(s) = \begin{cases} X_{b,h-1}(t) - x & \text{for } s < t \\ X_{b,h-1}(s) & \text{for } s \geq t \end{cases} . \tag{7}$$

for any $s \in \mathbb{R}$ given b, h, x and t. Otherwise, $X_{b,h}(t) = 1$ for all $t > b$. Intuitively speaking, $x_{b,h}^t$ represents for a given customer i the maximal expected share of $\eta - 1$ customers that arrive simultaneously at time t such that customer i yields at least the same expected utility by arriving immediately before time t compared to that of arriving at time b, assuming that an expected share of customers, $X_{b,h-1}(t) - x_{b,h}^t$, have arrived before the arrival of customer i, and that the remaining expected share of customers, $1 - X_{b,h-1}(t)$, arrives according to $X_{b,h-1}$. Note that $x_{b,h}^t$ is uniquely determined for each t since $\lim_{s \uparrow t} U_i(s, X_{b,h}^x)$ is strictly decreasing in x. This claim follows by a similar argument to that presented for $h = 0$. Note also that $A_h(b)$ exists and that $X_{b,h}$, by construction, is strictly increasing over the interval $[A_h(b), b]$ since the utility function V is strictly decreasing over both departure and waiting time.

The recursive process yields the sequence $X_{b,0}, X_{b,1}, X_{b,2}, \ldots$ which is bounded and monotonically decreasing with $X_{b,0}(t) \geq X_{b,1}(t) \geq \ldots$ over $h \in \mathbb{N}$ and for all $t \in \mathbb{R}$. It thus follows by the monotone convergence theorem that the sequence is convergent. Let $X_b(t) = \lim_{h \to \infty} X_{b,h}(t)$ denote the limit of the sequence at each t, and let $A(b) = \lim_{h \to \infty} A_h(b)$. Figure 3 graphically illustrates an example of a recursive sequence $X_{b,0}, X_{b,1}, \ldots$ that converges towards the limit X_b.

So far b has been fixed. We now define a family of functions $\{X_b\}_{0 < b < \infty}$ where for each b, X_b is the limit of the convergent and recursive sequence $\{X_{b,h} \mid 0 < b < \infty\}_{h \in \mathbb{N}}$. For each

10
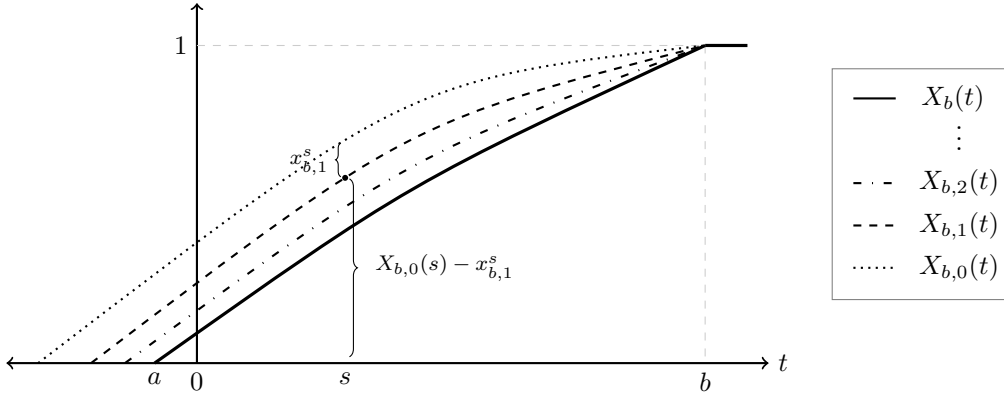
Fig. 3: *Example of a recursive sequence $X_{b,0}, X_{b,1}, \ldots$: As the number of iterations h increases, the hth recursively stated term $X_{b,h}$ converges towards the limit X_b. Note that X_b is continuous and strictly increasing over $[a, b]$ where $a = A(b)$.*

member of $\{X_b\}$, we examine whether it represents an equilibrium strategy. Specifically, we establish a value of b for which the member of $\{X_b\}$ satisfies the following three criteria: (1) X_b is non-decreasing and right-continuous over $\mathbb{R}$, (2) $\lim_{t \to \infty} X_b(t) = 1$, and lastly (3) for any customer i, $U_i(t, X_b) \geq U_i(s, X_b)$ for each $t \in \mathbb{R}$ where $0 < X_b(s) < X_b(t)$ for all $s < t$, and moreover, $U_i(b, X_b) \geq U_i(q, X_b)$ for all $q \geq b$.[3]

First we note that X_b satisfies criteria (1) and (2) for any value of $0 < b < \infty$, by construction. We also note that X_b only satisfies criteria (3) for values of b where $A(b) = 0$. We then make the following observations:

(i) For b being sufficiently close to ∞, then $A(b) > 0$. This follows from observing that A is monotonically increasing in b.

(ii) For b being sufficiently close to 0, then $A(b) < 0$. This follows by the inverse argument of (i).

(iii) $A(b)$ is continuous at all b. This follows by construction of X_b and A.

Combining (i), (ii) and (iii), there must exist a $b = b^*$ such that $X_{b^*}(0) = 0$ with $A(b^*) = 0$. Figure 4 graphically illustrates an example of such X_{b^*}. It thus immediately follows that X_{b^*} represents a strategy that constitutes a symmetric equilibrium. $\square$

We next address the uniqueness of an equilibrium strategy.

Lemma 6 *There exists at most one strategy F that constitutes a symmetric equilibrium for a queueing game $\mathcal{G}$.*

Proof. We prove by contradiction. Let F and $\tilde{F}$ be two distinct equilibrium strategies such that $F \neq \tilde{F}$. Let $b = \min\{t \mid F(t) = 1\}$ and $\tilde{b} = \min\{t \mid \tilde{F}(t) = 1\}$. In what follows, we distinguish between the following three cases of b and $\tilde{b}$:

---

[3]Note that property (3) is an alternative statement of the equilibrium definition (Definition 1) where X_b need not have a well-defined support set.

11
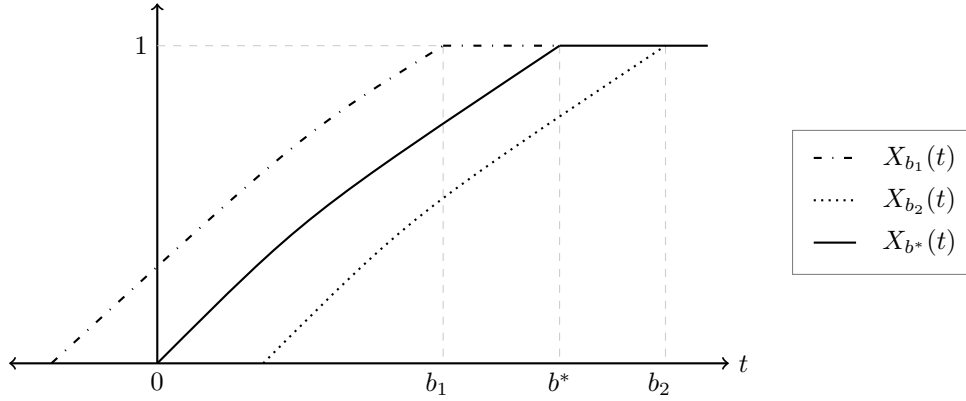
Fig. 4: *Example of $\{X_b\}_{0<b<\infty}$: All members of $\{X_b\}$ for which $b \neq b^*$ yields $A(b) \neq 0$. Only the member of $\{X_b\}$ for which $b = b^*$ represents an equilibrium strategy as $A(b^*) = 0$.*

$b < \tilde{b}$: In this case, it immediately follows that $U_i(t, F) > U_i(t, \tilde{F})$ for all $t \in [0, b]$ and any $i \in \mathcal{N}$. Let $s = \max\{t \mid F(t) = \tilde{F}(t), 0 \leq t < b\}$ be the latest point in time at which the two strategies intersect. Note that s exists and is uniquely determined since F and $\tilde{F}$ are continuous, $F(0) = \tilde{F}(0)$, and $b < \tilde{b}$. It then follows that the expected share of customers arriving from time s up until time b is strictly larger under F than that under $\tilde{F}$; hence, $D_i(d \mid s, F) \leq D_i(d \mid, s, \tilde{F})$ for all d with strict inequality at some d. Thus $U_i(s, F) < U_i(s, \tilde{F})$ for any i by Lemma 1. This contradicts the assumption that F provides higher expected utility than $\tilde{F}$, thus proves that F and $\tilde{F}$ cannot both be an equilibrium strategy.

$b > \tilde{b}$: The case is symmetric to that of $b < \tilde{b}$ and thus omitted.

$b = \tilde{b}$: In this case, it immediately follows that $U_i(t, F) = U_i(t, \tilde{F})$ for all $t \in [0, b]$ and any $i \in \mathcal{N}$. Let $s = \max\{t \mid F(t) \neq \tilde{F}(t), 0 \leq t \leq b\}$ be the latest point in time where F and $\tilde{F}$ do not intersect, and note that this point always exists since F and $\tilde{F}$ are continuous and $F \neq \tilde{F}$. Moreover, $f_+(s)$ and $\tilde{f}_+(s)$ denote the right-derivative of F and $\tilde{F}$ at point s, respectively. Then one of the following cases must hold:

(i) $F(s) > \tilde{F}(s)$ and $f_+(s) < \tilde{f}_+(s)$

(ii) $F(s) < \tilde{F}(s)$ and $f_+(s) > \tilde{f}_+(s)$

If case (i) holds, then there must exist an $\epsilon > 0$ (sufficiently small) such that $F(s) - F(s-\epsilon) < \tilde{F}(s) - \tilde{F}(s-\epsilon)$. For such ϵ, it follows that $D_i(d \mid s-\epsilon, F) \geq D_i(d \mid s-\epsilon, \tilde{F})$ for all d with strict inequality at some d, hence $U_i(s - \epsilon, F) > U_i(s - \epsilon, \tilde{F})$. This contradicts that $U_i(t, F) = U_i(t, \tilde{F})$ for all $t \in [0, b]$. A symmetric argument holds in case (ii); hence, proving that F and $\tilde{F}$ cannot both be an equilibrium strategy.

We have thus shown that there cannot exist two distinct strategies F and $\tilde{F}$ with $F \neq \tilde{F}$ where both constitute a symmetric equilibrium, hence proven Lemma 6. □

Lemma 5 and 6 thus complete the proof of Theorem 1.

4   Computational Results

This section presents a numerical method to compute an equilibrium strategy for a queueing game $\mathcal{G}$, and moreover, demonstrates a numerically computed example of such equilibrium strategy (Section 4.1). We subsequently establish the social efficiency of the computed equilibrium example (measured via the price of anarchy) and compare it to the social efficiency obtained in a corresponding queueing system that employs the FCFS service discipline (Section 4.2). To obtain tractable numerical solutions for the LCFS-PR game, we shall restrict attention only to games of two customers with independent, identical and exponentially distributed service time requirements.

We start by deriving a mathematical expression of D_i for the LCFS-PR queueing game with two customers. The expression is stated in the following lemma.

Lemma 7 *For any queueing game $\mathcal{G}$ where $\eta = 2$ and S is identical, independently and exponentially distributed, let F be a strategy with $b = \min\{t \mid F(t) = 1\} < \infty$. Then the cumulative departure time distribution D_i can be expressed as*

$$D_i(d \mid t, F) = \sum_{a \leq t} I_a G(d - t; \mu) + \int_{-\infty}^{t} f_+(a) G(d - t; \mu) \mathrm{d}a$$
$$+ \sum_{t < a < b} I_a \left(G(a - t; \mu) G(d - t; \mu) + (1 - G(a - t; \mu)) H(d - a; 2, \mu) \right)$$
$$+ \int_{t}^{b} f_+(a) \left(G(a - t; \mu) G(d - t; \mu) + (1 - G(a - t; \mu)) H(d - a; 2, \mu) \right) \mathrm{d}a$$

for each $t \geq 0$ where G is the cdf of the exponential distribution, H is the cdf of the Erlang distribution, and I_a is the size of jump discontinuity of F at point a, in the sense that

$$G(x; \mu) = \begin{cases} 1 - e^{-\mu x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$H(x; k, \mu) = \begin{cases} 1 - \sum_{m=0}^{k-1} \frac{1}{m!} e^{-\mu x} (\mu x)^m & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$I_a = \begin{cases} F(a) - \lim_{s \uparrow a} F(s) & \text{if } F(a) - \lim_{s \uparrow a} F(s) > 0 \\ 0 & \text{otherwise} \end{cases}$$

for any $x, a \in \mathbb{R}$.

Proof. The proof is relegated to Appendix A.1 as it requires additional notation to describe the stochastic queueing process and its sample path relations.

13

4.1 Numerical Procedure and Results

We now present a numerical method to compute an equilibrium strategy. The method is a discretized variant of the constructive proof of Lemma 5. Figure 5 depicts a flowchart of the general numerical procedure. For a given set of inputs, the method performs a search for the value b that induces a function X_b which constitutes an equilibrium strategy. Note that the number of required iterations for the search of b to converge is a function of the tolerance parameter ϵ. For any equilibrium strategy with $b \neq 1$, the search method (which combines a linear and binary search) requires multiple, and possibly many, iterations of b before convergence.
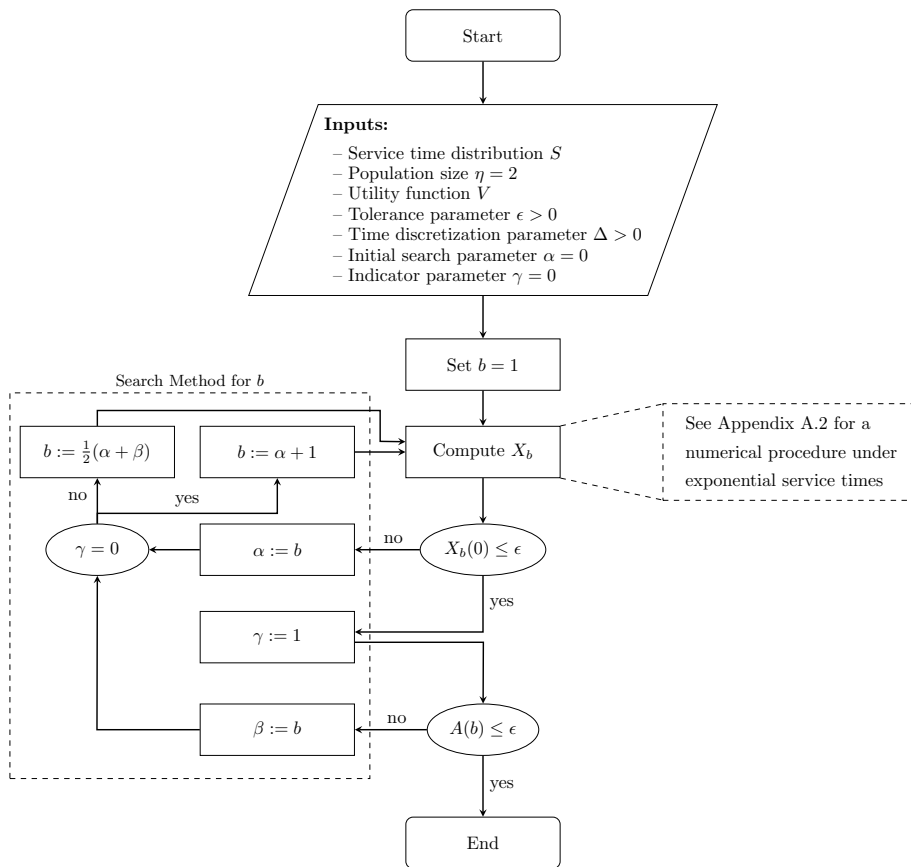
Fig. 5: *Flowchart of the numerical procedure:* Each geometric shape represents an action within the method. That is, the rounded squares are the start and ending, the trapezium is the exogenous inputs, the squares are steps in the process, and circles are binary decisions (yes/no) based on a question. The arrows indicate the flow from one action to another. Note that := is the assignment operator that changes an existing variable's value.

We next apply the numerical procedure to compute an equilibrium strategy under exponential service time which is depicted in Figure 6. The figure illustrates a recursive sequence $\{X_{b,h}\}_{h \in \mathbb{N}}$ for $b = 4.00$ that convergences at $h = 14$ and represents an equilibrium strategy. Intuitively speaking, the symmetric equilibrium prescribes a strategy for

which each customer arrives according to a strictly increasing probability distribution that extends over the interval from the opening time and up until time 4.00.
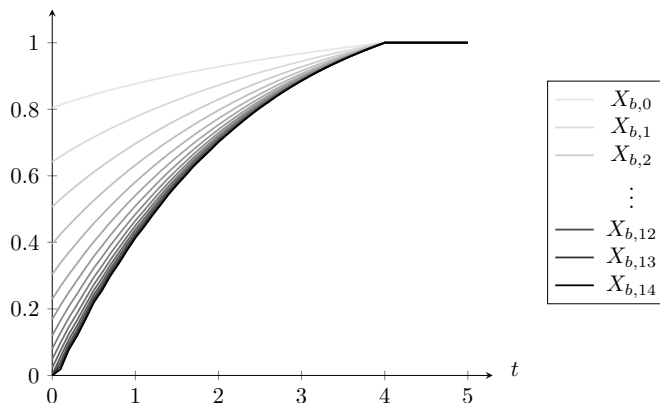
Fig. 6: *Numerically computed equilibrium strategy:* Example of a strategy that constitutes a symmetric equilibrium in a queueing game $\mathcal{G}$ where $\eta = 2$, $V(t,d) = -d^{0.5}(d-t)^{0.8}$, S is identical, independently and exponentially distributed with rate $\mu = 1$, and $\Delta = 0.1$, $\epsilon = 0.001$. The function $X_{b,14}$ represents an equilibrium strategy in the sense that it approximates the convergent limit of the recursive sequence $\{X_{b,h}\}_{h \in \mathbb{N}}$ with respect to the tolerance parameter ϵ.

4.2 Price of Anarchy (LCFS-PR vs. FCFS)

We measure the social inefficiency of the queueing game via the *price of anarchy* (PoA) which corresponds to the ratio of the aggregate expected utility in the Nash equilibrium, to that of the socially optimal solution. Naturally, the socially optimal solution depends on the restrictions imposed on the central planner. Similarly to Juneja and Shimkin (2013), we consider the case where the central planner is allowed to schedule arrivals based on observed service completions.[4] For the considered queueing game with only two customers, the socially optimal solution is that where one customer starts service at time 0 and the other starts service immediately after the departure of the first customer with no idleness at the server. Formally, let $\mathbf{W}$ denote the (random) sum of the two customers' utilities in the socially optimal solution. Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be the (random) independent, identical and exponentially distributed service time requirements of the customers. The expected value of $\mathbf{W}$ conditional on $\mathbf{S}_1$ and $\mathbf{S}_2$ is then given by

$$\mathrm{E}[\mathbf{W} \mid \mathbf{S}_1, \mathbf{S}_2] = V(0, \mathbf{S}_1) + V(\mathbf{S}_1, \mathbf{S}_1 + \mathbf{S}_2)$$

where E is the expectation operator. Note that the sum of two independent and identically exponentially distributed variables follows an Erlang distribution with shape 2. Let $g(x; \mu)$

---

[4]Another option is to assume that the arrival times must be prescheduled, with no feedback on service completions. While this may pose a reasonable choice for the present queueing game, finding the optimal schedule for this problem is a hard global optimization problem, and typically can only be solved using heuristic or approximation algorithms.

and $h(x; \mu)$ denote the density function at x for the exponential and Erlang distribution with rate μ and shape 2, respectively. Then the expected sum of customer utilities for the socially optimal solution is given by

$$\mathrm{E}[\mathbf{W}] = \int_0^\infty \int_0^\infty \left(V(0, s) + V(s, z) \right) g(s; \mu) h(z; \mu) \mathrm{d}s \mathrm{d}z. \tag{8}$$

Let U_i^* denote the expected utility for customer i induced by the equilibrium strategy for a queueing game with $\eta = 2$, then PoA is given by

$$\mathrm{PoA} = \frac{\sum_i U_i^*}{\mathrm{E}[\mathbf{W}]}. \tag{9}$$

Table 1 reports the approximated value of PoA in the specific queueing game considered in Figure 6 for which the utility function is given by $V(t, d) = -d^{0.5}(d - t)^{0.8}$ and the service time requirement S is exponentially distributed with rate 1. The table also reports the approximated PoA obtained by a corresponding queueing game with the FCFS service discipline as studied by Breinbjerg (2017).[5] For the FCFS queueing game, they numerically computed an equilibrium strategy for a similar queueing game with $\eta = 2$, $V(t, d) = -d^{0.5}(d - t)^{0.8}$, S is exponentially distributed with rate $\mu = 1$, and $\Delta = 0.1$, $\epsilon = 0.001$ (see Breinbjerg, 2017, Figure 6). For such a corresponding game, they find that the symmetric equilibrium prescribes a strategy for which both customers arrive with certainty at the opening time zero. When comparing the two approximated values of PoA, we find that the FCFS queueing game yields a lower PoA compared to that of the LCFS-PR queueing game. This means that arrival time incentives provided by FCFS service discipline lead to lower social inefficiency compared to those by the LCFS-PR discipline. Hence, the general social efficiency bounds established in Platz and Østerdal (2017) does not hold for the queueing game with a finite number of customers as considered by the present paper.

Table 1: Price of Anarchy for LCFS-PR and FCFS

	$\sum_i U_i^*$	$\mathrm{E}[\mathbf{W}]$	PoA
LCFS-PR	-2.230	-2.129	2.094
FCFS	-1.925	-2.129	1.808

Note: Approximated PoA values for a queueing game with $\eta = 2$, $V(t, d) = -d^{0.5}(d - t)^{0.8}$, S being exponentially distributed with rate $\mu = 1$, and $\Delta = 0.1$, $\epsilon = 0.001$ under the LCFS-PR and FCFS service discipline, respectively.

---

[5]Breinbjerg (2017) considers an almost identical queueing game to the present paper with the exception that customers arrive to a FCFS queueing system where they are not allowed to arrive before the opening time. However, this assumption does not affect the equilibrium properties in the LCFS-PR queueing game since customers will never choose to arrive before the opening time. Hence, the present queueing game can be directly compared to the FCFS queueing game.

5 Conclusion

We have examined the strategic choices of arrival times into a queueing system that employs the LCFS-PR service discipline, where each customer wants to complete service as early as possible while spending a minimum amount of time in the queue. Our main contributions include (1) establishing the existence and uniqueness of a strategy that constitutes a symmetric Nash equilibrium, and (2) numerically demonstrating an example for which the LCFS-PR service discipline induces lower social efficiency than that induced by a similar queueing game that employs the FCFS discipline.

The symmetric equilibrium in our LCFS-PR queueing environment prescribes a (mixed) arrival time strategy that represents a continuous and strictly increasing probability distribution that extends over a bounded interval of time. We note that these equilibrium properties are qualitatively similar to that established by Platz and Østerdal (2017) for the Last-Come First-Service discipline. While we restricted attention to queueing environments in which customers could arrive before the opening time and had preferences over early service completion and waiting time, the basic approach of the constructive proof of existence carries over to variants of such assumptions.[6]

Our numerical results for the symmetric equilibrium suggest that the LCFS-PR service discipline provides incentives for arrival times which lead to higher social inefficiencies compared to those provided by the FCFS discipline. This is in opposition to the result that FCFS constitutes the lower bound of social efficiency (Platz and Østerdal, 2017). A possible explanation for why the LCFS-PR discipline provides lower social efficiency in our queueing environment, may be the additional inefficiency caused by the property of preemption. As any newly arrived customer may preempt the service progress of another customer, the customers must arrive in equilibrium according to a probability distribution that extends over an "excessively" large interval of time in order to mitigate the expected disutility of being preempted after arrival. A further and more comprehensive study of the impact of preemption on social inefficiency propose an interesting challenge for future research.

Finally, the LCFS-PR queueing game may be further extended in other important directions. One is to extend the equilibrium analysis to allow for asymmetric solutions. This was recently done for a FCFS queueing game with no closing time, where it was shown that any equilibrium strategy is in fact symmetric and unique (Juneja and Shimkin, 2013). While such uniqueness does not necessarily hold for games with closing times, it may

---

[6]For example, customers may not only care about early service completion, but rather about the number of customers who arrived ahead of them. This is the case for concerts or flights with unmarked seats, where there is no actual penalty for being served late unless other customers have arrived and taken hold of the better seats. This disutility is modelled as an order penalty and was recently introduced by Ravner (2014).

extend to games with more general classes of preferences. Another direction is to comprehensively study the differences in social efficiency induced by the FCFS and LCFS-PR service disciplines. This could be done for various combinations of customer population sizes, service times distributions (e.g. non-exponential distributions with decreasing hazard rate, heavy tailed, etc.) and utility functions. A third direction is the consideration of heterogeneous (multiclass) customers. This has for example been considered lately in a FCFS queueing environment by Guo and Hassin (2012). A fourth direction is to consider target time preferences where customers have a target time at which they wish to get to their destination, and are accordingly penalized for being too early or too late. The lateness penalty considered in the present paper, is a special case of such preferences where the target time equals the opening time of the facility. A fifth direction is the consideration of queueing games with multiple servers. This has recently been considered by Haviv and Ravner (2015) who examine a multi-server system with no queue buffer, where customers are interested in maximizing the probability of obtaining service.

A   Appendix

A.1   Proof of Lemma 7

We start the proof by the following observation: A customer's waiting time when queueing under the LCFS-PR service discipline is independent of the queue length she faces upon arrival, since the discipline allows the customer to suspend the whole queue until after she completes her service. Thus, without loss of generality, we may say that the customer arrives at an idle server. A customer's waiting time in a LCFS-PR queue is then identical to the period of time between when the customer arrives to an empty system and when she departs while leaving behind an empty queue. A customer therefore only cares about the expected share of $\eta - 1$ customers that has arrived up until her arrival and their respective service time requirements.

To capture such a concept of waiting time, we start by introducing some notation. Let $\mathbf{A}$ denote the (random) arrival time of one of the two customers, and let $\{\mathbf{S}_j\}_{j\in\{1,2\}}$ be a sequence of (random) service time requirements, such that $\mathbf{S}_j$ is the service time of the jth customer to start service. For any customer $i \in \{1,2\}$, let $\mathbf{R}_i$ denote the (random) residual service time of customer i if preempted prior to service completion. Moreover, let $\mathbf{D}_i(t)$ denote the (random) departure time of customer i when she arrived at time t and the other customer arrived at $\mathbf{A}$. The departure time of customer i satisfies for each $t \geq 0$ the following sample path relation:

$$\mathbf{D}_i(t) = \begin{cases} \mathbf{S}_2 + t & \text{if } \mathbf{A} \leq t \\ \mathbf{1}_{\{\mathbf{S}_1+t\leq\mathbf{A}\}}(\mathbf{S}_1 + t) + \mathbf{1}_{\{\mathbf{S}_1+t>\mathbf{A}\}}(\mathbf{A} + \mathbf{S}_2 + \mathbf{R}_i) & \text{if } \mathbf{A} > t \end{cases}$$

Intuitively, the sample path above describes the cases for which customer i is preempted or not prior to service completion. That is, in the event of $[\mathbf{A} \leq t]$, customer i possibly preempts the customer already residing in the queue and completes her service after $\mathbf{S}_2$ time units. In the event of $[\mathbf{A} > t]$, customer i is the first to arrive at the system and is *possibly* preempted prior to service completion. That is, in the event that $[\mathbf{S}_1 + t \leq \mathbf{A}]$ then customer i departs the system at time $\mathbf{S}_1 + t$ before the other customer arrives at $\mathbf{A}$. Otherwise, if $[\mathbf{S}_1 + t > \mathbf{A}]$, then customer i is preempted and does not depart the system until the other customer has competed her service and i has competed her residual service requirement, i.e. she departs at time $\mathbf{A} + \mathbf{S}_2 + \mathbf{R}_i$.

Fix a strategy F and let $\mathbf{A} \sim F$. Moreover, let $\mathbf{S}_j \sim G$ for any j where G is the exponential probability distribution, such that for any $x \in \mathbb{R}$

$$G(x; \mu) = \begin{cases} 1 - e^{-\mu x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

and $g(x; \mu)$ denote the density of G at x. We then characterize the probability of the event $[\mathbf{D}_i(t) \leq d]$ conditional on $\mathbf{A}$, $\mathbf{S}_1$ and $\mathbf{S}_2$, which equals zero for any $d < t$, and otherwise is given by

$$\begin{aligned}
\Pr\{\mathbf{D}_i(t) \leq d \mid \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}\} &= \mathrm{E}\left[\mathbf{1}_{\{\mathbf{D}_i(t) \leq d\}} \mid \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}\right] \\
&= \mathbf{1}_{\{\mathbf{A} \leq t\}} G(d - t; \mu) \\
&\quad + \mathbf{1}_{\{\mathbf{A} > t\}}\left[\mathbf{1}_{\{\mathbf{S}_1 + t \leq \mathbf{A}\}} G(d - t; \mu) + \mathbf{1}_{\{\mathbf{S}_1 + t > \mathbf{A}\}} H(d - \mathbf{A}; 2, \mu)\right]
\end{aligned}$$

for any $0 \leq t \leq d$, where H denotes the Erlang probability distribution defined by

$$H(x; k, \mu) = \begin{cases} 1 - \sum_{m=0}^{k-1} \frac{1}{m!} e^{-\mu x} (\mu x)^m & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}.$$

Note that the memoryless property of the exponential distribution implies that the distribution of the residual service times does not depend on how long a customer has been in service prior to preemption since the remaining time is still probabilistically the same as in the beginning of her arrival time. That means that customer i's departure time is the sum of two independent, identically and exponentially distributed variables (or equivalently, Erlang distributed with shape 2) with location at time $\mathbf{A}$ in the case she is preempted. Consequently, the conditional probability $\Pr\{\mathbf{D}_i(t) \leq d \mid \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}\}$ is therefore independent of $\mathbf{S}_2$.

We next characterize D_i by marginalizing out the variables $\mathbf{A}$ and $\mathbf{S}_1$ such that

$$\begin{aligned}
D_i(d \mid t, F) &= \Pr\{\mathbf{D}_i(t) \leq d\} \\
&= \int_{\mathcal{S}(F)} \int_{\mathcal{S}(G)} \Pr\{\mathbf{D}_i(t) \leq d \mid \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}\}\, dG(s - t; \mu) dF(a) \\
&= \int_{-\infty}^{b} \int_{0}^{\infty} \Pr\{\mathbf{D}_i(t) \leq d \mid \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}\}\, g(s - t; \mu) ds\, dF(a)
\end{aligned}$$

for each $t \geq 0$ where all integrals are Lebesgue integrals and the supremum of the support $\mathcal{S}(F)$ is given by b. Since F might have points of discontinuity, let I_a denote the jump size of F at the point in time a, so

$$I_a = \begin{cases} F(a) - \lim_{s \uparrow a} F(s) & \text{if } F(a) - \lim_{s \uparrow a} F(s) > 0 \\ 0 & \text{otherwise} \end{cases}$$

for any $a \in \mathbb{R}$. Then we may express D_i as follows

$$D_i(d \mid t, F) = \sum_{a \leq b} I_a \int_0^\infty \Pr\{\mathbf{D}_i(t) \leq d \mid \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}\} g(s - t; \mu) \mathrm{d}s$$

$$+ \int_{-\infty}^b \int_0^\infty \Pr\{\mathbf{D}_i(t) \leq d \mid \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}\} g(s - t; \mu) f_+(a) \mathrm{d}s \mathrm{d}a$$

We next insert the expression for $\Pr\{\mathbf{D}_i(t) \leq d \mid \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}\}$ and divide the expression in the two intervals of $(-\infty, t]$ and $(t, b]$, respectively, so

$$D_i(d \mid t, F) = \sum_{a \leq t} I_a G(d - t; \mu) + \int_{-\infty}^t f_+(a) G(d - t; \mu) \mathrm{d}a$$

$$+ \sum_{t < a \leq b} I_a \left(\int_0^a g(s - t; \mu) G(d - t; \mu) \mathrm{d}s + \int_a^\infty g(s - t; \mu) H(d - a; 2, \mu) \mathrm{d}s \right)$$

$$+ \int_t^b f_+(a) \left(\int_0^a g(s - t; \mu) G(d - t; \mu) \mathrm{d}s + \int_a^\infty g(s - t; \mu) H(d - a; 2, \mu) \mathrm{d}s \right) \mathrm{d}a$$

The claim of Lemma 7 now follows immediately once we note that $\int_0^a g(s - t; \mu) = G(a - t; \mu)$ and $\int_a^\infty g(s - t; \mu) = 1 - G(a - t; \mu)$. □

A.2 Numerical Procedure

We here present a numerical method to compute X_b for a given b. Note that V, η, μ, ϵ, Δ and b are exogenous inputs to the procedure:

1. Let $\mathcal{T}_\Delta^b = \{t \in \{0, 1, 2, \dots\} : t\Delta < b\}$ be a discretization of the interval $[0, b)$ wrt. Δ.

2. Let $X_{b,h}(s) = 1$ for all $s \geq b$ and all $h \in \mathbb{N}$

3. Compute $U_i(b, X_{b,h}) = \int_b^\infty V(b, d) \mathrm{d}D_i(d \mid b, X_{b,h})$ according to the expression of D_i in Lemma 7 (note that $U_i(b, X_{b,h})$ is the same for all $h \in \mathbb{N}$).[7]

---

[7]We can further reduce the expression of D_i in the context of the constructive procedure in Lemma 5. That is, for a given customer i the procedure considers the other customer's strategy to be such that only one jump occurs immediately after the point in time that customer i arrives. Thus, the D_i in this context can be expressed as

$$\lim_{s \uparrow t} D_i(d \mid s, F) = \lim_{s \uparrow t} F(s) G(d - t; \mu) + I_t H(d - t; 2, \mu)$$

$$+ \int_t^b f_+(a) \left[G(a - s; \mu) G(d - s; \mu) + (1 - G(a - s; \mu)) H(d - a; 2, \mu) \right] \mathrm{d}a$$

4. Let $h = 0$ and sequentially compute $X_{b,0}(t)$ for each $t \in \mathcal{T}^b_\Delta$ according to equation (2).

5. Assign $h := h + 1$ and sequentially compute $X_{b,h}(t)$ for each $t \in \mathcal{T}^b_\Delta$ according to equation (5).

6. If $X_{b,h}(t) - X_{b,h-1}(t) \leq \epsilon$ for all $t \in \mathcal{T}^b_\Delta$ then let $X_b = X_{b,h}$ and stop procedure

7. Else, go back to step (5) and begin the next iteration of h.

Remark 1 Although the considered queueing game allows customers to arrive on the real line, we restrict ourselves only to compute the image of $X_{b,h}$ at non-negative points in time. This is done to save computing time, as we know the equilibrium strategy constitutes a probability distribution with support on the non-negative real line.

References

Arnott, R., de Palma, A., and Lindsey, R. (1989). Schedule delay and departure time decisions with heterogeneous commuters. *Transportation Research Record*, 1197:56–67.

Arnott, R., de Palma, A., and Lindsey, R. (1993). A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *American Economic Review*, 83(1):161–179.

Breinbjerg, J. (2017). Equilibrium arrival times to queues with general service times and non-linear utility functions. *European Journal of Operational Research*, DOI: 10.1016/j.ejor.2017.03.010.

Breinbjerg, J., Sebald, A., and Østerdal, L. P. (2016). Strategic behavior and social outcomes in a bottleneck queue: experimental evidence. *Review of Economic Design*, 20(3):207–236.

de Palma, A. and Fosgerau, M. (2013). Random queues and risk averse users. *European Journal of Operational Research*, 230(2):313–320.

Glazer, A. and Hassin, R. (1983). ?/M/1: On the equilibrium distribution of customer arrivals. *European Journal of Operational Research*, 13(2):146–150.

Glazer, A. and Hassin, R. (1987). Equilibrium arrivals in queues with bulk service at scheduled times. *Transportation Science*, 21(4):273–278.

Guo, P. and Hassin, R. (2012). Strategic behavior and social optimization in markovian vacation queues: The case of heterogeneous customers. *European Journal of Operational Research*, 222(2):278–286.

Since H and G is everywhere continuous and have $\lim_{a \downarrow t} G(a - t; \mu) = 0$ and $\lim_{a \downarrow t} H(a - t; 2, \mu) = 0$, respectively. We also note that D_i is continuous and everywhere differentiable w.r.t. d. Let $D'_i(d \mid t, F) = \frac{d}{dd} D_i(d \mid t, F)$ be the derivative of D_i w.r.t. d, then $U_i(t, F) = \int_t^\infty V(t, d) D'_i(d \mid t, F) dd$.

Hassin, R. (1985). On the optimality of first come last served queues. *Econometrica*, 53(1):201–202.

Hassin, R. (2016). *Rational Queueing*. Chapman and Hall/CRC.

Hassin, R. and Kleiner, Y. (2011). Equilibrium and optimal arrival patterns to a server with opening and closing times. *IIE Transactions*, 43(3):820–827.

Haviv, M. (2013). When to arrive at a queue with tardiness costs? *Performance Evaluation*, 70(6):387–399.

Haviv, M. and Ravner, L. (2015). Strategic timing of arrivals to a finite queue multi-server loss system. *Queueing Systems*, 81(1):71–96.

Jain, R., Juneja, S., and Shimkin, N. (2011). The concert queueing game: To wait or to be late. *Discrete Event Dynamic Systems*, 21(1):103–138.

Juneja, S. and Shimkin, N. (2013). The concert queueing game: Strategic arrivals with waiting and tardiness costs. *Queueing Systems*, 74(4):369–402.

Platz, T. T. and Østerdal, L. P. (2017). The curse of the first-in-first-out queue discipline. *Games and Economic Behavior (forthcoming)*.

Rapoport, A., Stein, W. E., Parco, J. E., and Seale, D. A. (2004). Equilibrium play in single-server queues with endogenously determined arrival times. *Journal of Economic Behavior and Organization*, 55(1):67–91.

Ravner, L. (2014). Equilibrium arrival times to a queue with order penalties. *European Journal of Operational Research*, 239(2):456–468.

Seale, D. A., Parco, J. E., Stein, W. E., and Rapoport, A. (2005). Joining a queue or staying out: Effects of information structure and service time on arrival and staying out decisions. *Experimental Economics*, 8(2):117–144.

Stein, W. E., Rapoport, A., Seale, D. A., Zhang, H., and Zwick, R. (2007). Batch queues with choice of arrivals: Equilibrium analysis and experimental study. *Games and Economic Behavior*, 59(2):345–363.

Verhoef, E. (2003). Inside the queue: Hypercongestion and road pricing in a continuous time - continuous place model of traffic congestion. *Journal of Urban Economics*, 54(3):531–565.

Vickrey, W. S. (1969). Congestion theory and transportation investment. *American Economic Review*, 59(2):251–260.