ARTICLE IN PRESS
JID: EOR [m5G;March 24, 2017;13:15]
European Journal of Operational Research 000 (2017) 1–11
Contents lists available at ScienceDirect
European Journal of Operational Research
journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

Equilibrium arrival times to queues with general service times and non-linear utility functions☆

Jesper Breinbjerg

Department of Business and Economics, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

ARTICLE INFO

Article history:
Received 12 October 2016
Accepted 2 March 2017
Available online xxx

Keywords:
Queueing
Strategic arrival times to a queue
Non-cooperative queueing games,

ABSTRACT

We examine a non-cooperative queueing game where a finite number of customers seek service at a bottleneck facility which opens at a given point in time. The facility serves one customer at a time on a first-come, first-serve basis and the amount of time required to service each customer is identically and independently distributed according to some general probability distribution. The customers must individually choose when to arrive at the facility, and they prefer to complete service as early as possible, while minimizing the time spent waiting in the queue. These preferences are captured by a general utility function which is decreasing in the waiting time and service completion time of each customer. Applications of such queueing games range from people choosing when to arrive at a grand opening sale to travellers choosing when to line up at the gate when boarding an airplane. We develop a constructive procedure that characterizes an arrival strategy which constitutes a symmetric Nash equilibrium, and we show that there exists at most one symmetric equilibrium. We accompany the equilibrium characterization with numerically computed examples of several symmetric equilibria induced by a non-linear utility function.

1. Introduction

This paper considers the following scenario: a number of customers want to acquire a good or service supplied by a bottleneck facility. The facility is initially closed for service but opens at a commonly known point in time. For the customers to acquire the good or service, they must simultaneously and independently choose a point in time where they submit a request for acquisition. We shall refer to such a submission as an *arrival*. If the number of arrivals at some point in time exceeds the capacity of the facility, then a queue is created and customers will have to wait in line. The time at which each customer eventually acquires the good or service does not only depend on her own arrival time, but also the arrival times of others. As customers might have preferences over outcomes of queueing, the arrival time decision constitutes a non-cooperative *queueing game*.

There exist many real-life applications of such queueing games. These range from physical queueing situations, e.g. customers choosing when to arrive at a grand opening sale, or travellers choosing when to line up at the gate when boarding an airplane, to virtual queueing situations, e.g. customers choosing when to phone a call center during opening hours, or foodies choosing when to join an online queue for table reservations at a popular gourmet restaurant.

The strategic choice of arrival times has been extensively studied since the initial formulation of the bottleneck model by Vickrey (1969). A sizable part of this growing body of literature is summarized in a recent survey (Hassin, 2016). The central approach to examine the strategic behavior is to characterize the equilibrium arrival times, presuming that self-optimizing customers make strategic decisions about their arrival. Given the wide range of queueing game applications, various queueing games have been studied which differ in characteristics such as the number of customers, customer preferences over outcomes, the service time required to process a customer, (dis)allowance of arrivals prior to the opening time, etc. In regards to customer preferences, a natural issue to consider is the desire to avoid congestion. This disutility is typically modeled as a *waiting time penalty* and was first considered by Glazer and Hassin (1983). They investigated the equilibrium arrival times of a (random) number of homogeneous customers who wish to minimize their individual

☆ The author thanks Lars Peter Østerdal, Refael Hassin, Moshe Haviv, Liron Ravner, Nahum Shimkin and an anonymous referee for valuable comments and suggestions. The author also appreciates fruitful discussions with conference and seminar participants at Tel Aviv, Tokyo, Paris and Odense. The financial support from the Danish Council for Independent Research – Social Sciences (Grant ID: DFF-1327-00097) is gratefully acknowledged. Lastly, a special thanks to Refael Hassin and the School of Mathematical Sciences at Tel Aviv University for the warm hospitality provided during my research visit.
E-mail address: jnb@sam.sdu.dk

http://dx.doi.org/10.1016/j.ejor.2017.03.010
0377-2217/© 2017 Elsevier B.V. All rights reserved.

Please cite this article as: J. Breinbjerg, Equilibrium arrival times to queues with general service times and non-linear utility functions, European Journal of Operational Research (2017), http://dx.doi.org/10.1016/j.ejor.2017.03.010

waiting time when the service time requirement of each customer is exponentially distributed. They showed that, in the symmetric equilibrium, each customer arrives according to a continuous probability distribution that extends over a bounded interval of time. This model was further extended to games where the facility has pre-scheduled service times (Glazer & Hassin, 1987), and where customers cannot arrive prior to opening (Hassin & Kleiner, 2011). Moreover, theoretical predictions of related models have been compared to empirical findings in laboratory experiments, which provide support for the symmetric equilibrium solution (Rapoport, Stein, Parco, & Seale, 2004; Seale, Parco, Stein, & Rapoport, 2005; Stein, Rapoport, Seale, Zhang, & Zwick, 2007). However, customers may not only want to avoid waiting time but also be interested in completing their service at an early time. An example of this could be a commuter driving home from work. He wishes to avoid traffic but is not willing to stay at work until midnight in order to achieve this. This additional disutility is often modeled as a *lateness penalty* that increases the later one completes service. Such preferences were first studied by Jain, Juneja, and Shimkin (2011) who analyzed equilibrium arrival times of a continuum of heterogeneous customers. This preference structure was further extended to queueing games with other characteristics, such as systems employing the last-come first-serve rather than the canonical first-come first-serve queue discipline (Platz & Østerdal, 2017), general customer populations with exponentially distributed service times (Juneja & Shimkin, 2013), finite closing times of the facility (Haviv, 2013), and discrete arrival time choices (Breinbjerg, Sebald, & Østerdal, 2016).

While the above mentioned studies uncover many important insights into the strategic choices of arrival time in queueing games, they limit their attention to some rather restrictive assumptions regarding customer preferences and service time requirements. Specifically, the preferences are assumed to be represented by a linear function of its penalty attribute(s), and moreover, the service time requirement of each customer is often assumed to be either deterministic or exponentially distributed. Both assumptions greatly simplify the equilibrium analysis and allow for an explicit characterization of the equilibrium arrival times. However, these simplifications raise some important questions about the extendability of the equilibrium solution. From a management perspective, it is important to understand and predict the arrival behavior within a more general class of queueing games in order to anticipate the arrival incentives induced by various queueing situations.

This paper extends the equilibrium analysis to queueing games with more general classes of customer preferences and service times. Specifically, we focus on queueing games where a finite number of customers, with identical preferences composed of waiting and lateness penalties, arrive to a single-server facility that opens at a given point in time. We shall also restrict attention only to a facility that serves customers on a first-come, first serve (FCFS) basis with no closing time and does not allow customer arrivals prior to the opening. For this class of queueing games, we provide a constructive procedure for finding a symmetric equilibrium for any game and show that the symmetric equilibrium is in fact unique. We note that the basic concept of the constructive procedure carries over to other variants of queueing games that modify the assumptions of closing time, early arrivals, and the priority order discipline.[1] We further show that the symmetric equilibrium is characterized by a continuous probability distribution that extends over a bounded interval of time, admits a positive probability of arrival at the opening time, and has an open interval of time immediately after opening where there are no arrivals. Lastly, we provide numerically computed examples of the symmetric equilibrium induced by a non-linear utility function.

The paper is organized as follows: Section 2 introduces the queueing game and model assumptions. Section 3 first defines the relevant notion of the equilibrium solution (Section 3.1), and thereafter establishes some preliminary results (Section 3.2) used to formulate a constructive procedure for findings such an equilibrium (Section 3.3). Subsequently, some properties of the equilibrium solution are presented and proved (Sections 3.4 and 3.5). Section 4 presents some numerical computations of the equilibrium which have been computed via a discretized version of the constructive procedure. We conclude the paper in Section 5 with a brief summary and future research directions. To facilitate an intuitive presentation of the paper, proofs that require technical notation for the transient queueing dynamics are relegated to the Appendix.

2. A queueing game[2]

A finite set of customers $\mathcal{N} = \{1, 2, \ldots, \eta\}$, $\eta \geq 2$ must obtain service by a single-server facility. The facility admits customers for service within the time interval $\mathbb{R}_0^+ = [0, \infty)$ such that the facility opens for admission at time 0 and does not close before all customers have been served. We assume that the facility serves one customer at a time according to a work-conserving FCFS regime. Moreover, we assume that the time required for each customer to complete her service is independent and identically distributed according to the cumulative distribution function S which is absolutely continuous, strictly increasing and has finite moments. If several customers arrive at exactly the same time, then they are admitted in a uniformly randomized order.

Strategy of arrival. Each customer $i \in \mathcal{N}$ independently chooses her time of arrival according to the same (*mixed*) *strategy* F which represents a cumulative probability distribution (cdf) that assigns to each point in time $t \in \mathbb{R}_0^+$ the probability $F(t)$ that a customer has arrived by time t. Let $\mathcal{S}(F)$ denote the *support* of strategy F which is the smallest closed set of probability 1, namely $\int_{\mathcal{S}(F)} \mathrm{d}F(t) = 1$.

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 for good. Let D_i denote the ex-ante *cumulative departure time distribution* for any customer i such that $D_i(d|t, F)$ is the probability that i has departed the system by time $d \in \mathbb{R}_0^+$ given that she arrived at time $t \in \mathbb{R}_0^+$ 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 FCFS is work-conserving. Note also that $D_i(d \mid t, F) = 0$ for all $d \leq t$. An explicit expression of D_i is given in Appendix A using standard queueing relations.

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 denote a utility function such that the value $V(t, d)$ is 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 at 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 $\lim_{t\to\infty} V(t, t) = -\infty$.

[1] Breinbjerg and Østerdal (2017) derive equilibrium arrival times under the Last-Come, First-Serve Preemptive-Resume (LCFS-PR) discipline and numerically compare these with those established in the present paper under the FCFS discipline.

[2] To facilitate an intuitive formulation of game, we let script letters denote sets or collections, capital letters denote functions, small letters denote set-elements or generic variables, greek letters denote exogenous parameters, and capital bold letters denote random variables.

Please cite this article as: J. Breinbjerg, Equilibrium arrival times to queues with general service times and non-linear utility functions, European Journal of Operational Research (2017), http://dx.doi.org/10.1016/j.ejor.2017.03.010