Operations Research Letters 45 (2017) 488–492

Contents lists available at ScienceDirect

Operations Research Letters

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

On threshold routing in a service system with highest-bidder-first and FIFO services

Tejas Bodas [a,*], D. Manjunath [b]

[a] LAAS CNRS, France
[b] IIT, Bombay, India

ARTICLE INFO

Article history:
Received 20 February 2016
Received in revised form 28 June 2017
Accepted 20 July 2017
Available online 2 August 2017

Keywords:
Routing
Wardrop equilibrium
Admission price
Bidding
Priority scheduling
FIFO

ABSTRACT

In this paper, we consider a two server system serving heterogeneous customers. One of the server has a FIFO scheduling policy and charges a fixed admission price to each customer. The second queue follows the highest-bidder-first (HBF) policy where an arriving customer bids for its position in the queue. Customers make an individually optimal choice of the server and for such system, we characterize the equilibrium routing of customers. We specifically show that this routing is characterized by two thresholds.

1. Introduction

Consider two make-to-order firms manufacturing an identical product. Upon receiving an order, the firms must assemble the product and deliver it to the customer ordering it. Each firm can assemble only one quantity at a time and the time taken to assemble the product need not be deterministic. During an assembly of a product, if there are more orders being placed by other customers, then these orders have to be fulfilled by the firm by suitably scheduling the subsequent orders. The two firms differ in their pricing strategy and the scheduling policy for choosing subsequent orders. One of the firm charges a fixed admission price for the product and maintains a FIFO scheduling discipline. The second firm employs a bidding policy where subsequent customers place a bid and their queue position in the schedule is proportional to the bid placed. The customers that order the product may differ in their cost for unit delay and are hence sensitive to the delay in receiving the product. When placing an order, the customer does not know the number of pending orders but may be informed about the service rate and the arrival rate for the orders. When ordering the product, the customers have to decide which firm to choose and if they choose the bidding firm, then what is the optimal bid to be made such that the cost of obtaining the product (the sum of the monetary and the delay cost) is minimized. Motivated by this problem, our interest is to characterize the equilibrium choice of the firm made by the heterogeneous customers.

Applicable to more general setting, a formal description of the problem considered in this paper is as follows. Consider a two server service system with customers arriving according to a homogeneous Poisson process. The customers are heterogeneous in their cost for unit delay. The service system consists of two servers and the customers are required to obtain service at one of these two servers. There is no dispatcher available to route the customers and hence each arriving customer has to make an individually optimal queue join decision. Each server in the service system has an associated queue and the two queues differ from each other in their scheduling policy. One of the queue has the standard FIFO scheduling policy and to monetize the offered service, it charges a fixed admission price to its customers. The other queue has a non-preemptive priority scheduling discipline where after the current service completion, a customer with the highest priority level is next chosen for service from the pool of customers waiting in the queue. The priority of a customer in this queue is determined by the bid paid by each arriving customer. Naturally, a higher bid corresponds to a higher priority in the queue. Such a scheduling policy is also known as the highest-bidder-first (HBF) policy and was introduced by Kleinrock [4]. In this paper, our primary interest is to characterize the equilibrium routing satisfying the Wardrop conditions [6] and determine the bidding decision made by those customers choosing the HBF server.

Such a system with parallel HBF and FIFO services was first analyzed in [1]. To investigate the effect on the revenue from

* Correspondence to: LAAS CNRS, 7 Avenue du Colonel Roche, 31400 Toulouse, France.
E-mail addresses: tbodas@laas.fr (T. Bodas), dmanju@ee.iitb.ac.in (D. Manjunath).

http://dx.doi.org/10.1016/j.orl.2017.07.008
0167-6377/

an HBF server, a free FIFO service was introduced in the system. Further a minimum bid was made mandatory for those choosing the HBF server. For such a system, the equilibrium routing and bidding strategy was analyzed in [1]. It was shown that the Wardrop equilibrium routing is characterized by a single threshold and customers with delay sensitivity (cost per unit delay) above the threshold choose the HBF server while the rest choose the FIFO server. Two scenarios were considered for modeling the system; in the first scenario, a free FIFO server was added in parallel to an existing HBF server. In the second scenario, the total service capacity was shared between the HBF and the FIFO server. Assuming that the customers cannot balk, it was shown that an addition of a free FIFO server decreases the system revenue. On the contrary, with the help of numerical examples, it was conjectured that sharing capacity with a FIFO server improves the revenue from the HBF server. For a summary on queues with HBF server and other similar queueing models with Wardrop equilibrium, refer the recent book by Hassin [3].

The primary difference between the system model considered in this paper and that of [1] is as follows. We assume that the FIFO server is not free but in fact comes with an admission price. This assumption makes the model more naturally applicable to a variety of revenue based service systems such as the above example for make-to-order firms. We relax the assumption of a minimum bid and analyze the equilibrium routing and bidding rule for this problem. This analysis is the primary objective of the paper. We begin by analyzing whether a single threshold routing function as in [1] satisfies the equilibrium routing conditions. To our surprise, this is not the case. We then check for the threshold routing where customers with sensitivity above a threshold choose the FIFO server while the rest choose the HBF server. We show that such a candidate for equilibrium routing also does not satisfy the necessary conditions for Wardrop equilibrium. In our main result, we prove that under (Wardrop) equilibrium routing, when customers use both servers, there exist two thresholds; customers with sensitivity between the two thresholds choose the FIFO server while the rest choose the HBF server. To the best of our knowledge, the result is novel and has a useful insight. While the 'middle class' of the population (based on their sensitivities) choose the FIFO service, the remaining customers (specifically those with high and low delay sensitivity) choose the HBF server.

The rest of the paper is organized as follows. In Section 2, we shall introduce the notation and recall some preliminary results from [1]. We prove our main results in Section 3 and this is followed by a discussion summarizing the results.

2. Notation and preliminaries

In this section, we shall introduce the basic notation and recall some preliminary results from [1]. We assume as in [1] that the customers arrive to the service system according to a homogeneous Poisson process of rate λ. Service times of customers are i.i.d. random variables with distribution $G(\cdot)$ and unit mean. Each arriving customer has an associated parameter β which is a realization of the random variable $\boldsymbol{\beta}$, $0 \leq a \leq \boldsymbol{\beta} \leq b < \infty$. $\boldsymbol{\beta}$ represents a customers cost per unit delay. Let $F(\beta)$ denote the distribution of $\boldsymbol{\beta}$ which is assumed to be absolutely continuous in (a, b). We also call $\boldsymbol{\beta}$ as the type of the customer and call $F(\beta)$ as the type profile.

The first server uses the non preemptive HBF discipline and serves at rate μ_1. The second server uses the FIFO discipline and serves at rate μ_2. Customers choosing the HBF server will have to place a bid before joining its queue while those choosing the FIFO server have to pay a fixed admission price denoted by c. We will assume that all arrivals will have to receive service from one of the two servers and they cannot balk. Thus an arriving customer now has to make the following decisions on arrival; which server to use, and, if it chooses the HBF server, then the value of its bid. As in [1], we assume oblivious decisions and let $p(\beta) : [a, b] \to [0, 1]$ denote the probability that a customer of type β chooses the FIFO server. Further, let $X(\beta)$ be the equilibrium bid if such a customer chooses the HBF server. For a preliminary analysis of the HBF queue, refer [4]. Lui [5] and Glazer and Hassin [2] were the first to consider the case with heterogeneous customers (characterized by $\boldsymbol{\beta}$) and have determined the equilibrium bidding function $X(\beta)$. The function $X(\beta)$ determines the optimal value of the bid to be made by a customer of type β such that the sum of the bid and the expected waiting cost in the queue is minimized. Specifically, it was shown that $X(\beta)$ is given by

$$X(\beta) = \int_0^{\beta} \frac{2\rho W_0 y}{(1-\rho+\rho F(y))^3}\, dF(y) \tag{1}$$

where ρ denotes the traffic intensity, $F(\cdot)$ denotes the underlying distribution of $\boldsymbol{\beta}$ and W_0 denotes the expected waiting time in the HBF server added to that of an arriving customer due to the residual service time of an existing customer. This is given by,

$$W_0 = \frac{\lambda}{2}\int_0^{\infty} \tau^2 dG(\mu\tau)$$

where λ and μ denote the arrival rate of customers and the service rate of the HBF server respectively. It was further shown that for a customer of type β, its expected waiting time $W(\beta)$ is given by

$$W(\beta) = \frac{\mu^2 W_0}{(\mu - \lambda(1-F(\beta)))^2}. \tag{2}$$

Now, for our system with HBF and FIFO service, for a given $p(\beta)$, it is easy to see that the arrival rate to the FIFO server is $\lambda_2 := \lambda\int_0^{\infty} p(\beta)dF(\beta)$ while the arrival rate to the HBF server is $\lambda_1 := \lambda - \lambda_2$. Let $\rho_i := \lambda_i/\mu_i$. A customer of type β that chooses the HBF server experiences a bid-dependent waiting time that will be denoted by $W_1(\beta)$ where

$$W_1(\beta) = \frac{\mu_1^2 W_0}{(\mu_1 - \lambda(1-F_1(\beta)))^2}$$

and where $F_1(\cdot)$ denotes the type profile of customers choosing HBF. The customers choosing the FIFO server experience an expected waiting time denoted by $W_2(\lambda_2)$. Continuing with the notation of [1], let $D_1(\beta) := W_1(\beta) + \frac{1}{\mu_1}$ and $D_2 := W_2(\lambda_2) + \frac{1}{\mu_2}$ be the expected sojourn times in, respectively, the HBF and the FIFO servers. Since the FIFO queue is an $M/G/1$ system, we have $W_2(\lambda_2) = \frac{W_0}{1-\rho_2}$. Refer Fig. 1 for an illustration of the system model.

In this paper, the primary interest is to obtain the equilibrium strategy henceforth denoted by $(p^E(\beta), X^E(\beta))$. Note that $X^E(\beta)$ needs to be determined for only those customers that under equilibrium decide to join the HBF queue. Clearly, the system considered is non-atomic and all customers choose individually optimal strategies. The equilibrium attained is a Wardrop equilibrium that was first described in [6] and used extensively in transportation systems. The Wardrop equilibrium routing condition on $p^E(\beta)$ for all β is that

$$p^E(\beta) \geq 0 \quad \text{implies that} \quad c + \beta D_2 \leq X^E(\beta) + \beta D_1(\beta). \tag{3}$$

Further $0 < p^E(\beta) < 1$ implies $c + \beta D_2 = X^E(\beta) + \beta D_1(\beta)$.

The following theorem recalls the equilibrium strategy $(p^E(\beta), X^E(\beta))$ for the system model considered in [1]. The key difference between the models is that the FIFO server in [1] charges no admission price and a customer joining the HBF server is required to pay a minimum bid M.