



ELSEVIER

Operations Research Letters 17 (1995) 41–45

**operations
research
letters**

Equilibrium strategies for queues with impatient customers

Refael Hassin^{a,*}, Moshe Haviv^b

^a Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel

^b Department of Econometrics, The University of Sydney, Sydney, NSW 2006, Australia and Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel

Received 15 September 1993; revised 1 July 1994

Abstract

We consider a memoryless queue in which the reward of service completion for an individual reduces to zero after some time. Customers, while comparing expected holding costs and the rewards have to decide if to join the system at all and if they do when to renege. We show that a unique Nash equilibrium exists in which each of the customers joins with some probability and reneges as soon as the reward is zero.

Keywords: Balking and reneging, Queues, Nash equilibrium

1. Introduction

We consider a single server queueing model in which the arrival process is Poisson with rate λ , service is exponential with expected value $1/\mu$, and the service order is first-come first-served. Each customer incurs a holding cost of C per unit of time in the system (in queue or in service). Our model deviates from the common one with respect to the value of service. We assume that this value is $R > 0$ if service is completed within T units of time of the customer's arrival. If service is completed later, it has zero value. We assume that a customer does not know how many customers are in front of him in the system. Moreover, we assume that he does not even know whether his demand is being processed at any given time. This property is common in queues where customers do not actually "stand in line" but only submit their request for service and wait until the end of service is announced.

When a customer's need for service arises (that is, upon his *arrival*) the customer has the option of not joining the queue at all and thus *balk* without ever returning. If however, he decides to join, he still has the option of leaving the system later at any time if his service has not been completed until then (this is the option of *reneging*). Since the customer obtains no new information while waiting, until his service is completed, we may assume that he selects a deadline, say τ , such that if his service does not end within τ units of time of his arrival, then he reneges and leaves the system for good.

* Corresponding author.

Customers are to decide (1) whether to join the system or not and (2) what reneging deadline to pick. We will investigate this model under the assumption that a fraction p of the population of customers decides to join the queue arrival, and the distribution of the reneging deadline is $F(\cdot)$. Since the *utility* of a customer (measured as zero for balking customers, and as the expected service value minus the expected holding cost for a joining customer) depends on what others do, we look for pairs $(p, F(\cdot))$ associated with a Nash equilibrium strategy. In the case that $F(\cdot)$ implies reneging after a deterministic waiting time of τ , the model will be denoted by (p, τ) .

The paper is organized as follows. Some relevant properties of the (p, τ) model are derived in Section 2. In Section 3 we show that there is a unique Nash equilibrium of the form (p, T) , namely once a customer joins he stays until his service ends or until he was held for T units of time (whatever comes first). An equation for p is derived. In Section 4 we consider social optimization and in Section 5 we conclude with a short discussion.

2. The (p, τ) model

We first observe that in the (p, τ) model, because of the first-come first-served service mechanism, it is not possible that a customer reneges while waiting. Therefore, a customer either completes service or reneges while in service, but he never reneges while waiting. Under these circumstances, for $0 < x \leq \tau$, let $f_\tau^p(\cdot)$ be the density function of the stationary waiting time (excluding service) of a customer who joins the system. Based on [3, 8] (see also [6]), we conclude that if $\mu > \lambda p$ then

$$f_\tau^p(x) = P_0 \lambda p e^{-(\mu - \lambda p)x}, \quad 0 < x \leq \tau, \quad (1)$$

where P_0 is the probability of the empty queue (and hence of no waiting before service commencement). Since

$$P_0 + \int_{0+}^{\tau} f_\tau^p(x) = 1$$

we conclude that

$$P_0 = \frac{\mu - \lambda p}{\mu - \lambda p e^{-(\mu - \lambda p)\tau}}.$$

The corresponding distribution function is

$$F_\tau^p(x) = \frac{P_0}{\mu - \lambda p} \left[\mu - \lambda p e^{-(\mu - \lambda p)x} \right], \quad 0 \leq x \leq \tau. \quad (2)$$

Lemma 2.1. *In the (p, τ) model, the waiting time for one who joins the queue, is with an increasing failure rate (IFR) distribution.*

Proof. We have to show that the hazard function $f_\tau^p(x)/(1 - F_\tau^p(x))$ is with a positive derivative for $0 \leq x \leq \tau$. We omit the details of the straightforward differentiation but note that this derivative is indeed positive due to the fact that $P_0 - (1 - \lambda p/\mu) > 0$. This inequality is true since $1 - \lambda p/\mu$ is the probability that a similar system but without reneging is empty. Of course, this probability is smaller than P_0 . \square

Consider a customer who joins the system and never reneges while the rest follow the (p, τ) strategy. His holding time is then the convolution between the distribution stated in Eq. (2) and exponential distribution with parameter μ . Note that this distribution is without atoms and the nonnegative numbers constitutes its

support. Denote the resulting density function by $g_{\tau}^p(x)$ where $0 \leq x < \infty$, and let $G_{\tau}^p(x)$ be the corresponding distribution function. It is a simple exercise to show that

$$g_{\tau}^p(x) = P_0 \mu e^{-\mu x} + \int_{t=0+}^{\min\{x, \tau\}} f_{\tau}^p(t) \mu e^{-\mu(x-t)} dt, \quad 0 \leq x \leq \infty,$$

or that

$$g_{\tau}^p(x) = P_0 \mu e^{-(\mu x - \lambda p \min\{x, \tau\})}. \tag{3}$$

In particular, for $0 \leq x \leq \tau$,

$$G_{\tau}^p(x) = P_0 \frac{\mu}{\mu - \lambda p} \left[1 - e^{-(\mu - \lambda p)x} \right], \quad 0 \leq x \leq \tau. \tag{4}$$

Lemma 2.2. *The holding time of a customer who joins the system and never reneges while the rest follow the (p, τ) strategy is with an increasing failure rate distribution.*

Proof. The lemma follows immediately from Lemma 2.1 coupled with the fact that the service time is with a constant hazard rate. Specifically, it is well-known that the sum of two independent IFR random variables is also IFR. See, e.g., [1, p. 100]. (Alternatively, one could use directly the functions that appear in Eqs. (3) and (4).) \square

Let $U_{\tau}^p(t)$ be the expected utility of a customer who joins the line with a deadline of $t \leq T$ when the rest of the customers follow the strategy (p, τ) . Clearly,

$$U_{\tau}^p(t) = \int_{x=0}^t (R - Cx) g_{\tau}^p(x) dx - (1 - G_{\tau}^p(t)) Ct. \tag{5}$$

In particular, $U_{\tau}^p(0) = 0$.

Suppose now that all customers follow the (p, τ) strategy. It is clear that a necessary condition for a Nash equilibrium is that

$$\tau = \arg \max_{t > 0} U_{\tau}^p(t). \tag{6}$$

If $p < 1$ then in addition, it is required that

$$U_{\tau}^p(\tau) = 0, \tag{7}$$

Since zero is what one receives if he balks. Recall that in the case that a Nash equilibrium specifies a strict mixing between two (or more) pure strategies, then one is indifferent between each of these pure strategies. Finally, if $p = 1$, it is required that

$$U_{\tau}^p(\tau) \geq 0. \tag{8}$$

3. The Nash equilibrium

Lemma 3.1. *For any pair (p, τ)*

$$\arg \max_t U_{\tau}^p(t) \in \{0, T\}.$$

Proof. Reneging at time zero, namely never joining the system, is certainly a candidate for an optimal action. What the lemma says is that if zero is not the best response, then T is. Clearly, the optimal reneging time for an individual is smaller than or equal to T . Also, by Lemma 2.2, the longer the customer is in the system, the more likely he is to end service during the next (infinitesimal) unit of time. Since the reward function is constant in the interval $[0, T]$, if an individual did not renege during the time interval $[t - \Delta t, t]$ then he should not renege during the time interval $[t, t + \Delta t]$. \square

Lemma 3.1 leads to the following theorem which is our main result. The assumption that $R \geq C/\mu$ is made in order to avoid trivialities, since if $R < C/\mu$, balking is a dominant strategy for all customers.

Theorem 3.1. Suppose that $R \geq C/\mu$. Then, the pair (p, T) is a Nash equilibrium if and only if

$$\frac{R(\mu - \mu e^{-(\mu - \lambda p)T}) + \frac{\lambda p C T (\mu - \lambda p) e^{-(\mu - \lambda p)T} - C \mu (1 - e^{-(\mu - \lambda p)T})}{\mu - \lambda p}}{\mu - \lambda p e^{-(\mu - \lambda p)T}} \begin{cases} = 0 & \text{if } p < 1, \\ \geq 0 & \text{if } p = 1. \end{cases} \quad (9)$$

Moreover, p is unique.

Proof. Lemma 3.1 implies that under Nash equilibrium, a customer either balks or he joins and reneges later at time T . Hence, the only question left is what portion does what. Straightforward algebra indicates that the condition stated in the theorem is just a repetition of Eqs. (7) and (8) when the reneging time is T . Finally, the reasoning behind the existence of a unique p is as follows. Aiming for a contradiction, suppose there were two equilibrium solutions p_1 and p_2 with $p_1 < p_2 \leq 1$. The utility of one who joins the system is zero under p_1 and nonnegative under p_2 . However, under p_2 the system is more congested, leading to a smaller utility. This is certainly a contradiction. \square

For example when $\lambda = 0.9$, $\mu = 1.0$, $T = 5$, $C = 1$ and $R = 2$ we have found numerically, that the value of the Nash equilibrium strategy is $p = 0.654$.

Remark. It is possible to argue informally that the Nash equilibrium (p, τ) is also an *evolutionarily stable strategy* (ESS). See [9, p. 16] for a definition of an ESS. Under (p, T) with $p < 1$, an individual is indifferent between balking and joining and then reneging at T . Other reneging times will be strictly suboptimal. If they all switch to any other (q, T) strategy then strategy (p, T) seems to be a better response for an individual than (q, T) is. This is the condition required for a Nash equilibrium in order to become an ESS. The reason for that is as follows. Suppose $q > p$ then if all customers use (q, T) than the system is now more congested than under (p, T) , leading to a negative utility for one who joins the queue. Since from balking the utility is zero, an individual is now better off under strategy (p, T) for himself than under (q, T) . The opposite is the case when $q < p$. What we have here is a typical case in which if all deviate from a Nash equilibrium strategy in one direction, then it is optimal to go against the market and deviate to the opposite direction.

4. Social optimization

Assuming that customers either balk upon arrival or renege at time T , the social optimization question is in finding the joining probability that maximizes social welfare. Specifically, we look for p^* such that

$$p^* \equiv \arg \max_p \{ p R G_T^p(T) - p C W_T^p \}, \quad (10)$$

where W_T^p is the expected holding time under the (p, T) model for a customer who joins the system. Note that

$$W_T^p = T(1 - G_T^p(T)) + \int_{t=0}^T t g_T^p(t) dt.$$

The reasoning behind the definition of p^* is simple: $pG_T^p(T)$ is the proportion of customers who end up getting service and pCW_T^p is the expected holding cost per customer.

Let p^{**} be the Nash equilibrium probability of joining the system. It is clear that $p^* \leq p^{**}$. Specifically, suppose $p^{**} < 1$, then the utility of an individual under this strategy is zero. A better utility for an individual (and hence a better social welfare) can be achieved only under a less congested system. Thus, $p^* \leq p^{**}$. This is a typical case in the sense that self-optimization leads to a more congested system. Self-optimization ignores the externalities imposed by an arrival on the rest of the society. Of course, by levying the right toll (namely reducing R for the individual but not to the society) the resulting new p^{**} can coincide with (the old) p^* .

5. Discussion

The common assumption used in queueing models with reneging is that the distribution of the time until reneging occurs is predetermined. See, for example, [2, 3, 4] or [5]. In this paper we derived the reneging behavior of customers from basic assumptions on the structure of their cost functions. Specifically, we assumed that the waiting cost is linear (fixed cost per time unit) while the value of service is fixed for the first T units of time after arrival and zero afterwards.

As an interesting open problem for future research we see the analysis of similar models with other waiting costs and service value functions. For example, the value of service completed t time units after arrival may be a monotone nonincreasing function, $R(t)$. We expect that the solutions under these more general assumptions will be considerably more complex. In particular, uniqueness of the Nash equilibrium may not hold anymore.

A related, but not equivalent, model is the one where the service value is constant but the holding costs are not linear [7]. Also in this model one may be interested in computing an equilibrium strategy of joining and reneging.

References

- [1] R.E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, New York, 1975.
- [2] D.Y. Barrer "Queueing with impatient customers and indifferent clerks", *Oper. Res.* **5**, 644–649 (1957).
- [3] D.Y. Barrer "Queueing with impatient customers and ordered services", *Oper. Res.* **5**, 650–656 (1957).
- [4] D.J. Daley "General customer impatience in the queue GI/G/1," *J. Appl. Probab.* **2**, 186–205 (1965).
- [5] A.G. De Kok and H.C. Tijms, "A queueing system with impatient customers", *J. Appl. Probab.* **22**, 688–696 (1985).
- [6] B.V. Gnedenko and I.N. Kovalenko, *Introduction to Queueing Theory*, 2nd ed., Birkhauser, Boston, 1989.
- [7] R. Haji and G.F. Newell, "Optimal strategies for priority queues with nonlinear costs of delay", *SIAM J. Appl. Math.* **20**, 224–240 (1971).
- [8] I.N. Kovalenko, "Study of many-server queueing system with limiting holding time", *Ukrainian Math. J.* **12** (3) (1960).
- [9] J. Maynard Smith, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, 1982.