Effect Of in Queue Waiting on Decision Making
Gök kem Sarıyer

Abstract-- Strategic customers take their waiting time into consideration upon making decisions. While satisfactorily completed service increases utility of a customer, waiting to be served decreases this utility in a queueing type service system. In this system, waiting can be classified into two: in queue waiting and in service waiting. Unlike the literature, proposed model assumes waiting in a queue, while other customers are being served, differs from waiting in service where the former one decreases the utility of a customer more. This paper analyzes decision making of a customer under this assumption. Comparing decision of the customer under proposed assumption and literature assumption, this paper shows that in queue waiting differs in service waiting.

Index Term-- Strategic customer, M/M/1, in queue waiting, in service waiting, observable queues, unobservable queues

INTRODUCTION
Customers generally wait before they receive service in their daily life. Since time is a scarce resource as expressed in [8], this waiting decreases the utility level of the customer of the service he received. Thus a strategic customer takes this waiting into consideration upon taking his rational decision of joining the system or to balk.

In queueing literature, different factors are assumed to affect these joining or balking decisions of customers. As a first viewpoint, the longer the length of the queue or the waiting, the less willing are the customers to join the queue. That is, when the waiting cost in the system is high, customers decide not to join the queue, since this waiting neutralizes or suppresses the value of the service that they will receive. In other words, a strategic customer decides to join the queue as long as the queue length or the expected waiting time is not too long. [14] is the first known one which examines the individual joining strategies by an economic aspects of queueing. He concludes that when a customer observes the length of the queue before he takes his decision of whether or not to join, he decides based on a threshold strategy. Specifically, since the expected waiting cost of this new customer is increasing in the queue length, he only joins if the length of the queue is smaller than this threshold level. Equilibrium decisions of the customers when the queue length is not observable, where customers decide based on their expected waiting time, are analyzed in [9]. There are many other of them which follow the similar assumption (e.g: [4], [5], [7], [12], [16], [18]-[23]). For a detailed review see [11].

More recent view in queueing literature suggests that joining probabilities may not be decreasing in queue length. [6] shows, equilibrium joining strategies may emerge where joining probabilities increases locally in the length of the queue. Referring this equilibria as sputtering equilibria, [6] discusses that this equilibria exists not only for discrete but also with continuously distributed priors on the expected service time where this expected service time is positively correlated with the value of the service. With a similar view, [15] shows the purchase incidence is not monotonic in length of the queue. [3] discusses that consumers can join longer queues with a higher affinity. [10] experimentally shows that waiting times can be a signal about the increasing quality. [2] also suggests that service value is positively correlated with waiting time. [1], [13], and [17] are some of the other papers assuming that the value of the service increases in service time.

Combining these two perspectives, while waiting to be served decreases the utility of the customer, higher service times increases this utility level. Thus in this paper, instead of assuming a fixed unit waiting cost in the system, cost of waiting is classified into two parts. First one covers cost of waiting in queue, and the other one covers cost of waiting in service where unit cost of in queue waiting is exactly higher than the unit cost of in service waiting. Under this scenario, the individually optimal decision and the equilibrium decision of a strategic customer in observable and unobservable queues are respectively analyzed. These decisions are also compared with the ones given in [14] and [9] which assume fixed unit waiting cost in system. The first conclusion is cost of waiting in queue has higher effect on decision of individual in both observable and unobservable queues. Numerical results also suggest that the effect of differentiating these waiting costs is more evident in unobservable queues compared to observable queues.

Decision of Individual in Observable Queues
Individual decision of a strategic customer who tries to maximize his own utility function when he observes the length of the queue prior to taking his decision is first analyzed in [14]. Systems into the consideration are basic M/M/1 queues with Poisson distributed arrivals having rate λ and exponentially distributed service times having rate μ. To express the utility function of a strategic customer, reward R and unit cost C are used in monetary values denoting the value of the received service and the cost of waiting in [14]. In this paper, keeping all other assumptions as the same, unit waiting cost is differentiated into two. C₁ and C₂ are respectively used to denote unit cost of waiting in the queue, and the unit cost of waiting in service where C₁>C₂. Based on this modification, denoting with Uᵢ(.), utility function of a customer who finds i customers in the queue upon his arrival is re-derived as:

Uᵢ(i+1) = R - C₁ \frac{i}{μ} - C₂ \frac{i}{μ} \quad (1)

where the second and third terms of equation (1) show total expected waiting cost in the queue and in the service respectively.
Lemma 1: \( U_o(i+1) \) is decreasing in \( i \).

Proof:

\[
\frac{d U_o(i + 1)}{d i} = - \frac{C_1}{\mu} < 0
\]

Thus result follows.

Since utility function given in equation (1) is decreasing in \( i \), and a strategic customer will only join if his utility function is positive, a threshold strategy is defined to express the decision of the customer. Denoting the threshold level of the customer with \( n_{ind}^i \), decision of the \( i \)th customer is given as:

\[
i = \begin{cases} 
\text{joins, if } n_{ind}^i \geq i \\
\text{does not join, o.w.}
\end{cases} \tag{2}
\]

Where

\[
n_{ind}^i = \frac{\mu - C_2}{C_1}
\]

In order to differentiate these unit costs properly and to compare the results with the fixed unit waiting cost case, equal total waiting costs are assumed.

\[
\frac{i C_1 + C_2}{\mu} = \frac{(i + 1) C}{\mu} \tag{3}
\]

While left side of equation (3) shows the total waiting cost under differentiated unit waiting costs, right side expresses this cost under fixed unit waiting cost. Derivation of \( C \) in terms of \( C_1 \) and \( C_2 \) is:

\[
C = \frac{i C_1 + C_2}{i + 1} \tag{4}
\]

Proposition 1: \( C \) is more affected by the change in \( C_1 \), compared to change in \( C_2 \).

Proof

\[
\frac{d C}{d C_1} = \frac{i}{i + 1} \geq \frac{d C}{d C_2} = \frac{1}{i + 1} \rightarrow i > 0
\]

So, result follows.

Proposition 1 shows that, decision of the individual is affected more by in queue waiting compared to in service waiting.

Proposition 2: Threshold level is lower when the waiting costs are differentiated compared to fixed unit waiting cost.

Proof

Using equation (4) expressing \( C \) in terms of \( C_1 \) and \( C_2 \), we need to show:

\[
\frac{R \mu - C_2}{C_1} \leq \frac{R \mu - C}{C} \tag{5}
\]

where left side of equation (5) shows the threshold level under differentiated unit waiting costs, and right side shows the threshold level under fixed unit waiting cost. Rearranging equation (5) we have:

\[
(R \mu - C_2)C \leq (R \mu - C)C_1 \rightarrow C(R \mu + C_1 - C_2) \leq R \mu C_1 \tag{6}
\]

Plugging \( C \) in terms of \( i, C_1, \) and \( C_2 \) in place,

\[
\frac{i C_1 + C_2}{i + 1} (R \mu + C_1 - C_2) \leq R \mu C_1 \rightarrow (C_1 - C_2)(i C_1 + C_2) \leq R \mu (C_1 - C_2) \tag{7}
\]

Since, \( C_1 > C_2 \rightarrow C_1 - C_2 > 0 \) we have:

\[
i C_1 + C_2 \leq R \mu \tag{8}
\]

Since condition given in equation (8) always holds, otherwise nobody joins, proof follows.

Proposition 2 is interpreted as: Customers are more willing to join the system when the unit waiting cost in queue and unit waiting cost in service are different.

Decision of Individual in Unobservable Queues

In this section, queue length is assumed as unobservable, thus customers cannot observe the queue before they decide. Besides, service rate and expected waiting time in the queue are assumed to known by customers. In unobservable queues customer decides based on an equilibrium strategy as given in literature (e.g. [9]). This strategy depends on the effective joining rate to the system \( \alpha \), where \( 0 \leq \alpha \leq 1 \). Denoting the utility function of the individual in unobservable case with \( U_u(\alpha) \), function is rewritten as:

\[
U_u(\alpha) = R - C_1 \left( \frac{\lambda \alpha}{\mu - \lambda \alpha} \right) - C_2 \frac{1}{\mu} \tag{9}
\]

Second term of equation (9) shows the total expected cost of waiting in the queue, and the third term shows expected cost of waiting in service in unobservable queues.

Lemma 2: \( U_u(\alpha) \) is concave in \( \alpha \).

Proof:

\[
\frac{d U_u(\alpha)}{d \alpha} = - \frac{C_1 \lambda \mu^2}{(\mu^2 - \mu \lambda \alpha)^2} \frac{d^2 U_u(\alpha)}{d \alpha^2} = - \frac{2 C_1 \lambda^2 \mu^3}{(\mu^2 - \mu \lambda \alpha)^3} < 0 \tag{10}
\]

Since the concavity follows as given in Lemma 2, the joining strategy of the individual based on the equilibrium joining probability, \( \alpha^{eq} \), is defined as in literature:
In order to write the fixed unit waiting cost $C$ as a function of $C_1$ and $C_2$, total expected waiting cost functions of two cases (fixed unit waiting cost, and differentiated waiting costs), when it is optimal for everybody to decide to join, will be equalized.

$$\frac{C}{\mu - \lambda} = \frac{\lambda C_1}{\mu (\mu - \lambda)} + \frac{C_2}{\mu} \rightarrow C = \frac{\lambda (C_1 - C_2)}{\mu} + C_2 \quad (12)$$

Left and right sides of equation (12) respectively show total expected waiting cost in fixed unit waiting cost and differentiated waiting cost cases.

**Proposition 3:** $C$ is increasing in $C_1$, $C$ is increasing in $C_2$ if $\lambda < \mu$.

**Proof**

$$\frac{dC}{dC_1} = \frac{\lambda}{\mu} > 0 \quad (13)$$

$$\frac{dC}{dC_2} = \left\{ \begin{array}{ll} 1 - \frac{\lambda}{\mu} & , \quad \text{if } \lambda < \mu \\ 0 & , \quad \text{if } \lambda \geq \mu \\ \end{array} \right. \quad (14)$$

Thus result follows.

The interpretation of Proposition 3 is as follows: The fixed unit waiting cost increases when the waiting unit cost in the queue increases. It is also increasing with the unit waiting cost in service if $\mu > \lambda$, since the server has enough capacity to capture all possible arrivals. However when $\mu \leq \lambda$, since the capacity of the server is not enough to serve all customers, then the unit fixed waiting cost is decreasing in unit cost of waiting in service.

**Proposition 4:** The joining probability of differentiated waiting costs is at least equal to or greater compared to fixed unit waiting cost case.

**Proof:**

Assume $a_{eq}$ and $a_{eq}^f$ respectively denote the equilibrium joining probability in differentiated waiting costs and fixed unit waiting cost cases. Based on the possible regions, the comparison between them is as follows:

$$a_{eq}^f, a_{eq} = \begin{cases} 0, & \text{if } R \leq \frac{C_2}{\mu} \\ \frac{\mu (R\mu - C_2)}{\lambda (C_1 - C_2 + R\mu)}, & \text{if } \frac{C_2}{\mu} < R \leq \frac{C_2}{\mu} - \frac{C_1\lambda}{\mu (\mu - \lambda)} + C_2 \\ 1, & \text{if } \frac{C_2}{\mu} - \frac{C_1\lambda}{\mu (\mu - \lambda)} + C_2 < R \end{cases} \quad (15)$$

In the $1^{st}$, $2^{nd}$, $4^{th}$ and $5^{th}$ regions result follows. We now compare the equilibrium joining probabilities in the $3^{rd}$ region, assuming $a_{eq}^f \leq a_{eq}$, and $C = \frac{\lambda (C_1 - C_2)}{\mu} + C_2$

$$a_{eq}^f \leq a_{eq} \rightarrow \frac{R\mu - C}{R\lambda} \leq \frac{R\mu^2 - C_2\mu}{\lambda (C_1 - C_2 + R\mu)} \rightarrow R\mu C_1 \leq C (C_1 - C_2 + R\mu) \quad (16)$$

Plugging $C$ value in place, inequality (16) is rewritten as:

$$R\mu C_1 \leq \left( \frac{\lambda C_1 - \lambda C_2 + C_2}{\mu} \right) (C_1 - C_2 + R\mu) \rightarrow (\mu - \lambda)(R\mu - C_2) \leq \lambda C_1 \quad (17)$$

The above condition holds in the $3^{rd}$ region, since otherwise it is optimal for all the customers to join the system. Thus the result follows.

The interpretation of Proposition 4 is similar to Proposition 2. Both of these results show that differentiating the waiting costs, positively affect the joining decision of the strategic customer in both of the observable and unobservable queues.

**NUMERICAL RESULTS**

In this section, numerical analysis corresponding both of the observable and unobservable queues is given.

Table I shows numerical examples for observable queues. $N$ shows the maximum number of customer in the queue that a new arriving customer can observe upon his arrival. Second column of the Table shows the corresponding fixed unit waiting cost given the values of $C_1$ and $C_2$. Case-1 and Case-2 are used to denote differentiated waiting cost, and fixed waiting cost cases. Third and fourth columns respectively show the percentage joining rates in Case-1, $PJR_{Case-1}$ and in Case-2, $PJR_{Case-2}$.
In a similar way, Table II represents numerical analysis corresponding unobservable queues. In the 3rd and 4th columns of the Table, $q_e^{eq}$ and $q_f^{eq}$ are used to denote the equilibrium joining probabilities of Case-1 and Case-2 respectively.
In the case, when the server spends time by providing service for the strategic customer setting in this paper.

The interpretation of Observation 1 shows that, the effect of differentiating unit waiting costs is clear in unobservable queues compared to observable queues.

In Table 1, in only 6 experiments between the total of 32, differentiating the unit waiting costs show higher joining percentage compared to the fixed waiting costs; 8/32=0.25. In Table 2, same ratio is: 26/32=0.81.

The theoretical analysis of the observable queues shows that, fixed unit waiting cost, C, is increased with the increase in queue and in service waiting costs, C1 and C2 where the effect of the former one is more evident compared to latter one. Additionally, joining percentage of the strategic customer when the waiting costs are differentiated is at least equal to or greater than the joining percentage of the fixed waiting cost case, since the threshold level is lower in the former one.

Similar analysis for the unobservable queues represents that, C is directly affected with the change in C1 where the relation between C and C2 depends on the server utilization, λ/µ. In addition, customers are more willing to join the system when the waiting costs are differentiated as seen in observable case.

Based on the numerical analysis, it is observed that the effect of differentiating waiting costs is more evident in unobservable queues, since decision of the customer highly depends on the cost of waiting in the queue in unobservable queues.

declaration of interest
All authors declare that there is no conflict of interest

REFERENCES


