# The Concert/Cafeteria Queueing Problem: A Game of Arrivals

Sandeep Juneja
School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai, India - 400 005
juneja@tifr.res.in

Rahul Jain
EE & ISE Departments
University of Southern California
Los Angeles, CA 90089
rahul.jain@usc.edu

## ABSTRACT

We introduce the concert or the cafeteria queueing problem: Fixed but a large number of users arrive into a queue which provides service starting at time 0. Users may arrive before 0. They incur a queued waiting cost $\alpha \cdot W$, where $W$ is the time to wait in the queue until service, and service time cost $\beta \cdot (t + W)$, where $t$ is the arrival time and $t + W$ is the total time until service. Each user picks a mixed strategy for arrival to minimize $E[\alpha W + \beta(t + W)]$. We analyze the system in an asymptotic regime and develop fluid limit for the resultant queueing system. The limiting system may be modeled as a non-atomic game for which we determine an equilibrium arrival strategy. In particular, we note that the equilibrium arrival strategy is to arrive uniformly between some $\tau_0 < 0$ and $\tau_1$ selected so that the queue is never empty. We note that larger the $\beta/\alpha$, larger the queue. Furthermore, we note that the 'price of symmetric anarchy' of this system equals 2. In addition to modeling queue formation at large concerts or cafeterias in certain settings, the model may be relevant more generally, for instance, in explaining queue formation in DMV offices at opening time, and at retail stores at opening time during peak shopping season.

## General Terms

Queueing, Game theory

## Keywords

Games of timing, Nash equilibrium

## 1. INTRODUCTION

In this paper we introduce the concert or the cafeteria queueing problem. Many people before going to a concert where a large queue may be anticipated choose their arrival time strategically. There is a cost to going very late as then the best seats may be taken. Going early may involve a large and annoying wait in the queue. Similarly, people may like to eat their lunch early as soon as the cafeteria opens for

lunch. However, for this very reason, the delays may be large initially. Similar trade-offs govern customer decision making in many queueing situations such as visiting a retail store on the day of a huge sale; going to the DMV office, to a movie theater, in communications and computer networks, etc. In this paper we make an initial attempt to study the strategic trade-offs faced by customers in such queueing situations.

In our model we assume that there are a large but fixed number of customers that need to be served by a server in a first come first serve manner. The server at the queue becomes active at a particular time. We allow the flexibility that the customers can choose to arrive before that time. Under the assumption that each customer implements a mixed strategy, i.e., selects her time of arrival as a sample from a probability distribution, we develop fluid limit for the resultant queueing system. This fluid limit offers a great deal of analytical simplification. In particular, we show that in the resulting limiting fluid system the arrival distribution can be modeled as a non-atomic game [10] for which we solve for a symmetric Nash equilibrium arrival distribution for each customer under linearity assumptions on costs associated with the time of service and with the time spent in the queue. We also show that under this symmetric equilibrium the queue never becomes empty until every customer has been served. Importantly, we show that more the customers value early service, the larger the resultant queue size is under this equilibrium. We also identify the cost of the global welfare solution associated with our game. This allows us to infer the interesting fact that the price of symmetric anarchy of this game equals 2 for all parameter values.

We note that in many settings, costs for a given customer may be modeled as a sum of a term proportional to the delay in the queue and another term proportional to the number of customers that have arrived before this customer. The latter differs from our earlier assumption of cost depending on the time at which service is received. This, for instance, is relevant in concert hall, movie theater etc. settings where more people have entered before a given customer, the worst the seats that may be available to that customer. We note that similar results hold under these modified assumptions as well.

Strategic models involving queues are not new. They were first considered in [9] using pricing as a means to queueing stability. Mendelson and Whang [8] introduced a stylized model for a single queuing service provider with multiple priority classes, and each user has private information about job parameters (such as delay cost and expected service time).

For such a model, they introduced an incentive-compatible priority pricing rule for the $M/M/1$ queue which is efficient as well (maximizes the social welfare). In [11], a general equilibrium model with congestion externality (i.e., queueing delays) is considered. They conclude that the competitive price is social welfare maximizing. A dynamic pricing model is considered in [7] with adaptive learning by the users. Various models of learning of congestion delays are considered: rational expectations, Markovian expectations and exponentially smoothed expectations and conditions of network stability are established. In [4], a *cost sharing* perspective is taken on sharing the total delay cost, and the Aumann-Shapley mechanism is used to determine individual user payments. Most of the work discussed above is summarized in the book [3]. In [6], a multi-server extension of the Mendelson-Whang model is considered. However, a competitive setting is considered where there are enough servers in the market, and the influence of each is negligible, i.e., the servers act as price takers. Thus, the existence of competitive equilibrium is established. While there has been a lot of work on studying pricing by queues to study stability, games of timing involving queues, where the arriving users choose time of arrival strategically are not so well-understood. The earliest such work is [2] which introduced a model where users choose strategically the time of their arrival into a ?/M/1 queue. However, they considered a discrete population model where each user is interested in minimizing its queueing delay, and service times were assumed to be exponential. In contrast, we consider a more general framework: Each user has a linear cost function of waiting delay as well as a linear cost function of the time of service of the user or the number of people who arrive before that user, a significant motivation for people in many scenarios for arriving early. Thus, our cost structure is more realistic in many settings. On the other hand, we consider a fluid approximation of the discrete population model. A related paper is [5] which addresses a problem with a similar motivation but the model is different. They consider discrete time, and users arriving at any time, may be served in any random order, and thus experience only average delay. Moreover, there is no queueing: All users that arrive in any time slot, are served in that time slot. In [13], Tian, Huang and Yang consider a related problem of equilibrium behavior of discrete population of morning commuters arriving to a mass transit system.

The organization of this paper is as follows: In Section 2 we spell out the mathematical framework for the queueing system that we consider. Here we conduct asymptotic analysis and derive the fluid limit of this queueing system. In Section 3, we show existence of a symmetric Nash equilibrium solution, and compute a price of symmetric anarchy. Finally, we end with a brief conclusion in Section 4.

## 2. MATHEMATICAL FRAMEWORK AND THE FLUID LIMIT

We consider a series of queueing systems indexed by $n$ that we analyze as $n \to \infty$. For the system $n$, we assume that there are $n$ customers, each of whom independently picks an arrival time as a sample from a distribution $F_n(\cdot)$ where $F_n(\cdot)$ denotes a distribution with support within $[-nT, nT]$ for a constant $T > 0$ that is sufficiently large (as specified later). The queue begins to serve at time 0. The service

time of each customer is an i.i.d process $(V_i : 1 \leq i \leq n)$ with rate $EV_i = 1/\mu$. We assume that $T > 1/\mu$ so that as $n \to \infty$, all the customers are served before time $nT$.

We first develop a fluid limit for this system. Some additional assumptions and notation is needed for this. We restrict $F_n$ to satisfy the limit

$$\sqrt{n}\left(F_n(nt) - F(t)\right) \to 0 \qquad (1)$$

uniformly on compact sets (uoc), where $F(\cdot)$ is a distribution function with support in $[-T, T]$. Thus, for instance, $F_n(\cdot)$ may correspond to a uniform distribution between $[-nT, nT]$ while $F(\cdot)$ corresponds to a uniform distribution between $[-T, T]$.

Let $A_n(t)$ denote the number of arrivals by time $t$ in system $n$. Similarly, for $t > 0$, let $S_n(t)$ denote the number of service completions in queue if the server is busy throughout the time $t$, i.e.,

$$S_n(t) = \sup\{m : \sum_{i=1}^{m} V_i \leq t\}.$$

Then, the queue length process $Q_n(t)$ for $t \leq 0$ equals $A_n(t)$, and for $t > 0$,

$$Q_n(t) = A_n(t) - S_n(B_n(t))$$

where $B_n(t)$ denotes the time that the queue has been busy between $[0, t]$ and

$$B_n(t) = \int_0^t 1(Q_n(s) > 0)ds.$$

For $t \leq 0$, we may re-express $Q_n(t)$ to equal $X_n(t)$ where, the 'net input process'

$$X_n(t) = nF_n(t) + (A_n(t) - nF_n(t)).$$

For $t > 0$,

$$Q_n(t) = X_n(t) + Y_n(t)$$

where,

$$X_n(t) = (nF_n(t) - \mu t) + (A_n(t) - nF_n(t)) - (S_n(B_n(t)) - \mu B_n(t))$$

and 'the regulator process'

$$Y_n(t) = \mu(t - B_n(t)).$$

It then follows (see Chen and Yao 2001) that for $t > 0$,

$$Y_n(t) = \sup_{0 \leq s \leq t} [-X_n(s)]^+$$

and

$$Q_n(t) = X_n(t) + \sup_{0 \leq s \leq t} [-X_n(s)]^+.$$

Let $V(m) = \sum_{i=1}^{m} V_i$. Then the workload at the queue at time $t > 0$ equals

$$Z_n(t) = V(A_n(t)) - B_n(t).$$

It equals $V(A_n(t))$ for $t < 0$.

We define the normalized values of arrival, service and related process as follows:

$$\bar{A}_n(t) = \frac{A_n(nt)}{n},$$

$$\bar{Z}_n(t) = \frac{Z_n(nt)}{n},$$

and

$$\bar{Q}_n(t) = \frac{Q_n(nt)}{n}.$$

For $t > 0$,

$$\bar{S}_n(t) = \frac{S_n(nt)}{n}.$$

and

$$\bar{B}_n(t) = \frac{B_n(nt)}{n}.$$

Then, under mild regularity conditions, $\bar{A}_n(t) \to F(t)$ and for $t > 0$, $\bar{S}_n(t) \to \mu t$ uniformly on compact sets. It then follows that the processes

$$(\bar{Q}_n, \bar{B}_n, \bar{Z}_n) \to (\bar{Q}, \bar{B}, \bar{Z}),$$

where for $t \leq 0$,

$$\bar{Q}(t) = \bar{X}(t),$$

where

$$\bar{X}(t) = F(t),$$

and for $t > 0$,

$$\bar{Q}(t) = \bar{X}(t) + \bar{Y}(t)$$

where,

$$\bar{X}(t) = F(t) - \mu t,$$

$$\bar{Y}(t) = \sup_{0 \leq s \leq t} [-\bar{X}(s)]^+,$$

$$\bar{Z}(t) = \bar{Q}(t)/\mu,$$

and

$$\bar{B}(t) = t - \bar{Y}(t)/\mu.$$

The proof for these results essentially follows, for instance, from the proof of Theorem 6.5 in [1], and is omitted.

REMARK 1. The arrival process $A(\cdot)$ has an interesting diffusion limit. To see this, define

$$\hat{A}_n(t) = \sqrt{n}(\bar{A}_n(t) - F(t)) = \frac{A(nt) - nF(t)}{\sqrt{n}}.$$

This may be re-expressed as

$$\hat{A}_n(t) = \sqrt{n}(\bar{A}_n(t) - F_n(nt)) + \sqrt{n}(F_n(nt) - F(t))$$

Then, since the latter term converges to zero, by the converging together theorem $\hat{A}_n(t)$ has the same distribution limit as

$$\sqrt{n}(\bar{A}_n(t) - F_n(nt)).$$

The latter has mean zero and variance $F_n(nt)(1 - F_n(nt))$ that converges to $F(t)(1 - F(t))$ the variance of a Brownian bridge evaluated at $F(t)$. Under mild restrictions it can be shown that the process $\{\hat{A}_n(t)\}$ converges to an appropriate time changed Brownian bridge. (Similar convergence is used in Kolmogorov-Smirnov test; see for instance [14]). Furthermore, it can be shown that the diffusion limit of the queueing process, wherever the queue is positive is a Gaussian process.

## 3. GAME OF ARRIVALS

Consider a tagged arrival $s$ in System $n$. Suppose that $W_n(t)$ denotes her expected waiting time in the queue if she arrives at time $t$ given that the arrival process of all the other customers is described by the CDF $F_n$. Further suppose that her cost structure is as follows: She incurs a cost $\beta(s)$ if she is served at time $s$. In addition, if she spends time $t$ in the queue then that costs her $\alpha t$. Suppose she samples her arrival (she selects a mixed strategy) from a probability distribution $G_n(\cdot)$ whose support is within $[-nT, nT]$. To ease the analysis we assume that $\beta(t) = \beta t$, that is, $\beta(\cdot)$ is a linear function. Then her expected cost is

$$C_{s,n}(G_n, F_n) := \int_{-nT}^{nT} (\beta(t + W_n(t)) + \alpha W_n(t)) \, dG_n(t) \tag{2}$$

and her problem is to find $G_n$ that minimizes this. This clearly corresponds to putting all the probability mass at times that minimize

$$\beta t + (\alpha + \beta) W_n(t).$$

This problem can be solved once we can evaluate $W_n(t)$ for each $t$, which in turn would depend on the probability distribution $F_n$ used by other customers to select their arrival times.

We analyze this problem asymptotically. Setting $\tilde{t} = t/n$, the equation (2) may be re-expressed as:

$$\int_{-T}^{T} (\beta(n\tilde{t} + W_n(n\tilde{t})) + \alpha W_n(n\tilde{t})) \, dG_n(n\tilde{t}). \tag{3}$$

Consider the case where $\frac{W_n(n\tilde{t})}{n} \to \bar{W}(\tilde{t})$ as $n \to \infty$ for some function $\bar{W}(\cdot)$, and $G_n(n\tilde{t}) \to G(\tilde{t})$ as $n \to \infty$ for some distribution function $G(\cdot)$ with support within $[-T, T]$. Then, asymptotically the problem of the tagged arrival reduces to finding a distribution $G$ that minimizes

$$\int_{-T}^{T} (\beta(\tilde{t} + \bar{W}(\tilde{t})) + \alpha \bar{W}(\tilde{t})) \, dG(\tilde{t}).$$

Again, the solution corresponds to assigning the probability mass to times $t \in [-T, T]$ that minimize

$$(\alpha + \beta)\bar{W}(t) + \beta t.$$

### 3.1 Game of Arrivals: The Fluid Problem

This asymptotic 'fluid' problem can be modeled as a non-atomic game where each customer is a point in the unit interval $[0, 1]$. All customers have identical cost functions. This is given by

$$C(G, F) = \int_{-T}^{T} (\beta(t + \bar{W}(t)) + \alpha \bar{W}(t)) \, dG(t)$$

for a customer who chooses her arrival using distribution $G$, when everyone else selects their arrival using distribution $F$. Note that this determines the waiting time $\bar{W}(t)$ for an arrival at time $t$ in the limiting fluid model.

We first define a symmetric Nash equilibrium for such a game.

DEFINITION 1. *A mixed strategy profile $F$ is a symmetric Nash equilibrium if for any player $s \in [0, 1]$ who plays $F_s$,*

$$C(F, F) \leq C(F_s, F), \forall F_s.$$

That is, if all the other customers use mixed strategy $F$, then an arbitrary player $s$'s best response is to also play $F$ (though the best response may not be unique).

Suppose that everyone else aside from the tagged customer $s$ follows the distribution $F(\cdot)$ (this may be the limiting distribution in (1)). Then, $\bar{W}(t)$ for $t < 0$ equals

$$-t + F(t)/\mu = F(t)/\mu - t,$$

where $-t$ is the customer wait before the server becomes active, and $F(t)/\mu$ denotes the remaining queueing delay, once the server is active. For $t > 0$, $\bar{W}(t)$ equals the asymptotic workload at time $t$, that is, $\bar{Q}(t)/\mu$. So the solution to the tagged customer's arrival distribution problem corresponds to finding times t that minimize

$$(\alpha + \beta)F(t)/\mu - \alpha t$$

for $t < 0$ and

$$(\alpha + \beta)\bar{Q}(t)/\mu + \beta t$$

for $t > 0$.

Let

$$t^* = \inf\{t \geq 0 : F(t) < \mu t\}.$$

This denotes the first time the server starts to serve at less than the full rate $\mu$ after it starts serving at time zero. Note that $t^* \leq \tau_1 = 1/\mu$ since, otherwise, there exists an $\epsilon > 0$ such that $F(\tau_1 + \epsilon) \geq \mu(\tau_1 + \epsilon) > 1$, a contradiction.

Consider the case where $t^* = \tau_1$. This corresponds to the case where the server continuously serves customers at her full rate $\mu$ till time $\tau_1$ and at that time all the customers are served. In this case for $t \in [0, \tau_1]$,

$$\bar{Q}(t) = F(t) - \mu t,$$

so that

$$\bar{W}(t) = F(t)/\mu - t \qquad (4)$$

for all $t \in [-T, \tau_1]$. Hence, in this case tagged customer's arrival distribution is selected so that all the probability mass is assigned to times in $[-T, \tau_1]$ that minimize

$$(\alpha + \beta)F(t)/\mu - \alpha t$$

for $t \in [-T, \tau_1]$.

### 3.1.1 Game Theoretic Analysis

Let $\mathcal{G}$ denote the space of all probability distributions on $[-T, T]$. Recall that $F$ denotes the arrival distribution followed by everyone else in the non-atomic game described earlier. For this $F$ to be a symmetric equilibrium, we need that

$$F \in \arg\min_{G \in \mathcal{G}} \int_{-T}^{T} \left((\alpha + \beta)\bar{W}(t) + \beta t\right) dG(t).$$

Recall that

$$t^* = \inf\{t \geq 0 : F(t) < \mu t\}.$$

The following lemma helps characterize such an $F$.

LEMMA 1. *If $F$ is a symmetric Nash equilibrium for our non-atomic game, then $t^* = \tau_1$ so that $F(\tau_1) = 1$.*

**Proof:** To see this, suppose that for a given $F$, $t^* < \tau_1$. Then, it is easy to see that $F$ cannot be a symmetric Nash equilibrium. To see this, note that at time $t^*$, there is no

queue and hence $\bar{W}(t^*) = 0$. Furthermore, since $t^* < \tau_1$, all the customers have not arrived and hence $F(t^*) < 1$.

Now select an arrival distribution $G$ for a tagged customer as follows: Set $G(t) = F(t)$ for $t < t^*$ and $G(t^*) = 1$. This distribution puts a positive mass at time $t^*$ when the waiting time is zero. It is easy to see that $C(G, F) < C(F, F)$ so that $F$ cannot be a symmetric Nash equilibrium. $\square$

Therefore, in our search for equilibrium solution we restrict ourselves to $F$ with $t^* = \tau_1$ and $F(\tau_1) = 1$. In particular $\bar{B}(t) = t$ for all $t \in [0, \tau_1]$.

It follows from (4) that for such an $F$, the cost at any time $t \in [-T, \tau_1]$ equals

$$(\alpha + \beta)F(t)/\mu - \alpha t. \qquad (5)$$

Let $\tau_0 = -\frac{1}{\mu}\frac{\beta}{\alpha}$. (We select $T$ such that $T > \frac{1}{\mu}\frac{\beta}{\alpha}$.) The equation (5) becomes independent of $t$ for $t \in [\tau_0, \tau_1]$ if we select $F = F^*$ where $F^*(t) = 0$ for $t \in [-T, \tau_0]$,

$$F^*(t) = \frac{t - \tau_0}{\tau_1 - \tau_0}$$

$t \in [\tau_0, \tau_1]$, and $F^*(t) = 1$ for $t \in [\tau_1, T]$. In that case, (5) equals $\beta/\mu$ for $t \in [\tau_0, \tau_1]$. Hence, if each customer samples their arrival from $F^*$, the cost to each customer equals $\beta/\mu$.

THEOREM 1. *The distribution $F^*$ corresponds to a unique symmetric Nash equilibrium solution to our non atomic game.*

Suppose that $F$ is a symmetric Nash equilibrium. Let $S$ denote its support (smallest closed set so that its complement has probability zero under $F$). Then, along this support the cost is a.s. constant $c$ for some $c$, and it is $\geq c$ a.s. elsewhere. Let $t_0$ be the left boundary of $S$ and $t_1$ be its right boundary. Clearly $t_0 \leq 0$ and from Lemma 1, we have that $t_1 \leq \tau_1$. Note that the cost to arrival at time $t$ equals $(\alpha + \beta)F(t)/\mu - \alpha t$. For this cost to a.s. equal $c$ in $S$, $S$ must be an interval $[t_0, t_1]$ (note that there cannot be an open interval where $F$ is constant and the $(\alpha+\beta)F(t)/\mu - \alpha t$ is also constant).

Also note that $t_1 < \tau_1$ is not feasible as then

$$c = (\alpha + \beta)/\mu - \alpha t_1 > (\alpha + \beta)1/\mu - \alpha\tau_1 = \beta/\mu$$

and this is greater than the cost $\beta/\mu$ incurred by a customer that comes at time $\tau_1$ with probability 1. Thus, $t_1 = \tau_1$ and $c = \beta/\mu$. This, further implies that $t_0 = -\beta/(\mu\alpha) = \tau_0$. Note that an arrival outside interval $[\tau_0, \tau_1]$ incurs a cost $> \beta/\mu$. This is because the cost of customer coming before $\tau_0$ is higher than her coming at $\tau_0$ (due to the cost of increased wait) . Similarly, the cost of coming after $\tau_1$ is more than coming at $\tau_1$ (in both cases the waiting cost is zero). Therefore, $F = F^*$ is the unique symmetric Nash equilibrium to our game. $\square$

## 3.2 Global Welfare Solution and Price of Symmetric Anarchy

We now discuss the global welfare solution for the limiting fluid problem. Here we identify a distribution $F$ followed by all the customers that minimizes the overall cost. The associated fluid problem cost equals

$$\int_{-T}^{T} \left((\alpha + \beta)\bar{W}(t) + \beta t\right) dF(t).$$

At the global welfare solution, each customer selects a distribution in a such way so that the total cost to all the

customers is minimized. We first develop a lower bound on this cost and then propose a distribution that achieves this. To get a lower bound, note that the smallest value of $\bar{W}(t)$ is zero for all $t$. Also note that the total time to service is minimized if the server serves at the fastest possible rate. That is, it functions at full rate $\mu$ starting at time zero and serves all the customers by time $\tau_1$. Then, the average time to service for each customer is $\tau_1/2$. So the lower bound on the overall cost is $\beta\tau_1/2 = \beta/(2\mu)$.

It is easy to see that $F(t) = t/\tau_1$ for $0 \le t \le \tau_1$ achieves this lower bound. Note that under this arrival distribution, the queue always equals zero and there is no waiting so $\bar{W}(t) = 0$. Furthermore, under this distribution, the service is provided to all at the fastest possible rate: All customers are served by time $\tau_1$, so the average time to service of all customers is also minimized. Therefore, the minimum possible cost in the system corresponds to everyone following the distribution $F(t) = t/\tau_1$ for $0 \le t \le \tau_1$ and equals $\beta\tau_1/2 = \beta/(2\mu)$.

Recall that the cost associated with everyone following a symmetric equilibrium distribution $F^*$ equalled $\frac{\beta}{\mu}$ so that the equilibrium solution is twice as expensive as the global welfare solution. Therefore the price of symmetric anarchy defined as the ratio of maximum cost over all symmetric Nash equilibria with the global welfare solution, equals 2.

REMARK 2. In our cost structure, we have thus far considered two components: The cost of time at which service is received plus the cost of waiting in a queue. In many settings, such as at the concert hall, while the cost of waiting is appropriate, the other component of cost is better modeled as proportional to the number of customers that have received service before the tagged customer gets served. Fortunately, that leads to only minor changes in our fluid analysis.

To see this note that in the fluid model this change corresponds to replacing the cost

$$\beta(t + \bar{W}(t)) + \alpha\bar{W}(t)$$

with

$$\beta F(t) + \alpha\bar{W}(t).$$

Again, consider the case where $t^* = \tau_1$. As discussed earlier, in this setting

$$\bar{W}(t) = F(t)/\mu - t.$$

so the overall cost equals

$$\frac{F(t)}{\mu}(\alpha + \hat{\beta}) - \alpha t,$$

where $\hat{\beta} = \beta\mu$. This differs from the previously analyzed cost function in that $\hat{\beta}$ replaces $\beta$. It is then easily seen that if we define $\hat{\tau}_0 = -\frac{1}{\mu}\frac{\hat{\beta}}{\alpha} = -\frac{\beta}{\alpha}$. (We select $T$ such that $T > \frac{\beta}{\alpha}$.) and select $\hat{F}(t) = 0$ for $t \in [-T, \hat{\tau}_0]$,

$$\hat{F}(t) = \frac{t - \hat{\tau}_0}{\tau_1 - \hat{\tau}_0}$$

$t \in [\hat{\tau}_0, \tau_1]$ and $\hat{F}(t) = 1$ for $t \in [\tau_1, T]$. Then, $\hat{F}$ is a Nash equilibrium solution to the fluid problem. This follows essentially from the proof of Nash equilibrium of $F^*$ in the previous case. One simply needs to note that an arrival before time $\hat{\tau}_0$ incurs extra waiting cost compared to

an arrival at time $\hat{\tau}_0$, both see no customers in the queue. Furthermore, an arrival at any time $\ge \tau_1$ sees all the customers arrive before her and incurs no waiting. So the total cost is independent of her arrival time once it is $\ge \tau_1$.

Under this Nash equilibrium, it can be seen that the cost to a customer equals $\beta$ (as compared to $\beta/\mu$ in the previous settings). Furthermore, it is again easy to see that $\beta$ bounds from above the cost of any other symmetric Nash equilibrium. Furthermore, the global welfare solution corresponds to any strategies employed by customers that leads to zero waiting. In that case the cost to each customer equals

$$\beta\int_0^T F(t)dF(t) = \beta\int_0^1 xdx = \beta/2,$$

so that the price of symmetric anarchy again equals 2.

## 4. CONCLUSION

In this paper we considered the queueing problem that may arise in settings such as concert halls, movie theaters, cafeterias etc., where a large number of customers may queue up before a facility that opens for service at a particular time. The customers strategically selected their arrival time distributions. We developed a queueing framework for this problem for which we identified the fluid limit. We observed that fluid limit allows a great deal of tractability in analyzing the strategic arrival problem faced by each customer. We identified the unique symmetric Nash equilibrium strategy for each customer and showed that the price of symmetric anarchy equalled 2 in our framework. In our ongoing work, we study the equilibrium properties of the fluid model under general cost functions. We generalize our analysis to multiple classes of customers and we discuss structural changes in the queueing discipline that may reduce the price of anarchy.

We hope that this analysis motivates further research in strategic analysis of queues using fluid model simplifications.

## 5. REFERENCES

[1] H. Chen and D. Yao, *Fundamentals of Queueing Networks*, Springer, 2001.

[2] A. Glazer and R. Hassin, "?/M/1: On the equilibrum distribution of customer arrivals", *European J. of Oper. Research* 13:146-150, 1983.

[3] R. Hassin and M. Haviv, *To Queue or Not to Queue*, Kluwer Academic Publishers, 2003.

[4] M. Haviv, "The Aumann-Shapley price mechanism for allocating congestion costs", *Operations Research Letters* 29:211-215, 2001.

[5] M.A. Lariviere and J. A. van Mieghem, "Strategically seeking service: How competition can generate Poisson arrivals", *Manufacturing and Service Operations Management*, 6(1):23-40, 2004.

[6] P. Lederer and L. Li, "Pricing, Production, Shceduling, and Delivery-time Shceduling", *Operations Research* 45(3):407-420, 1997.

[7] Y. Masuda and S. Whang, "Dynamic pricing for network service: equilibrium and stability", *Management Science*, 45(6):857-869, 1999.

[8] H. Mendelson and S. Whang, "Optimal incentive-compatible priority pricing for the M/M/1 queue", *Operations Research*, 38(5):870-883, 1990.

[9] P. Naor, "The regulation of queue size by levying tolls", *Econometrica*, 37(1):15-24, 1969.

[10] D. Schmeidler, "Equilibrium points of nonatomic games", J. Stat. Physics 7(4):295-300, 1973.

[11] D. Stahl and A. Whinston, "A general economic equilibrium model of distributed computing", in *New directions in computational economics*, eds. W. Cooper and A. Whinston, Kluwer Acad. Pub., pp.175-189, 1994.

[12] J. Van Mieghem, "Price and service discrimination in queueing systems: incentive compatibility of $Gc\mu$ scheduling", *Management Science*, 46(9):1249-1267, 2000.

[13] Q. Tian, H. Huang, and H. Yang, "Equilibrium properties of the morning peak-period commuting in a many-to-one mass transit system", *Transportation Research Part B*, 41, 6: 616-631, 2007.

[14] W. Whitt, *Stochastic Process Limits*, Springer, 2002.