The Concert Queueing Game with a Random Volume of Arrivals

Sandeep Juneja
Tata Institute of Fundamental Research
Mumbai
juneja@tifr.res.in

Tushar Raheja
Indian Institute of Technology
Delhi
tushar@raheja.org

Nahum Shimkin
Technion - Israel Institute of Technology
Haifa
shimkin@ee.technion.ac.il

***Abstract*—We consider the concert queueing game in the fluid framework, where the service facility opens at a specified time, the customers are particles in a fluid with homogeneous costs that are linear and additive in the waiting time and in the time to service completion, and wish to choose their own arrival times so as to minimize their cost. This problem has recently been analyzed under the assumption that the total volume of arriving customers is deterministic and known beforehand. We consider here the more plausible setting where this volume may be random, and only its probability distribution is known beforehand. In this setting, we identify the unique symmetric Nash equilibrium and show that under it the customer behavior significantly differs from the case where such uncertainties do not exist. While, in the latter case, the equilibrium profile is uniform, in the former case it is uniform up to a point and then it tapers off. We also solve the associated optimization problem to determine the socially optimal solution when the central planner is unaware of the actual amount of arrivals. Interestingly, the Price of Anarchy (ratio of the social cost of the equilibrium solution to that of the optimal one) for this model turns out to be two exactly, as in the deterministic case, despite the different form of the social and equilibrium arrival profiles.**

I. Introduction

Customers going to a rock concert or a movie theater need to resolve the following dilemma: Going early involves encountering a rush to get the best seats, going late involves sacrifice in the viewing experience. Evening commuters often face the trade-off between reaching home late from work or getting caught in the evening rush hour. Similar trade-offs govern queueing behavior in a busy cafeteria: People may prefer to eat as soon as the cafeteria opens at lunch time or they may choose to stay hungry and eat later when the waiting is less but the food quality may deteriorate. We refer to this 'queue arrival timing problem' as the concert queueing problem (see [10] and [9]). This problem is especially important when the number of potential customers involved is large. Typically, even when the size of population coming to a queue is large, it may still be substantially variable.

In this paper we consider this concert queueing problem in the fluid framework. Here each customer is a particle or a point in a continuum that needs to decide when to arrive to a queue where the server opens service at a specified time. The arrivals are non-cooperative, their cost structure is homogeneous and is linear and additive in the waiting time and in the time to service. The customers can arrive before or after the server opens for service, and are served in a first come first serve manner. This problem was recently considered in [10], where the total volume of customers is assumed to be fixed and known beforehand to arriving customers. This fluid model approximates the actual scenario where the total number of customers is finite but large and more or less constant (see [11] for a proof of convergence of the equilibrium profile in the discrete queueing model to that of the associated fluid model as the number of customers increases to infinity). This basic fluid model has been extended to multiple classes of customers [9], parallel and serial queues [8], and different opening and closing conditions [7]. In this paper, we analyze a more realistic scenario in the fluid setting, where the volume of arriving customers may be random, and only its probability distribution is known upfront.

In [10] and [9], the authors show that there exists a unique Nash equilibrium arrival profile that corresponds to customers arriving uniformly over a specified interval. They further show that the price of anarchy (the ratio of the social cost of the worst Nash equilibrium to the optimal one) in their framework equals 2. As mentioned above, we extend this framework to allow for random arrival volume. Under this extension, we derive the unique symmetric equilibrium profile for customer arrival instances. We note that this differs significantly from the arrival profile when volume of arrivals is fixed. Specifically, we show that in the random setting, the unique Nash equilibrium profile is uniform only up to a point and then it tapers off as a function of time. Thus, customers have a higher arrival density in the beginning of the arrival period than at its end. We also explicitly evaluate the cost incurred by each customer in equilibrium, and verify that uncertainly in the arrival volume tends to increase this cost.

We also consider the problem of determining the socially optimal solution in this setting when the central planner is unaware of the volume of the arriving traffic, but can dictate the distribution of arrival times for those who do arrive. This problem may be of independent interest in various settings. For instance, when a central planner gives appointments to arriving customers and a random amount of customers show up. It is also useful in ascertaining the level of inefficiency of the equilibrium profile through the computation of PoA. We note that unlike in the case where the arrival volume is fixed, when it is allowed to be random, the social optimal solution may

VALUETOOLS 2012, October 09-12, Cargèse, France

DOI 10.4108/valuetools.2012.250166

involve queueing under certain scenarios. Interestingly, unlike in the equilibrium solution, under the social optimal solution the arrival profile of each customer is indeed uniform. Also, the PoA turns out to exactly equal two, as in the deterministic arrival volume case.

Regarding related literature, a comprehensive overview of game theoretic (or strategic) decision problems in queueing systems may be found in the monograph [5]. However, it does not address fluid models. Equilibrium flows in transportation and communication networks (also known as the selfish routing problem) have been extensively studied following Wardrop's seminal 1952 paper [18]; see [16] for a survey of that literature. This model essentially considers a fluid flow problem which views users as infinitesimal and selfish, similar to our model, but does not address timing decisions which are the focus of this paper. Bottleneck fluid models similar to ours have been extensively studied the transportation setting, starting with a seminal paper by Vickrey [17]. In the basic model, also known as the morning commute problem, a known volume of infinitesimal users are served on a FCFS basis by a fixed-rate server, and need to choose their staring time so that their service ends as close as possible to a nominal arrival time. Here a cost is typically incurred both for late and for early arrival. For later developments on this model see, e.g., [12], [13] and their references. The specific effects of uncertainty in population size have been considered in [1], which is perhaps the closest work to ours. In their set-up penalties are imposed for early and late service completions (relative to a nominal target time, common to all), while we only penalize for lateness. They also explicitly model demand and supply to determine the distribution of volume of customers that show up. Our analysis on the other hand is substantially more detailed. We also determine explicitly the socially optimal solution not considered in those papers. In a non-fluid setting, we finally mention the work in [4], [6], [11] on similar strategic arrival timing problems in queues with a finite customer population and stochastic service times. As may be expected, the analysis there becomes more complicated and the results less explicit.

The organization of this paper is as follows: In Section 2, we develop the mathematical framework for the concert queuing game involving random volume of arrivals having homogeneous and linear costs. In Section 3, we identify a unique symmetric Nash Equilibrium in this framework. In Section 4 we identify the socially optimal profile and calculate the PoA. We explicitly solve for the equilibrium and social profiles for a few examples in Section 5. We end with brief conclusion in Section 6.

II. Mathematical Framework

Consider the following fluid model: The population of potential arrival is represented by the continuous interval $[0,\infty)$, with each customer considered a point in that interval. The volume of customers that actually arrive to the queue is a random variable $\Lambda \in [0,\infty)$, with distribution function G_0. We assume that Λ has a finite mean. All arriving customers are admitted to the queue, and are served in a first come first serve manner.

Service starts at time zero, and commences thereafter as a constant rate* $\mu > 0$. The costs incurred by each customer are taken to be linear and additive in waiting time and time to service. For ease of analysis we restrict ourselves to customers behaving symmetrically, in that they select their arrival times from the same distribution F (this in fact is not crucial here since in the fluid setting it is only the aggregate arrival profile that matters; for more on this see [9]). The aim then is to look for a common equilibrium distribution. Now, if random Λ amount of customers that arrive, we can map them into the interval $[0,\Lambda]$. If each of them samples its arrival time from F, then, $\Lambda F(t)$ denotes the random amount of arrivals by time t. Let $W_{\Lambda,F}(t)$ be the waiting time of an arrival at time t in this scenario. Conditioned that the amount of arrivals equals Λ, the cost of an arrival at t to the serving facility is given by

$$C_{\Lambda,F}(t) = \alpha W_{\Lambda,F}(t) + \beta(t + W_{\Lambda,F}(t))$$

where $t + W_{\Lambda,F}(t)$ is the time to service of a customer who arrives at time t. $\alpha > 0$ is the unit cost of waiting time in the system and $\beta > 0$ is the unit cost of time to service. It will be convenient to normalize the cost so that $\alpha + \beta = 1$; in particular, $0 < \alpha, \beta < 1$.

Let $Q_{\Lambda,F}(t)$ denote the queue size at time t when Λ amount of customers arrive (recall that Λ is a random variable). To relate this to F, let

$$X_{\Lambda,F}(t) = \Lambda F(t) - \mu t 1_{\{t\geq 0\}}$$

denote the net input process to the system when Λ customers arrive. Then, it is well known that (see Chapter 6.2 in [2]):

$$Q_{\Lambda,F}(t) = X_{\Lambda,F}(t) - \min\{0, \inf_{s\leq t} X_{\Lambda,F}(s)\}. \qquad (1)$$

Note that $Q_{\Lambda,F}(t) \geq 0$, and that it can only have upward jumps, that match those of F. In particular, if F does not have a jump at time t, then

$$W_{\Lambda,F}(t) = Q_{\Lambda,F}(t)/\mu + \max\{0, -t\}.$$

If F has a jump at time t, then the position of an arriving customer would be uniformly distributed in $[Q_{\Lambda,F}(t-), Q_{\Lambda,F}(t)]$, and its expected waiting time conditional on amount of arrivals Λ,

$$W_{\Lambda,F}(t) = \bar{Q}_{\Lambda,F}(t)/\mu + \max\{0, -t\}$$

where

$$\bar{Q}_{\Lambda,F}(t) = \frac{1}{2}(Q_{\Lambda,F}(t-) + Q_{\Lambda,F}(t)).$$

Also note that the distribution of the volume of arrivals as seen by an arriving customer differs from G_0, and is given by the tilted distribution G defined by

$$dG(\lambda) = \frac{\lambda dG_0(\lambda)}{\int_0^\infty \lambda dG_0(\lambda)}$$

*We consider here the service rate to be deterministic. However, most of the following results are applicable to the case of stochastic μ, as the ratio Λ/μ is the main quantity that appear in the analysis.

(see [3] or [1], for example). This length biased distribution captures the fact that a particular arrival is more likely when the total number of arrivals is large.

The unconditional expected cost seen by a customer if she arrives at time t then equals (recalling that $\alpha + \beta = 1$)

$$EC_F(t) = \int_0^\infty W_{\lambda,F}(t)\, dG(\lambda) + \beta t,$$

and the expected cost of a customer who selects her arrival time by sampling from probability distribution H is

$$EC_{H,F} = \int_{-\infty}^{\infty} \left[\int_0^\infty W_{\lambda,F}(t)\, dG(\lambda) + \beta t \right] dH(t).$$

Note that the expectation here is taken with respect to the length-biased distribution G.

Our arrival game therefore corresponds to a volume Λ of arrivals showing up at the server facility and each selecting her arrival time as an independent sample from F, a probability distribution over the reals. We refer to F as the *arrival profile*. The following definition of symmetric Nash equilibrium is standard:

Definition 1: An arrival profile F is a Symmetric Nash Equilibrium (SNE) if, for every distribution H,

$$EC_{F,F} \leq EC_{H,F}\,.$$

Equivalently, there exists a set $\mathcal{T}_F$ of F-measure 1 and a constant c_e such that

$$(i) \quad EC_F(t) \geq c_e \quad \text{for all } t, \tag{2}$$

$$(ii) \quad EC_F(t) = c_e \quad \text{for all } t \in \mathcal{T}_F. \tag{3}$$

Here c_e denotes the expected cost incurred by a customer that arrives with probability 1 along the set $\mathcal{T}_F$. Customer, were it to arrive at any other time, will incur expected cost that is at least c_e.

To see the equivalence, first suppose that for a given F and $\mathcal{T}_F$, (i) and (ii) above hold. Then, $EC_{H,F} \geq c_e$ for every distribution H, while $EC_{F,F} = c_e$. On the other hand, if given a candidate F for SNE, violation of (i) clearly implies that F is not an SNE. Violation of (ii) again implies that there exists a set of positive F-measure where the cost is less than it is at another set of positive F-measure. Again, it is easy to that such an F is not an SNE.

Recall that the support of a probability measure is the smallest closed set that has probability 1. Let $\overline{\mathcal{T}}_F$ denote the support of the probability measure associated with an arrival profile F.

The following regularity assumption will be invoked in parts of our analysis. A similar assumption was used in [11]. While imposing reasonable restrictions on the arrival-time distributions that may be employed by the customers, it makes our search for the equilibrium distribution substantially simpler.

Assumption 1: The support $\overline{\mathcal{T}}_F$ of SNE profile F can locally (i.e., on any finite interval) be represented as a finite union of closed intervals and points.

III. Equilibrium Analysis

In this section through a series of lemmas we develop necessary conditions that an SNE must satisfy. We then show the existence of a unique SNE. Let $t_b = \{\inf x : x \in \overline{\mathcal{T}}_{\mathcal{F}}\}$ and $t_e = \{\sup x : x \in \overline{\mathcal{T}}_F\}$ be the end points of the support of F, corresponding to the first and last arrival times. The following properties of an SNE are easily seen.

Lemma 1: An SNE profile F is a continuous function of t (i.e., the corresponding probability measure has no point masses). In addition, the expected cost $EC_F(t)$ is constant over the support $\overline{\mathcal{T}}_F$. Furthermore, $-\infty < t_b < 0$ and $0 < t_e < \infty$ (hence, $0 < F(0) < 1$).

Proof: The first claim is easily seen as if the profile had a point mass at time t, then there must exist an $\epsilon > 0$ such that $EC_F(t-\epsilon) < EC_F(t)$.

To see the second claim note that since F is continuous, the waiting time $W_{\lambda,F}(t)$ is continuous for each λ so that the cost $EC_F(t)$ is a continuous function of t. Hence, (3) extends to the support $\overline{\mathcal{T}}_F$.

Next, $t_b > -\infty$ follows as $EC_F(t)$ increases to ∞ as $t \downarrow -\infty$. To see that $t_b < 0$ note that if it were larger than or equal to zero, then a customer arriving at time zero would incur zero wait, hence would incur a cost that is strictly smaller than any customer arriving at a positive time, leading to a contradiction. The assertion that $0 < t_e < \infty$ can be verified similarly. ■

Let

$$T_\Lambda = \inf\{t \geq 0 : \Lambda F(t) < \mu t\} \tag{4}$$

denote the first time after zero when the server starts to serve at less than full rate μ, given that the arrival volume is Λ. Note that $Q_{\Lambda,F}(T_\Lambda) = 0$. Since $F(0) > 0$, it follows that $T_\Lambda > 0$ for $\Lambda > 0$. Lemma 2 below is important for our analysis: It states that no queue will build up beyond T_Λ.

Lemma 2: For an SNE profile F, $Q_{\Lambda,F}(\tau) = 0$ for all $\tau \geq T_\Lambda$.

Proof: Suppose that there exists $\tau > T_\Lambda$, so that $Q_{\Lambda,F}(\tau) > 0$. Without loss of generality we may assume that $\tau \in \overline{\mathcal{T}}_F$. Then due to continuity of F, (and hence through continuity of $Q_{\Lambda,F}$), there exists $s \geq T_\Lambda$ denoting the last time before τ that $Q_{\Lambda,F}(t)$ equals zero. Clearly, if it were known that Λ is the arrival volume, then arriving at s would be preferable to arriving at τ, contradicting $\tau \in \overline{\mathcal{T}}_F$. We need to show that this is the case also when the arrival volume is stochastic.

From (1), it follows that

$$X_{\Lambda,F}(\tau) - X_{\Lambda,F}(s) = Q_{\Lambda,F}(\tau) - Q_{\Lambda,F}(s) > 0,$$

and hence,

$$\Lambda(F(\tau) - F(s)) > \mu(\tau - s). \tag{5}$$

Now, the desired contradiction follows by comparing the costs at times τ and s and showing that $EC_F(\tau) > EC_F(s)$. To

see this, recall that,

$$EC_F(t) = \int_0^\infty W_{\lambda,F}(t)\, dG(\lambda) + \beta t.$$

We claim that $W_{\lambda,F}(\tau) \geq W_{\lambda,F}(s)$ for all λ. Indeed, for any λ such that $W_{\lambda,F}(s) = 0$, this holds trivially. Otherwise, for λ such that $W_{\lambda,F}(s) > 0$ (that is, if $\lambda > \Lambda$), we have by (5) that $W_{\lambda,F}(\tau) > W_{\lambda,F}(s)$. Now, since $\tau > s$, we obtain that $EC_F(\tau) > EC_F(s)$, providing the desired contradiction. ■

Corollary 1: For an SNE profile F,

$$W_{\Lambda,F}(t) = \Lambda F(t)/\mu - t \tag{6}$$

for $t \leq T_\Lambda$, and $W_{\Lambda,F}(t) = 0$ otherwise.

Proof: Observe that for $t < 0$, $Q_{\Lambda,F}(t) = \Lambda F(t)$ and hence (6) follows, as $-t$ is the customer wait before the server becomes active, and $\Lambda F(t)/\mu$ is the remaining queueing delay. For $t \geq 0$ the required equality follows from Lemma 2, which implies that the server will be working at full rate on $0 \leq t \leq T_\Lambda$. ■

Lemma 3: $\frac{\mu t}{F(t)}$ strictly increases as a function of t for all $t \in \overline{\mathcal{T}}_F$.

Proof: From Corollary 1, it follows that for $\lambda < \frac{\mu t}{F(t)}$, the associated waiting time $W_{\lambda,F}(t) = 0$. Hence,

$$EC_F(t) = \int_{\lambda \geq \frac{\mu t}{F(t)}} \left(\lambda \frac{F(t)}{\mu} - t\right) dG(\lambda) + \beta t.$$

Through integration by parts, letting $\bar{G}(\lambda) = 1 - G(\lambda)$ for each λ, this may be re-expressed as

$$EC_F(t) = t\left(\frac{F(t)}{\mu t}\int_{\frac{\mu t}{F(t)}}^{\infty_h} \bar{G}(\lambda)d\lambda + \beta\right).$$

Now, $x\int_{1/x}^\infty \bar{G}(\lambda)d\lambda$ is clearly a non-decreasing function of x. Since $EC_F(t)$ is constant for all $t \in \overline{\mathcal{T}}_F$, the result follows. ■

Lemma 4: An SNE profile F has a right continuous derivative in $\overline{\mathcal{T}}_F$ given by

$$F'(t) = \mu \frac{\bar{G}(\frac{\mu t}{F(t)}) - \beta}{\int_{\frac{\mu t}{F(t)}}^\infty \lambda dG(\lambda)} \tag{7}$$

for each $t \in \overline{\mathcal{T}}_F$.

Proof: Recall that

$$EC_F(t) = \frac{F(t)}{\mu}\int_{\frac{\mu t}{F(t)}}^\infty \bar{G}(\lambda)d\lambda + \beta t.$$

Note that on $\overline{\mathcal{T}}_F$, the derivative of $EC_F(t)$ in t equals zero. Hence, through simple manipulations it follows that for $t \in \overline{\mathcal{T}}_F$, wherever F is differentiable (that is, almost everywhere)

$$F'(t) = \mu \frac{\bar{G}(\frac{\mu t}{F(t)}) - \beta}{\int_{\frac{\mu t}{F(t)}}^\infty \lambda dG(\lambda)} = \mu \frac{\bar{G}(\frac{\mu t}{F(t)}) - \beta}{\int_{\frac{\mu t}{F(t)}}^\infty \bar{G}(\lambda)d\lambda + \frac{\mu t}{F(t)}\bar{G}(\frac{\mu t}{F(t)})}.$$

Since, the RHS is right continuous, there exists a right continuous version of F' in $\overline{\mathcal{T}}_F$. ■

Remark 1: Let $\mathcal{G}$ denote the set of points of discontinuity of the length-biased volume distribution G (which is countable at most). Then the points of discontinuity of F' correspond to times t at which $\frac{\mu t}{F(t)} \in \mathcal{G}$.

Lemma 5: Under Assumption 1, the support $\overline{\mathcal{T}}_F$ of an SNE profile F is an interval, denoted $[t_b, t_e]$.

Proof: Clearly, $\overline{\mathcal{T}}_F$ does not consist of isolated points as it cannot have point mass at any point. Suppose there exist $t_1 < t_2 < t_3 \in \overline{\mathcal{T}}_F$ such that $0 < F(t_1) = F(t_2) < 1$ and $F'(t_3) > 0$. We show that under Assumption 1, this leads to a contradiction. Specifically, we argue that for such a t_1 we must have

$$\bar{G}\left(\frac{\mu t_1}{F(t_1)}\right) \leq \beta, \tag{8}$$

and

$$\bar{G}\left(\frac{\mu t_2}{F(t_2)}\right) \geq \beta. \tag{9}$$

Then, since $\frac{\mu t}{F(t)}$ is strictly increasing with t, and $\bar{G}$ is a non-increasing function, this implies that $\bar{G}\left(\frac{\mu t_1}{F(t_1)}\right) = \bar{G}\left(\frac{\mu t_2}{F(t_2)}\right) = \beta$. However, since $F'(t_3) > 0$ implies from (7) that $\bar{G}\left(\frac{\mu t_3}{F(t_3)}\right) > \beta$, since $\frac{\mu t}{F(t)}$ strictly increases with t, we have the desired contradiction.

To see (8), note that since (t_1, t_2) is not in $\overline{\mathcal{T}}_F$, it follows from its definition that $EC_F(t)$ is differentiable along this interval (with $F(t)$ set as a constant independent of t) so that $EC_F{}'(t_1^+) \geq 0$. It then follows that $\bar{G}\left(\frac{\mu t_1^+}{F(t_1)}\right) \leq \beta$, and therefore (8) follows since $\bar{G}$ is right continuous.

To see (9), note that under Assumption 1, $F'(t_2^+) \geq 0$, so that

$$\bar{G}\left(\frac{\mu t_2^+}{F(t_2^+)}\right) \geq \beta.$$

Again, since F is continuous and $\bar{G}$ is right continuous, (9) follows. ■

Let $\lambda_l \geq 0$ denote the left limit of support of G, corresponding to the minimal possible arrival volume. Recalling the definition of T_Λ from (4), T_{λ_l} is then the first time beyond 0 that the server starts serving at less than full capacity, for some arrival volume. Also define

$$\lambda^* = \inf\{\lambda : G(\lambda) \geq \alpha\} \tag{10}$$

(recall that $0 < \alpha < 1$ is the normalized waiting cost coefficient). Evidently $G(\lambda^*) = \alpha$, unless λ^* is a discontinuity point of G.

Theorem 1: Under Assumption 1, there exists a unique SNE profile F satisfying the following properties:

(i) $t_e = \frac{\lambda^*}{\mu}$.

(ii) The equilibrium cost c_e is given by

$$c_e = \frac{1}{\mu}\int_{\lambda^*}^\infty \bar{G}(\lambda)d\lambda + \beta\frac{\lambda^*}{\mu}.$$

(iii) $t_b = -\frac{c_e}{\alpha}$.

(iv) For $t_b \le t \le T_{\lambda_l}$, F' is constant and specified by

$$F'(t) = \frac{\mu}{E\Lambda}\alpha.$$

(v) For $T_{\lambda_l} \le t \le t_e$,

$$F'(t) = \mu \frac{\bar{G}(\frac{\mu t}{F(t)}) - \beta}{\int_{\frac{\mu t}{F(t)}}^{\infty} \lambda dG(\lambda)}.$$

(vi) For $T_{\lambda_l} < t \le t_e$, $F'(t)$ is a non-increasing function. Hence $F(t)$ is a concave function on the interval $[t_b, \infty)$.

(vii) Finally,

$$F(0) = \frac{\mu}{E\Lambda}c_e = \frac{1}{E\Lambda}\left(\int_{\lambda^*}^{\infty} \bar{G}(\lambda)d\lambda + \beta\lambda^*\right),$$
$$T_{\lambda_l} = \frac{F(0)}{\mu}\left[\frac{1}{\lambda_l} - \frac{\alpha}{E\Lambda}\right]^{-1},$$
$$F(T_{\lambda_l}) = \frac{F(0)}{\lambda_l}\left[\frac{1}{\lambda_l} - \frac{\alpha}{E\Lambda}\right]^{-1}.$$

Proof: First we argue that an SNE profile has to satisfy properties $(i) - (vii)$. Then we show that a unique profile satisfying these conditions exists.

To see (i), note that since $F'(t_e^-) \ge 0$, we have

$$\bar{G}\left(\frac{\mu t_e^-}{F(t_e^-)}\right) \ge \beta. \tag{11}$$

Furthermore, since t_e is the largest point in the support of F, we have $\frac{d}{dt}EC_F(t_e^+) \ge 0$ (note that the cost for $t > t_e$ has $F(t) = 1$ and is differentiable) so that

$$\bar{G}\left(\frac{\mu t_e^+}{F(t_e^+)}\right) \le \beta. \tag{12}$$

From (11) and (12), (i) follows when $\bar{G}(\lambda) = \beta$ has at most one solution. When it has multiple solutions (which must lie on an interval), it follows that $\bar{G}(\mu t_e) = \beta$. Then, $\mu t_e = \lambda^*$ because by definition of t_e, there exists a sequence $t_n \uparrow t_e$ with $F'(t_n) > 0$ for all n sufficiently large. This implies that $\bar{G}\left(\frac{\mu t_n}{F(t_n)}\right) > \beta$ for all n sufficiently large, so that $\mu t_e = \lambda^*$.

(ii) follows by noting that

$$c_e = EC_F(t) = \frac{F(t)}{\mu}\int_{\frac{\mu t}{F(t)}}^{\infty} \bar{G}(\lambda)d\lambda + \beta t. \tag{13}$$

for all $t \in \overline{\mathcal{T}_F}$ and evaluating this cost at t_e.

(iii) is obvious. (iv) follows from (7), after noting that $\lambda_l F(t) \ge \mu t$ for $t \le T_{\lambda_l}$ so that for such a t, $\bar{G}\left(\frac{\mu t}{F(t)}\right) = 1$. (v) simply restates (7).

(vi) can be seen by differentiating $F'(t)$ in (iv) and (v). We get For $t_b \le t \le T_{\lambda_l}$ clearly $F''(t) = 0$. For $T_{\lambda_l} \le t \le t_e$, after simple manipulations, it follows that

$$F''(t) = -\mu^2 \frac{G'(\frac{\mu t}{F(t)})(1 - \frac{tF'(t)}{F(t)})^2}{F(t)\int_{\frac{\mu t}{F(t)}}^{\infty} \lambda dG(\lambda)} \le 0,$$

at points where $F'(t)$ is differentiable. It is not differentiable for t for which $\frac{\mu t}{F(t)} \in \mathcal{G}$. At these points $F'(t)$ is non-increasing.

In (vii), to evaluate $F(0)$, simply equate the equilibrium cost at time zero to c_e. T_{λ_l} and $F(T_{\lambda_l})$ are determined by noting that

$$F(T_{\lambda_l}) = F(0) + T_{\lambda_l}\frac{\mu}{E\Lambda}\alpha = \mu T_{\lambda_l}\lambda_l.$$

The conditions $(i) - (vii)$ specify the necessary conditions that must apply to any SNE. We now employ a monotonicity argument to show that there exists a unique arrival profile $F(t)$ that satisfies these conditions, and is in fact the unique SNE.

Consider the function

$$h(x,t) = \int_{\frac{\mu t}{x}}^{\infty}\left(\lambda\frac{x}{\mu} - t\right) dG(\lambda) + \beta t - c_e. \tag{14}$$

For $0 < t \le t_e$, the function $h(x,t)$ increases from less than zero to infinity as x increases from zero to infinity. In particular, for any $0 < t \le t_e$, there exists a unique $F(t)$ so that $h(F(t),t) = 0$. It is easy to see that for $0 < t < t_e$,

$$\frac{\partial}{\partial x}h(x,t) = \frac{1}{\mu}\int_{\frac{\mu t}{x}}^{\infty} \lambda dG(\lambda).$$

Since, for $0 < t < t_e$, $\frac{\partial}{\partial x}h(F(t),t) > 0$, by implicit function theorem (see, e.g., Luenberger 1984) this $F(t)$ satisfies the ode (7) for $0 < t < t_e$. The remaining conditions on $F(t)$ for $t \ge 0$ follow from simple algebraic manipulations in (14). ■

Remark 2: We may now examine the effect of randomized arrival volume on the equilibrium cost. Consider the deterministic model with a deterministic arrival volume Λ_0 that equals $E_0(\Lambda)$ (note that in computing the last expectation, we use the true distribution G_0 rather than the biased distribution G). It is easily seen from (13) that the equilibrium cost equals $\frac{\beta}{\mu}\Lambda_0$ (see also [10]). On the other hand, in the stochastic model,

$$\begin{aligned} c_e =& \frac{1}{\mu}\int_{\lambda^*}^{\infty} \bar{G}(\lambda)d\lambda + \beta\frac{\lambda^*}{\mu} \\ \ge& \frac{\beta}{\mu}\left(\int_{\lambda^*}^{\infty} \bar{G}(\lambda)d\lambda + \int_0^{\lambda^*} \bar{G}(\lambda)d\lambda\right) \\ =& \frac{\beta}{\mu}E\Lambda = \frac{\beta}{\mu}\frac{E_0\Lambda^2}{E_0\Lambda} \ge \frac{\beta}{\mu}E_0\Lambda, \end{aligned} \tag{15}$$

(note that E denotes the expectation with respect to the biased distribution G). Thus, the equilibrium cost with random Λ is larger that in the corresponding deterministic model.

IV. Social Optimality

The socially optimal solution to our problem may be considered under two scenarios: 1) The central planner knows the realized Λ and uses this information in selecting the arrival profile F for the arriving customers; 2) The central planner is only aware of the distribution G_0 of Λ, and plans the customer arrival profile F before observing Λ. In the first case, when there the arrival volume is Λ, the arrival profile corresponds to a uniform distribution along the interval $[0, \Lambda/\mu]$ and the

associated total cost equals $\beta\Lambda/\mu$ (see [9]). Its expected value equals $\beta E_0\Lambda/\mu$. The PoA in this case clearly exceeds 2 (see Remark 2 above). The second case arguably provides a more fair comparison in terms of the information available to the respective decision makers. It is also analytically more interesting, and requires the solution of a non-trivial variational problem. Our key observations are that the arrival profile remains uniform in this case, and the PoA exactly equals 2.

Consider then the second problem, where a central planner is given the distribution G_0 of the arrival volume, and wishes to specify the arrival profile $F(t)$ so as to minimize the expected social cost. It is easy to see that an optimal arrival profile would put zero mass before the opening time. Thus, our objective is to minimize

$$J_F = \int_\lambda dG_0(\lambda) \int_0^\infty C_{\lambda,F}(t)\lambda dF(t) \tag{16}$$

where

$$C_{\lambda,F}(t) = \alpha W_{\lambda,F}(t) + \beta(W_{\lambda,F}(t)+t) \tag{17}$$
$$= W_{\lambda,F}(t) + \beta t \tag{18}$$

and

$$W_{\lambda,F}(t) = Q_{\lambda,F}(t)/\mu \tag{19}$$

Theorem 2: The socially optimal arrival profile F that minimizes J_F is given by the uniform distribution

$$F'(t) = \frac{\mu}{\lambda^*}, \quad 0 \le t \le t_e = \frac{\lambda^*}{\mu}$$

where λ^* is defined in (10).

The proof is given below. We note that the socially optimal arrival profile shares the same endpoint t_e with the Nash equilibrium solution. However, the starting point and shape is different.

Under the derived solution, the actual queue size is given by $Q_{\lambda,F}(t) = (\frac{\lambda}{\lambda^*} - 1)^+\mu t$ for $0 \le t \le t_e = \lambda^*/\mu$. At time t_e the queue length equals $(\lambda - \lambda^*)^+$ and thereafter for $t \ge t_e$ it equals

$$(\lambda - \mu t)^+.$$

We illustrate this graphically in Figure IV.

We also point out that the last theorem and its proof are somewhat deeper than what may first meet the eye. Using essentially the same proof, it may be shown that the optimal arrival density $F'(t)$ is proportional to the instantaneous service rate $\mu(t)$ even if that rate is not constant in time. However, we will not deal here with this more general case.

A. Price of Anarchy

Substituting the expression for socially optimal $F(t)$ in (16), the socially optimal cost J^* can be seen to be

$$J^* = \frac{1}{2\mu}\int_{\lambda^*}^\infty \bar{G}(\lambda)d\lambda + \frac{\beta}{2\mu}\lambda^*.$$

From Theorem 1(ii), the Nash Equilibrium cost is given by c_e,

$$\frac{1}{\mu}\int_{\lambda^*}^\infty \bar{G}(\lambda)d\lambda + \beta\frac{\lambda^*}{\mu}.$$

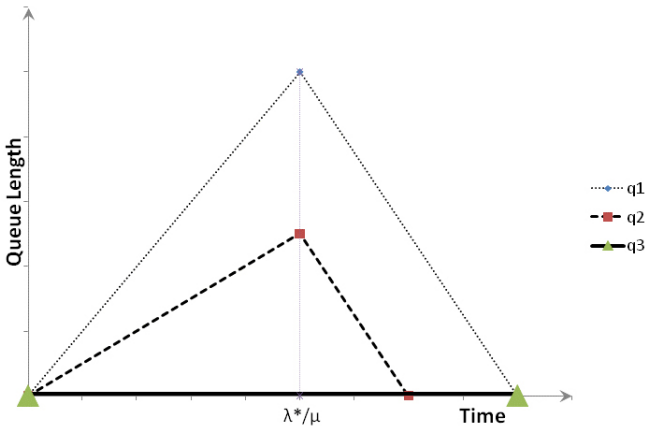

Fig. 1. Queue length process under socially optimal profile. $q1$ and $q2$ correspond to scenarios where arriving $\lambda > \lambda^*$. q3 corresponds to $\lambda \le \lambda^*$.

Therefore, $PoA = c_e/c_S = 2$ in this case.

B. Proof of Theorem 2

The proof proceeds through several steps.

1. *An alternative form for the cost function:* The treatment of the optimization problem is greatly simplified by expressing the cost differently. Start from

$$\int_0^\infty C_{\lambda,F}(t)\lambda dF(t) = \int_0^\infty (W_{\lambda,F}(t) + \beta t)\lambda dF(t). \tag{20}$$

For the integral over $W_{\lambda,F}$ we have:

$$\int_0^\infty W_{\lambda,F}(t)\lambda dF(t) = \int_0^\infty Q_{\lambda,F}(t)dt. \tag{21}$$

This is just the well-known relation between (linear) waiting cost and holding cost. It follow from

$$\int_0^\infty W_{\lambda,F}(t)\lambda dF(t)$$
$$= \int_{t=0}^\infty \int_{s=0}^\infty 1_{\{t\le s<t+W_{\lambda,F}(t)\}} ds\, \lambda dF(t), \tag{22}$$
$$= \int_{s=0}^\infty (\int_{t=0}^\infty 1_{\{t\le s<t+W_{\lambda,F}(t)\}}\lambda dF(t))ds, \tag{23}$$
$$= \int_0^\infty Q_{\lambda,F}(s)ds. \tag{24}$$

For the integral over t we have the standard formula for the expected value of a positive random variable:

$$\int_0^\infty tdF(t) = \int_0^\infty (1 - F(t))dt. \tag{25}$$

Therefore,

$$\int_0^\infty C_{\lambda,F}(t)\lambda dF(t) = \int_0^\infty (Q_{\lambda,F}(t) + \beta\lambda(1-F(t))dt, \tag{26}$$

and

$$J_F = \int_\lambda dG_0(\lambda)\int_0^\infty (Q_{\lambda,F}(t) + \beta\lambda(1-F(t))dt. \tag{27}$$

2. *A relaxed variational problem:* In order to minimize the cost (27), we first formulate and solve a relaxed optimization

problem, and show that the solution to that problem also solves the original one. Observe that

$$Q_{\lambda,F}(t) \geq \tilde{Q}_{\lambda,F}(t)^+, \tag{28}$$

where

$$\tilde{Q}_{\lambda,F}(t) \doteq \lambda F(t) - \mu t\,. \tag{29}$$

We note that $\tilde{Q}$ can be considered as a (possibly negative) queue size in a system that continues service at full rate μ even when the queue is negative.

Consider then the modified cost function $\tilde{J}_F \leq J_F$:

$$\tilde{J}_F = \int_\lambda dG_0(\lambda) \int_0^\infty (\tilde{Q}_{\lambda,F}(t)^+ + \beta\lambda(1-F(t))dt \tag{30}$$

or

$$\tilde{J}_F = \int_\lambda dG_0(\lambda) \int_0^\infty k_\lambda(F(t),t)dt, \tag{31}$$

where

$$k_\lambda(x,t) = (\lambda x - \mu t)^+ + \beta\lambda(1-x). \tag{32}$$

This can be written as

$$\tilde{J}_F = \int_0^\infty K(F(t),t)dt, \tag{33}$$

where

$$K(F(t),t) = \int_\lambda k_\lambda(F(t),t)\, dG_0(\lambda). \tag{34}$$

This can be seen to be in the standard form of a variational problem, with cost function K, optimizing over the (convex) set of probability distribution functions F (i.e, subject to $dF \geq 0$, $F(0)=0$, $F(\infty)=1$).

It is further easily seen that $k_\lambda(x,t)$ is a convex function in x (for any fixed t). It follows then that $\tilde{J}_F$ is a *convex* function of F. This implies that any solution that satisfies the first-order necessary conditions is a global optimum (e.g., see [15]). It is therefore sufficient to show that the proposed solution satisfies the first-order conditions, as we do below.

3. *The first variation:* Let $\epsilon H(t)$ be a continuous variation around $F(t)$, with $H(0) = H(\infty) = 0$. We will also require that $F'(t) + \epsilon H'(t) \geq 0$ for $\epsilon > 0$ small enough. From (31),

$$\tilde{J}_{F+\epsilon H} = \int_\lambda dG_0(\lambda) \int_0^\infty k_\lambda(F(t)+\epsilon H(t),t)dt \tag{35}$$

Now,

$$\frac{dk_\lambda(x,t)}{dx} = \lambda(1_{\{\lambda x - \mu t \geq 0\}} - \beta) \tag{36}$$

almost everywhere, hence

$$D_F \doteq \left.\frac{dJ_{F+\epsilon H}(t)}{d\epsilon}\right|_{\epsilon\downarrow 0} \tag{37}$$

$$= \int_\lambda dG_0(\lambda) \int_0^\infty \lambda(1_{\{\lambda F(t) - \mu t \geq 0\}} - \beta)H(t)dt \tag{38}$$

$$= E_0\Lambda \int_\lambda dG(\lambda) \int_0^\infty (1_{\{\tilde{Q}_{\lambda,F}(t)\geq 0\}} - \beta)H(t)dt \tag{39}$$

$$= E_0\Lambda \int_0^\infty (q_+(t) - \beta)\, H(t)dt, \tag{40}$$

where in (39) we used the relation $dG(\lambda) = \frac{\lambda dG_0(\lambda)}{E_0\Lambda}$, and

$$q_+(t) \doteq \int_\lambda 1_{\{\tilde{Q}_{\lambda,F}(t)\geq 0\}} dG(\lambda). \tag{41}$$

4. *First-order conditions:* Consider the proposed solution, namely $F(t) = \mu t/\lambda^*$ on $[0, t_e]$, so that $q_+(t) = \beta$ on that interval. Since $q_+(t)$ cannot increase in absence of arrivals, we can infer that $q_+(t) \leq \beta$ for $t > t_e$.

Since $H(t) \leq 0$ when $F(t) = 1$, i.e., for $t \geq t_e$, it follows from (40) that $D_F \geq 0$. Thus F satisfies the first-order conditions, and by the stated convexity F is a global solution of the relaxed problem.

5. *Back to the original problem:* We finally observe the solution $F^*(t) = \mu t/\lambda^*$, $0 \leq t \leq t^*$ of the relaxed problem is also an optimal solution to the original problem. Indeed, under this arrival profile the original queue size $Q_{\lambda,F^*}(t)$ has at most one busy period that starts at $t = 0$ (the busy period exists if $\lambda > \lambda^*$, and otherwise $Q_{\lambda,F^*}(t) \equiv 0$), which implies that $Q_{\lambda,F^*}(t) = \tilde{Q}_{\lambda,F^*}(t)^+$, and $J_{F^*} = \tilde{J}_{F^*}$. However, since $J_F \geq \tilde{J}_F$ holds in general, it follows that F^* minimizes J_F as well. □

V. Examples

In this section we illustrate the derived equilibrium and socially optimal profiles on two examples: when G is a two point distribution as well as when G is uniformly distributed. Note that the latter corresponds to $G_0(\lambda)$ proportional to λ^{-1} over an interval.

A. Distribution G is Supported on Two Points

Consider the setting where the number of arrivals can take two possible values under the length biased distribution G: λ_l with probability p_{λ_l} or $\lambda_h > \lambda_l$ with probability $p_{\lambda_h} = 1 - p_{\lambda_l}$. The profiles depend on whether $p_{\lambda_h}\alpha > p_{\lambda_l}\beta$ or not.

Case 1: $p_{\lambda_h}\alpha > p_{\lambda_l}\beta$ (equivalently, $p_{\lambda_h} > \beta$). Here, the equilibrium profile is no longer uniform but is piecewise uniform with the density of the arrival profile taking two possible positive values, higher one first and then lower one in a contiguous interval.

Specifically, from Theorem 1, $\lambda^* = \lambda_h$, so $t_e = \frac{\lambda_h}{\mu}$. The equilibrium cost c_e equals

$$\frac{1}{\mu}\int_{\lambda_h}^{\lambda_h} \bar{G}(\lambda)d\lambda + \beta\frac{\lambda_h}{\mu} = \beta\frac{\lambda_h}{\mu}.$$

Then, $t_b = -\frac{\beta\lambda_h}{\alpha\mu}$. Furthermore, for $t_b \leq t \leq T_l$, $F'(t) = \frac{\mu}{\lambda_h p_{\lambda_h} + \lambda_l p_{\lambda_l}}\alpha$, and for $T_l \leq t \leq t_e$,

$$F'(t) = \mu\frac{(p_{\lambda_h} - \beta)}{\lambda_h p_{\lambda_h}} = \mu\frac{(p_{\lambda_h}\alpha - p_{\lambda_l}\beta)}{\lambda_h p_{\lambda_h}}.$$

This may be re-expressed as

$$\frac{\mu}{E\Lambda}\alpha * \frac{E\Lambda}{\lambda_h}\left(1 - \frac{\beta p_{\lambda_l}}{\alpha p_{\lambda_h}}\right) < \frac{\mu}{E\Lambda}\alpha.$$

Note that F' is piecewise constant and concave on $[t_b, \infty)$. Also,

$$F(0) = \frac{\beta\lambda_h}{\lambda_h p_{\lambda_h} + \lambda_l p_{\lambda_l}},$$
$$T_{\lambda_l} = \frac{\lambda_l\lambda_h}{\mu[p_{\lambda_h}\lambda_h + \lambda_l(\beta p_{\lambda_l} - \alpha p_{\lambda_h})]},$$
$$F(T_{\lambda_l}) = \frac{\lambda_h}{p_{\lambda_h}\lambda_h + \lambda_l(\beta p_{\lambda_l} - \alpha p_{\lambda_h})}.$$

Under the socially optimal profile in this parametric setting we have $\lambda^* = \lambda_h$, so $t_e = \frac{\lambda_h}{\mu}$. The social cost c_s equals $\beta\frac{\lambda_h}{2\mu}$. For $0 \le t \le t_e$, $F'(t) = \frac{\mu}{\lambda_h}$.

Case 2: $p_{\lambda_h}\alpha \le p_{\lambda_l}\beta$ (equivalently, $p_{\lambda_h} \le \beta$). Here the equilibrium profile turns out to be uniform. Again, from Theorem 1, $\lambda^* = \lambda_l$, so $t_e = \frac{\lambda_l}{\mu}$. The equilibrium cost c_e equals

$$\frac{1}{\mu}\int_{\lambda_l}^{\lambda_h} p_{\lambda_h} d\lambda + \beta\frac{\lambda_l}{\mu} = \frac{p_{\lambda_h}\lambda_h + \lambda_l(\beta p_{\lambda_l} - \alpha p_{\lambda_h})}{\mu}.$$

$$t_b = \frac{p_{\lambda_h}\lambda_h + \lambda_l(\beta p_{\lambda_l} - \alpha p_{\lambda_h})}{\alpha\mu}.$$

For $t_b \le t \le t_e$, $F'(t) = \frac{\mu}{\lambda_h p_{\lambda_h} + \lambda_l p_{\lambda_l}}\alpha$. In this case, F' is constant on $[t_b, t_e)$,

$$F(0) = \frac{p_{\lambda_h}\lambda_h + \lambda_l(\beta p_{\lambda_l} - \alpha p_{\lambda_h})}{\lambda_h p_{\lambda_h} + \lambda_l p_{\lambda_l}}, \quad \text{and} \quad T_{\lambda_l} = t_e = \frac{\lambda_l}{\mu}.$$

Under the socially optimal profile in this parametric setting we have $\lambda^* = \lambda_l$, so $t_e = \frac{\lambda_l}{\mu}$. The social cost c_s equals

$$\frac{p_{\lambda_h}\lambda_h + \lambda_l(\beta p_{\lambda_l} - \alpha p_{\lambda_h})}{2\mu}.$$

For $0 \le t \le t_e$, $F'(t) = \frac{\mu}{\lambda_l}$.

B. Distribution G is Uniform

Suppose that G corresponds to the uniform distribution between $[\lambda_l, \lambda_h]$. Under the equilibrium profile, from Theorem 1: $\lambda^* = G^{-1}(\alpha) = \lambda_h\alpha + \lambda_l\beta$, which implies that $t_e = \frac{1}{\mu}(\lambda_h\alpha + \lambda_l\beta)$. The equilibrium cost c_e equals

$$\frac{1}{\mu}\int_{\lambda^*}^{\lambda_h} \bar{G}(\lambda) d\lambda + \beta\frac{\lambda^*}{\mu},$$

which in turn equals

$$\frac{\beta^2}{2\mu}(\lambda_h - \lambda_l) + \beta\frac{\lambda^*}{\mu} = \beta^2\frac{\lambda_h + \lambda_l}{2\mu} + \alpha\beta\frac{\lambda_h}{\mu}.$$

Then, $t_b = \frac{2\lambda_h\alpha\beta + (\lambda_h + \lambda_l)\beta^2}{2\mu\alpha}$.
For $t_b \le t \le T_{\lambda_l}$, $F'(t) = \frac{\mu\alpha}{E\Lambda}$.
For $T_{\lambda_l} \le t \le t_e$, $F'(t) = 2\mu^2\frac{t_e - \frac{t}{F(t)}}{{\lambda_h}^2 - \frac{\mu t}{F(t)}^2}$.
For $T_{\lambda_l} < t \le t_e$, F' is a strictly decreasing function. Hence $F(t)$ is a strictly concave function on that interval, and concave on the larger interval $[t_b, \infty)$. Furthermore,

$$F(0) = \beta^2\frac{\lambda_h + \lambda_l}{2E\Lambda} + \alpha\beta\frac{\lambda_h}{E\Lambda},$$
$$T_{\lambda_l} = \frac{F(0)}{\mu}\left[\frac{1}{\lambda_l} - \frac{\alpha}{E\Lambda}\right]^{-1},$$
$$F(T_{\lambda_l}) = \frac{F(0)}{\lambda_l}\left[\frac{1}{\lambda_l} - \frac{\alpha}{E\Lambda}\right]^{-1}.$$

Figures 2, 3 and 4 illustrate the density of the equilibrium profiles graphically. Note that for mean $(\lambda_h + \lambda_l)/2$ fixed, equilibrium cost increases with λ_h and hence with variance. This also leads to increase in deviation from uniform distribution in the arrival profile. These are depicted in Figure 2. In Figure 3, we keep the variance of the uniform distribution the same but change the mean. Thus, $(\lambda_h - \lambda_l)$ is kept fixed while $(\lambda_h + \lambda_l)$ is increased. As the mean increases, the curves can be seen to become closer to the uniform distribution, since the randomness becomes relatively less significant. In Figure 4, λ_l and λ_h are fixed but we change the cost of time to service parameter β (in all these figures $\alpha + \beta = 1$). As β increases, the customers arrive earlier and have to wait more.

Under the socially optimal profile $\lambda^* = G^{-1}(\alpha) = \lambda_h\alpha + \lambda_l\beta$ which again implies that $t_e = \frac{1}{\mu}(\lambda_h\alpha + \lambda_l\beta)$. The social cost c_s equals

$$\beta^2\frac{\lambda_h + \lambda_l}{4\mu} + \alpha\beta\frac{\lambda_h}{2\mu},$$

and for $0 \le t \le t_e$, $F'(t) = \frac{(\lambda_h\alpha + \lambda_l\beta)}{\mu}$.

VI. Conclusion

In this article we considered the concert queueing problem in the fluid framework where the arriving volumes were random. We derived the unique equilibrium arrival profile in this setting and noted that while this profile is uniformly distributed when the arrival volume is fixed, when it is random, the arrival profile is constant up to a point and thereafter tapers down. We also derived the arrival profile that minimizes the overall social welfare cost. Interestingly, this turned out to be uniformly distributed. Somewhat surprisingly, the price of anarchy remained equal to 2 even in the scenario where the arrival volumes were random.

There are many directions related to presence of uncertainty in the fluid system that require further research. For instance, how does the system behavior change when the service rates are variable, both deterministically and randomly? It would be interesting to see how random server start times impact the system behavior. One generalization that is of obvious interest is to consider heterogeneous arrivals with non-linear costs.

VII. Acknowledgments

The first author would like to thank Yahoo Research India Lab for partially funding this research. This research was additionally supported by the Israel Science Foundation, grant No. 1319/11.

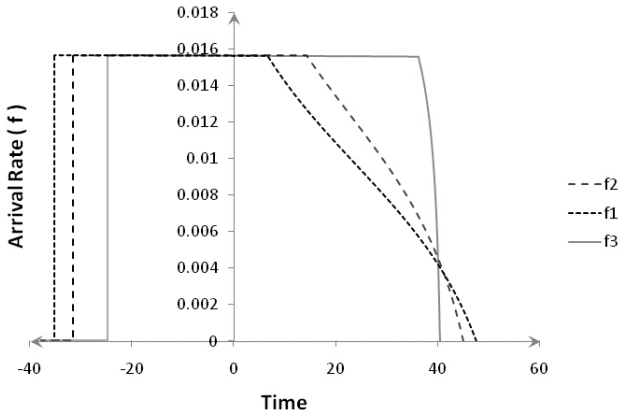

Fig. 2. G is uniformly distributed, f_1: $\lambda_l = 50$, $\lambda_h = 350$; f_2: $\lambda_l = 100$, $\lambda_h = 300$; f_3: $\lambda_l = 190$, $\lambda_h = 210$; $\beta = 3/8, \mu = 5$

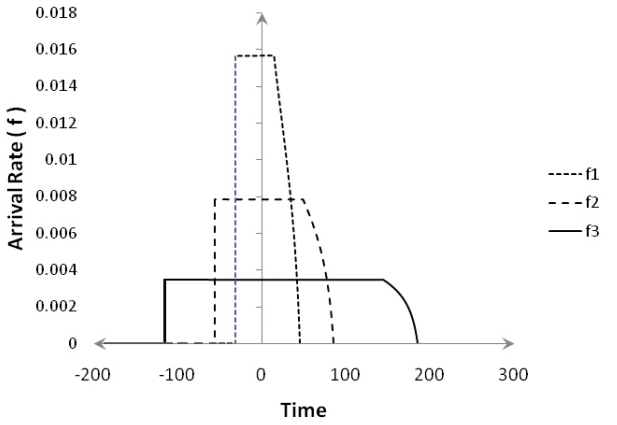

Fig. 3. G is uniformly distributed, f_1: $\lambda_l = 100$, $\lambda_h = 300$; f_2: $\lambda_l = 300$, $\lambda_h = 500$; f_3: $\lambda_l = 800$, $\lambda_h = 1000$; $\beta = 3/8, \mu = 5$

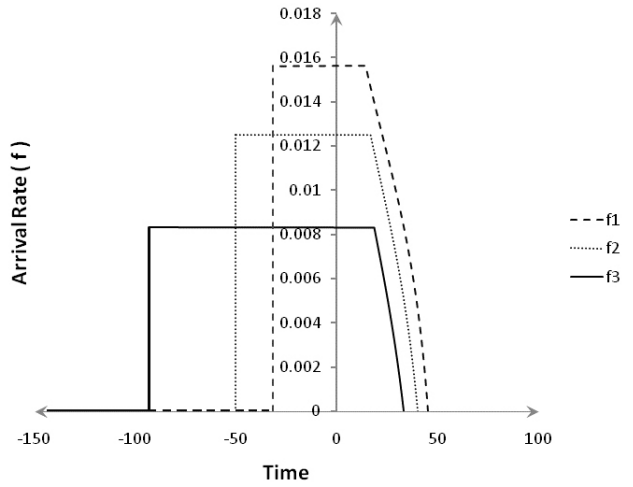

Fig. 4. G is uniformly distributed, f_1: $\beta/\alpha = 3/5$; f_2 :$\beta/\alpha = 1$;f_3 :$\beta/\alpha = 2$; $\lambda_l = 100$, $\lambda_h = 300, \mu = 5$

REFERENCES

[1] R. Arnott, A. Palma and R. Lindsey. 1999. Information and time-of-usage decisions in the bottleneck model with stochastic capacity and demand. *European Economic Review* **43**, 525-548.

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

[3] R.B. Cooper, 1981. *Introduction to Queueing Theory*, 2nd Ed., North-Holland, New York.

[4] A. Glazer and R. Hassin. 1983. ?/M/1: On the equilibrum distribution of user arrivals. *Eur. J. Oper. Res.* **13**, 146-150.

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

[6] R. Hassin and Y. Kleiner. 2011. Equilibrium and optimal arrival patterns to a server with opening and closing times. *IEE Transactions* **43**(3), 164-175.

[7] M. Haviv. 2010. When to arrive at a queue with tardiness costs? Preprint, Ocrober 2010.

[8] H. Honnapa and R. Jain. 2010. Strategic arrivals into queueing networks. *Proc. 48th Annual Allerton Conference*, Illinois, Oct. 2010, pp. 820-827.

[9] R. Jain, S. Juneja and N. Shimkin. 2011. The Concert Queuing Problem: To wait or to be late *Discrete Event Dyn. Syst.* **21**, 103-138.

[10] S. Juneja and R. Jain. 2009. The Concert/Cafeteria Queuing Problem: A game of arrivals. *Proc. ValueTools'09*, Pisa, Italy.

[11] S. Juneja and N. Shimkin. 2012. The Concert Queueing Game: Strategic arrivals with waiting and tardiness costs", to appear in *Queueing Systems*.

[12] A. Lago and C. F. Daganzo. 2007. Spillovers, merging traffic and the morning commute. *Transportation Research B* **41**, 670-683.

[13] R. Lindsey. 2004. Existence, uniquemess, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes. *Transportaion Research* **38**, 293-314.

[14] D. G. Luenberger. 1969. *Optimization by Vector Space Methods*, Wiley.

[15] D. G. Luenberger. 1997. *Linear and Nonlinear Programming*, Second Edition. Springer International.

[16] T. Roughgarden. 2005. *Selfish Routing and the Price of Anarchy*, MIT Press.

[17] W. S. Vickrey. 1969. Congestion Theory and Transport Investment. *The American Economic Review* **59**, 251-260.

[18] J. G. Wardrop. 1952. Some Theoretical Aspects of Road Traffic Research. *Proc. Inst. Civil Engineers*, Part 2, Vol. 1, 325-378.