Strategically Seeking Service: How Competition Can Generate Poisson Arrivals

Martin A. Lariviere
Jan A. Van Mieghem

Kellogg School of Management
Northwestern University
Evanston IL, 60208
www.kellogg.northwestern.edu/faculty/lariviere/research/
www.kellogg.northwestern.edu/faculty/vanmieghem/

January 2003

Abstract

We consider a simple timing game in which strategic agents select arrival times to a service facility. Agents find congestion costly and hence try to arrive when the system is under-utilized. Working in discrete time, we characterize pure strategy equilibria for the case of ample service capacity. In this case, agents try to spread themselves out as much as possible and their self-interested actions will lead to a socially optimal outcome if all agents have the same well-behaved delay cost function. For even modest sized problems, the set of possible pure strategy equilibria is quite large, making implementation potentially cumbersome. We consequently examine mixed strategy equilibria and show that there is a unique symmetric equilibrium. Not only is this equilibrium independent of the number of agents and their individual delay cost functions, the arrival pattern it generates approaches a Poisson process as the number of agents and arrival points gets large. Our results extend to the case of limited capacity given an appropriate initialization of the system. In this setting, we also argue that for any initialization, competition among customers will equate the expected workload across the horizon and thus move the system to steady state very quickly. Our model lends support to the traditional literature on strategic behavior in queuing systems. This work has generally assumed that customers arrive according to a renewal process but act strategically upon arrival. We show that assuming renewal arrivals is an acceptable assumption given a large population and long horizon.

1 Introduction

Much of economics is based on the observations that, all else being equal, people prefer the same goods at a lower cost. As most find waiting inconvenient (if not explicitly costly) a natural generalization is to assume that people prefer to avoid congestion and the concomitant delay. How rational agents respond to expected delays is consequently an active area of study. While this literature is extensive, it seems to ignore one of the most basic ways in which people attempt to avoid delays – trying to arrive when the system is under-utilized. Beginning with Naor (1969), researchers have assumed that customers arrive according to a renewal process. Customers have been assumed to be sophisticated in whether they join or balk from a queue (Naor, 1969; Yechiali, 1971; Lippman and Stidham, 1977), in how they submit work (Dewan and Mendelson, 1990; Stidham, 1992; Rump and Stidham, 1998), and in how they declare their priority class (Mendelson, 1985; Mendelson and Whang, 1990; Van Mieghem, 2000), but to our knowledge no one has considered how a delay-sensitive customer should choose an arrival time to avoid congestion. Even Rump and Stidham (1998), who examine stability as customers learn about system congestion, assume that the time between successive arrivals are independently and identically distributed over fixed intervals.

Here we present a model in which agents strategically choose *when* to seek service. Each agent wants to minimize her own delay cost. Delay costs may be agent specific. We only require that an agent's cost is an increasing function of the number of other agents seeking service at the same time. Consequently, each agent tries to pick a distinct arrival time. Working in discrete time to simplify the analysis, we show that many pure strategy equilibria exist. These equilibria are all independent of the particular delay cost function each agent has. In any equilibrium, the number of arrivals in any two periods differs by at most one, i.e., customers spread out their arrivals as much as possible. For symmetric problems, pure strategy equilibria minimize total system delay costs subject to a mild regularity condition. Self-interested agents thus implement socially efficient outcomes.[1]

While a pure strategy equilibria is attractive, its implementation is cumbersome. All

[1] This is essentially the reverse of Ostrovsky and Schwarz (2002), who suppose that all agents are needed for processing to start. Coordination in their model requires simultaneous arrivals but delay-sensitive, independent agents may choose to be late.

agents must agree on the specific equilibrium being played. Such coordination is possible if agents choose sequentially and choices are observable, i.e., an appointment system leads to a pure strategy equilibrium. However, such an arrangement may not be possible. We therefore consider mixed strategy equilibria. Again, there are possibly a large number of outcomes, but equilibria may now depend on agents' delay cost functions. We show that there is a unique symmetric equilibrium. In this equilibrium, each agent puts equal probability on every arrival time and thus the number of arrivals in any time period has a binomial distribution. The distribution of arrivals per time period is stationary over the horizon and converges to a Poisson distribution as the number of agents and time periods gets large. In the limit, the distribution of arrivals across periods is independent. Hence, the arrival pattern converges to a Poisson process as the number of agents and time periods gets large.

In our model, customers arrive according to a renewal process as a *consequence* of strategic interaction. Our work therefore complements the existing literature on strategic behavior in queuing systems by showing that assuming a renewal arrival pattern is reasonable for systems operating over a long horizon with a large number of potential customers. We also argue that when capacity is limited, competitive customers will force the expected workload to be constant over the horizon. Thus our model also supports the use of steady state analysis.

Below we first present the model and develop results assuming that agents must be served in the period in which they arrive. In the subsequent section, we suppose that agents may have to wait one or more periods in order to be served. Section 4 concludes.

2 Model basics and equilibria with ample capacity

We consider a finite horizon divided into $T \geq 2$ periods or "bins" (we will use the terms interchangeably). There are $M \geq 2$ agents or customers who seek service over the horizon by choosing an arrival bin. Let $\alpha_t = 1, ..., M$ be the number of customers arriving to bin t. Let $\lambda = M/T$ be the average number of customers per bin. For simplicity, we adopt the convention that all customers arrive at the start of the time period. For now we assume that the system has ample capacity, i.e., there is sufficient resources to serve all M customers in one time period. A customer arriving in time period t is therefore certain to receive service in that time period. No customers remain in the system from period t to period $t+1$, so the only customers in the system during period t are the α_t who arrive in that period.

Customer m values service at $V_m > 0$ regardless of the period in which she is served. All customers prefer to avoid congestion. Let $W_m(\alpha_t)$ denote agent m's expected disutility of being one of α_t arrivals in period t. (α_t includes agent m.) Agent m's objective is to maximize her net utility $U_m(\alpha_t) = V_m - W_m(\alpha_t)$ through her choice of arrival bin t. Equivalently, agent m seeks to minimize her expected congestion or delay cost $W_m(\alpha_t)$. To avoid trivialities, assume $U_m(1) > 0$; agent m expects a positive benefit if she is the only arrival.

This formulation embeds an important assumption. Since V_m is independent of t, agent m's net utility depends on the time bin only through the number of arrivals α_t. Thus we are assuming non-urgent or postponable service for which agents do not care about the time bin t; they only care about the wait they encounter after arriving.

We allow for significant flexibility in modeling the expected delay cost $W_m(\alpha_t)$; all we require is that $W_m(\alpha_t)$ is strictly increasing in α_t, the number of arrivals to bin t. We will emphasize two basic forms of W_m depending on whether the agents are served sequentially or in a batch. In the former case, agent m incurs a cost $C_m(k) \geq 0$ for $k = 1, ..., \alpha_t$ if she is the k^{th} customer to be served for some function C_m. Upon arrival, the agents are randomly ordered such that each agent has an equal probability of being in any position in line. Given $\alpha_t \geq 1$, agent m's expected delay costs are:

$$W_m(\alpha_t) = \frac{1}{\alpha_t} \sum_{k=1}^{\alpha_t} C_m(k). \tag{1}$$

We assume that $C_m(k)$ increases sufficiently fast that $W_m(\alpha_t)$ is increasing.[2]

We work with two special cases of (1). First suppose that delay costs are linear. $C_m(k) = \theta_m(k-1)$ for $\theta_m > 0$. We then have:

$$W_m(\alpha_t) = \frac{1}{\alpha_t} \sum_{i=1}^{\alpha_t} \theta_m(k-1) = \theta_m \frac{\alpha_t - 1}{2}. \tag{2}$$

Alternatively, delay costs can increase at an increasing rate. Suppose $C_m(k) = \rho_m^{(k-1)} - 1$ for $\rho_m > 1$. The resulting expected delay costs are:

$$W_m(\alpha_t) = \frac{1}{\alpha_t} \sum_{k=1}^{\alpha_t} \rho_m^{(k-1)} - 1 = \frac{\rho_m^{\alpha_t} - 1}{\alpha_t(\rho_m - 1)} - 1. \tag{3}$$

In a batch setting, all arrivals are served simultaneously but the quality of service falls

[2] This requires that $C_m(\alpha+1) \geq \alpha^{-1} \sum_{k=1}^{\alpha} C_m(k)$ for $\alpha = 1, ..., M-1$.

with the number of agents being served:

$$W_m(\alpha_t) = \theta_m - \frac{\theta_m}{\alpha_t} \text{ for } \theta_m > 0. \tag{4}$$

Note that if m's delay cost is (2), the agent only cares about the average number arrivals to bin t and thus behaves as if she were risk neutral. In contrast, (3) is convex in the number of arrivals, making the agent's net utility concave in arrivals.[3] Hence, such an agent is risk averse and prefers to avoid to variability. The reverse holds in a batch setting. The agent's delay cost is concave and her net utility is convex. With batch costs, an agent is risk seeking and prefers variability in the number of arrivals. We exploit these properties below.

2.1 Pure-strategy equilibria

Agent m would like to seek service in the bin that would maximize her expected net benefit, $U_m(\alpha_t)$. Of course, that benefit depends on the actions of the other agents. We must look for equilibrium arrival patterns. For now we restrict our attention to pure strategy equilibria in which each agent reports deterministically to one bin. Defining such an equilibrium requires some notation. Let e_j denote j^{th} $T\times 1$ unit vector and let π_m denote agent m's strategy. $\pi_m = e_t$ if agent m reports to bin t. Let $\Pi = (\pi_1, ..., \pi_M)$ and $\Pi_{-m} = (\pi_1, ..., \pi_{m-1}, \pi_{m+1}, ...\pi_M)$. Define $\alpha(\Pi)$ as the arrival vector that results from Π:

$$\alpha(\Pi) = (\alpha_1(\Pi), ..., \alpha_T(\Pi)) = \sum_{j=1}^{M} \pi_j$$

and $\alpha(\Pi_{-m})$ as the arrival vector of all agents but m:

$$\alpha(\Pi_{-m}) = (\alpha_1(\Pi_{-m}), ..., \alpha_T(\Pi_{-m})) = \sum_{j\neq m}^{M} \pi_j.$$

Finally, let $B_m(\Pi) = t$ if $\pi_m = e_t$. We can now state that Π^* is a pure strategy Nash equilibrium if the following holds for $m = 1, ..., M$:

$$V_m - W_m\left(\alpha_{B_m(\Pi^*)}(\Pi^*)\right) \geq V_m - W_m\left(\alpha_t\left(\Pi^*_{-m}\right) + 1\right) \quad \text{for } t = 1, ..., T.$$

In words, a pure strategy equilibrium requires that holding the proposed actions of others fixed, an agent cannot improve her payoff by unilaterally moving to a different arrival bin.

[3] Convexity [concavity] is overly restrictive. W_m is defined for natural numbers and need not be continuous. We use convexity [concavity] as short hand for increasing [decreasing] first differences, i.e., $W_m(\alpha+1) - W_m(\alpha) \geq [\leq] W_m(\alpha) - W_m(\alpha-1)$.

Theorem 1 *An arrival vector* Π^* *is a pure strategy Nash equilibrium if and only if* $\alpha_t(\Pi^*) - \alpha_s(\Pi^*) \leq 1$ *for all* $s, t = 1, ..., T$.

Proof: Suppose that Π^* is a Nash equilibrium and that $B_m(\Pi^*) = t$. Consider agent m's incentive to deviate. As W_m is increasing, she has no interest in moving to any bin s such that $\alpha_s(\Pi^*) \geq \alpha_t(\Pi^*)$. Suppose there is a bin s' such that $\alpha_t(\Pi^*) > \alpha_{s'}(\Pi^*)$. Deviating is profitable if $\alpha_t(\Pi^*) > \alpha_{s'}\left(\Pi^*_{-m}\right) + 1 = \alpha_{s'}(\Pi^*) + 1$. Thus if Π^* is an equilibrium, it must be the case that $\alpha_t(\Pi^*) - \alpha_s(\Pi^*) \leq 1$ for all $s, t = 1, ..., T$.

Now suppose $\alpha_t(\Pi^*) - \alpha_s(\Pi^*) \leq 1$ for all $s, t = 1, ..., T$ and $B_m(\Pi^*) = t$. If she were to move to bin s, the number of arrivals to s would be $\alpha_s\left(\Pi^*_{-m}\right) + 1 = \alpha_s(\Pi^*) + 1 \geq \alpha_t(\Pi^*)$. Hence she has no reason to unilaterally deviate from t, and Π^* is an equilibrium. ■

In a pure strategy equilibrium each bin must have a "Yogi-Berra" property. In equilibrium, no one goes there anymore; it's too crowded. Each agent is content to report to her assigned bin because, holding everyone else's decision constant, she cannot find a less congested one. Compared to her current assignment every other bin is either as crowded, more crowded, or will become as crowded if she were to deviate by moving to it. Just as no agent can single-handedly lower her congestion cost, there is no way a central planner can lower system cost assuming that all agents have the same well-behaved congestion cost function.

Theorem 2 *Suppose that all agents have the same delay cost function* $W(\alpha)$. *Define* $S(\alpha) = \alpha W(\alpha)$ *as the system cost of assigning* α *agents to one bin. If* $S(\alpha)$ *is convex, a pure strategy Nash equilibrium minimizes total system congestion costs.*

Proof: If $S(\alpha)$ is convex, $S(\alpha) + S(\alpha') \geq S(\alpha - 1) + S(\alpha' + 1)$ for $\alpha > \alpha' + 1$. Thus for any assignment of agents to bins such that $\alpha_t > \alpha_s + 1$ for some bins s and t, cost can be lowered by transferring agents from t to s. This is feasible unless $\alpha_t - \alpha_s \leq 1$ for all $s, t = 1, ..., T$, which is the condition for a Nash equilibrium. ■

Symmetric, self-interested agents will choose a socially efficient outcome that minimizes total waiting costs. The required condition is fairly weak. It is obviously satisfied by any weakly convex delay cost function. Some concave cost functions such as (4) also satisfy it.

While we have characterized pure strategy equilibria, we have not proved that one necessarily exists. Theorem 1 does offer guidance on how we may construct one. Let $\lfloor x \rfloor$ denote the smallest integer less than or equal to x. Let $\underline{\lambda} = \lfloor \lambda \rfloor$ and $\tau = M - \underline{\lambda} T$. Clearly, $0 \leq \tau < T$. In words, if one assigns $\underline{\lambda} T$ customers such that each bin has $\underline{\lambda}$ agents, there are τ agents still to be assigned. An equilibrium vector Π_1^* can then be formed by selecting τ bins and adding one agent to each of those bins. As all agents are assigned an arrival time and each agents is assigned to only one bin, the vector is feasible. As the maximum difference in arrivals is one, the vector is an equilibrium.

Since $T \geq 2$ and $M \geq 2$, it is immediate that the equilibrium Π_1^* is not unique. If one merely interchanges an agent assigned to bin t with an agent assigned to bin $s \neq t$, one has created a new equilibrium Π_2^*. Indeed, the set of pure strategy equilibria can be quite large.

Lemma 1 *Let $\bar{\Pi}$ denote the set of all possible pure strategy equilibria. Let $\bar{\Pi}_t^m$ denote the set of equilibria that assigns agent m to bin t. Let $\|X\|$ denote the cardinality of the set X.*

1. $\left\|\bar{\Pi}\right\| = \frac{M!}{\underline{\lambda}!(T-\tau)\max\{(\underline{\lambda}+1)!\tau,1\}}\binom{T}{\tau}$.
2. $\frac{\left\|\bar{\Pi}_t^m\right\|}{\left\|\bar{\Pi}\right\|} = \frac{1}{T}$.

Proof: The first part is proved by induction on T. Suppose that $T = 2$ and that $\underline{\lambda}$ agents are assigned to the first bin. $\frac{M!}{\underline{\lambda}!(T-\tau)\max\{(\underline{\lambda}+1)!\tau,1\}} = \binom{M}{\underline{\lambda}}$ gives the number of equilibria given that there are $\underline{\lambda}$ agents to the first bin while $\binom{T}{\tau}$ then accounts for the number of ways to select τ bins to have $\underline{\lambda}+1$ agents. Suppose the result holds for $T-1$ bins and consider the case of T bins. Fix the τ bins assigned $\underline{\lambda}+1$ agents and assume that the first bin has only $\underline{\lambda}$ agents. The number of equilibria vectors satisfying these criteria are:

$$\binom{M}{\underline{\lambda}}\frac{(M-\underline{\lambda})!}{\underline{\lambda}!\,(T-1-\tau)\max\{(\underline{\lambda}+1)!\tau,1\}} = \frac{M!}{\underline{\lambda}!\,(T-\tau)\max\{(\underline{\lambda}+1)!\tau,1\}}.$$

The second term relies on the inductive hypothesis, and the result follows.

For the second part, pick an element of $\bar{\Pi}_t^m$. Exchanging all agents assigned to bin t for all those assigned to bin $s \neq t$ creates an equilibrium outside of $\bar{\Pi}_t^m$. For each element of $\bar{\Pi}_t^m$, we can thereby create $T-1$ elements of $\bar{\Pi}$ outside of $\bar{\Pi}_t^m$. Running the procedure in reverse for each element of $\bar{\Pi}$ not in $\bar{\Pi}_t^m$ assures that no equilibria are unaccounted for. ■

With such a large number of potential equilibria, implementation is potentially an issue. The following offers a possible means for selecting an equilibrium.

Theorem 3 *Suppose agents select bins sequentially. Agent m observes all prior arrivals before choosing a bin. If agent m is indifferent between bins, assume she chooses randomly between bins with equal probability. This procedure leads to an element of $\bar{\Pi}$.*

Proof: For the last agent, there must be available at least one bin with $\underline{\lambda}$ or fewer agents. She prefers this to any bin with $\underline{\lambda}+1$ or more agents. Similarly, no earlier agent will choose a bin already occupied by $\underline{\lambda}+1$ agents. Thus, the final arrangement of agents will have bins with either $\underline{\lambda}$ or $\underline{\lambda}+1$ agents and must be an equilibrium by Theorem 1. ■

Customers benefit from an appointment system that enforces sequential choice with visibility because it always implements an equilibrium vector. If the initial sequencing of agents is random, the assumption that indifferent agents randomize implies that any element of $\bar{\Pi}$ is a feasible outcome. This does not hold for other behavioral assumptions. Suppose instead that an agent choosing between equally crowded bins always selects the earliest bin. Then if $M > T$, the first $\underline{\lambda}(T-\tau)$ agents always opt to arrive in the last $(T-\tau)$ time periods because they anticipate that the final τ agents will choose to arrive in one of the first τ periods. Hence, the first τ bins will always be the "crowded" bins with $\underline{\lambda}+1$ occupants.

2.2 Mixed strategy equilibria

Sequential choice requires less upfront coordination and information than picking an arbitrary pure strategy, but it may not always be feasible. To coordinate on a pure strategy, it is still necessary to sequence the agents and take appointments. In addition, the agents must know the total number of customers attempting to use the system. Consequently, a mixed strategy equilibrium may be a more plausible outcome. Here agents randomize over their possible actions, i.e., which bin to select, so an agent's strategy is a probability distribution over the set of bins. Hence, we now have $\pi_m = \left(\pi_m^1, ..., \pi_m^T\right)$ for $\pi_m^t \geq 0$ and $\sum_{t=1}^T \pi_m^t = 1$. Note that the pure strategies considered above are subsumed in this formulation by considering degenerate distributions (i.e., by allowing π_m to be a unit vector). The set of mixed strategy equilibria is thus at least as large as the set of pure strategy equilibria.

The number of arrivals to bin t is now a random variable that depends on the agents' strategies Π. Let $\alpha_t(\Pi)$ denote this random variable. Let $\alpha_t(\Pi_{-m})$ denote the number of arrivals to t excluding agent m. $\alpha_t(\Pi)$ takes values from zero to M while $\alpha_t(\Pi_{-1})$ takes values from zero to $M-1$. Note that if agent m's realized bin choice is t, her expected delay costs are $E\left[W_m\left(\alpha_t\left(\Pi_{-m}\right)+1\right)\right]$. We can therefore define $\mathcal{W}_m(\Pi)$, m's expected delay cost when all agents follow mixed strategies Π, as:

$$\mathcal{W}_m(\Pi) = \sum_{t=1}^{T} \pi_m^t E\left[W_m\left(\alpha_t\left(\Pi_{-m}\right)+1\right)\right].$$

We say that a random variable X is stochastically larger than a random variable Y if $P(X \leq \beta) \leq P(Y \leq \beta)$ for all β. X is stochastically larger than Y if and only if $E[\phi(X)] \geq$

$E[\phi(Y)]$ for all increasing ϕ such that the expectation is defined (Shaked and Shanthikumar, 1994). The following lemma is then an immediate consequence of W_m being increasing.

Lemma 2 *If the distribution of $\alpha_t(\Pi_{-m})$ is stochastically larger than $\alpha_s(\Pi_{-m})$, then agent m prefers arriving in bin s.*

For $\Pi^* = (\pi_1^*, ..., \pi_M^*)$ to be a mixed strategy equilibrium, we must have for $m = 1, ..., M$:

$$V_m - \mathcal{W}_m(\Pi) \geq V_m - E\left[W_m\left(\alpha_t\left(\Pi_{-m}^*\right) + 1\right)\right] \quad \text{for } t = 1, ..., T. \tag{5}$$

(See Fudenberg and Tirole, 1996.) This definition requires that an agent must weakly prefer randomizing over a set of actions to going to any one bin with certainty. Thus, she must be indifferent among any actions on which she puts positive probability.[4] We now consider an example to illustrate the nature of a mixed strategy equilibrium

Example 1 Suppose $T > 2$, and $\underline{\lambda} = \lambda \geq 2$. (The latter implies $\tau = 0$.) Four agents are selected to randomize, placing equal weight on bins 1 and 2. For each bin $t > 2$, λ agents are selected and deterministically report to bin t. If agents remain, $\lambda - 2$ agents deterministically report to bin 1 and $\lambda - 2$ report to bin 2. Note that $E[W_m(\alpha_t(\Pi))] = \lambda$ for all t. We claim this forms a mixed strategy equilibrium if all agents have an expected delay cost function as given in (4), i.e., $W(\alpha) = C - C/\alpha$.

To establish this we need to verify that three types of agents are willing to follow the proposed equilibrium. Type 1 agents report deterministically to bins 1 or 2. Type 2 agents randomize between 1 and 2. Type 3 agents report deterministically to some bin $t > 2$. Let m_i denote a representative agent of each type for $i = 1, 2, 3$. First consider a type 1 agent reporting deterministically to bin $j = 1$ or 2. Such an agent has no interest in moving to bin $3 - j$ by Lemma 2. Next, since $\alpha_t(\Pi_{-m_1})$ is weakly stochastically smaller than $\alpha_t(\Pi_{-m_2})$ for all t, a type 2 agent has higher delay costs from following the proposed equilibrium than a type 1 agent. Hence, if type 2 agents follow the equilibrium, so will type 1 agents.

For type 2 agents, $E[\alpha_t(\Pi_{-m_2})] = \lambda - \frac{1}{2}$ for $t = 1, 2$. The agent prefers the proposed equilibrium to having $\alpha_t(\Pi_{-m_2}) = \lambda - \frac{1}{2}$ deterministically (because W is concave). Thus she prefers participating in the equilibrium to deviating to $t > 2$ for which $\alpha_t(\Pi_{-m_2}) = \lambda$.

Finally, a type 3 agent's cost from following the strategy is $W(\lambda)$ with certainty. Additionally, $\alpha_t(\Pi_{-m_3}) - (\lambda - 2)$ has a binomial distribution with parameters $(4, 1/2)$. Therefore, if she deviates to bin 1 or 2, her costs increase by:

$$\begin{aligned}&\frac{W(\lambda-1) + 4W(\lambda) + 6W(\lambda+1) + 4W(\lambda+2) + W(\lambda+3)}{16} - W(\lambda) \\ &\qquad = C\left(-9 - 2\lambda + 6\lambda^2 + 2\lambda^3\right)/2(\lambda-1)\lambda(\lambda+1)(\lambda+2)(\lambda+3),\end{aligned}$$

which is positive since $\lambda \geq 2$. Hence, type 3 agents also follow the equilibrium.

[4] To see this classical result, multiply both sides of (5) by π_m^t and sum over t.

While all pure strategy equilibria are independent of the assumed delay cost functions, this mixed strategy equilibrium depends on them in a crucial way. Agents are assumed to have a concave cost function and so behave in a risk-seeking fashion. Suppose, on the other hand, that a type 2 agent has a cost function as given in (3) with $\rho = 10$, i.e., $W_{m_2}(\alpha) = \frac{10^{\alpha}-1}{9\alpha} - 1$. It is possible to show that such an agent always prefers to deviate to bin $t \geq 3$ for any $\lambda > 2$. Thus, the feasibility of an equilibrium depends critically on the delay cost function. In this case, the equilibrium can collapse if agents have convex costs and so act in a risk-averse manner. Given this observation, it useful to consider whether any mixed strategy equilibrium can be independent of the agents' delay costs.

Theorem 4 *An equilibrium Π^* is independent of all agents' delay cost functions if and only if for $m = 1, ..., M$:*

1. $\alpha_t\left(\Pi^*_{-m}\right)$ is weakly stochastically smaller than $\alpha_s\left(\Pi^*_{-m}\right)$ for all $s, t \in \{1, ..., T\}$ such that $\pi^t_m > 0$ and $\pi^s_m = 0$.
2. For all $t, u \in \{1, ..., T\}$ such that $\pi^t_m > 0$ and $\pi^u_m > 0$:
$$P(\alpha_t\left(\Pi^*_{-m}\right) = \beta) = P(\alpha_u\left(\Pi^*_{-m}\right) = \beta) \quad (6)$$
for $\beta = 0, ..., M-1$.

Proof: First, suppose both conditions hold. By Lemma 2, agent m prefers t to s regardless of W_m. In addition, the second condition assures that she is indifferent between between the bins over which she randomizes for all W_m. Hence, Π^* is an equilibrium for all W_m. To go the other way, $\alpha_t\left(\Pi^*_{-m}\right)$ is stochastically smaller than $\alpha_s\left(\Pi^*_{-m}\right)$ if and only if $E\left[\phi\left(\alpha_t\left(\Pi^*_{-m}\right)\right)\right] \leq E\left[\phi\left(\alpha_s\left(\Pi^*_{-m}\right)\right)\right]$ for all increasing ϕ (Shaked and Shanthikumar, 1994). Consequently, if Π^* is independent of W_m for any increasing W_m, condition 1 must hold. Next, suppose (6) fails for some agent m and some bins t and u. There must exist $\beta < \beta'$ such that

$$\begin{aligned} P(\alpha_t\left(\Pi^*_{-m}\right) &= \beta) > P(\alpha_u\left(\Pi^*_{-m}\right) = \beta) \\ P(\alpha_t\left(\Pi^*_{-m}\right) &= \beta') < P(\alpha_u\left(\Pi^*_{-m}\right) = \beta') \end{aligned}$$

Since m randomizes over t and u, they must result in the same expected congestion cost:

$$E\left[W_m\left(\alpha_t\left(\Pi^*_{-m}\right)+1\right)\right] = E\left[W_m\left(\alpha_u\left(\Pi^*_{-m}\right)+1\right)\right].$$

Pick a W_m such that this equality holds. Since W_m is strictly increasing, we can create a new cost function $\tilde{W}_m$ such that $\tilde{W}_m(\beta+1) = W_m(\beta+1)+\varepsilon$, $\tilde{W}_m(\beta'+1) = W_m(\beta'+1)-\varepsilon$, and $\tilde{W}_m = W_m$ otherwise. We must then have:

$$E\left[\tilde{W}_m\left(\alpha_t\left(\Pi_{-m}^*\right)+1\right)\right] > E\left[\tilde{W}_m\left(\alpha_u\left(\Pi_{-m}^*\right)+1\right)\right],$$

so the equilibrium is not independent of the cost function. ■

Neither condition of the theorem is trivial. The first asserts that for a mixed strategy to be independent of delay costs, it must be that an agent randomizes over what she perceives as less crowded bins. However, the bins cannot be too different. One can show that if Π^* is a mixed strategy equilibrium for any W, we must have $E\left[\alpha_t\left(\Pi^*\right)\right] - E\left[\alpha_s\left(\Pi^*\right)\right] \leq 1$ for all $s, t = 1, ..., T$. Thus a condition similar to that of Theorem 1 must hold in expectation. The second asserts that agents must view bins over which they randomize as identical. The following modifies the previous example to be independent of delay cost functions.

Example 2 Consider a setting similar to Example 1 but suppose that only two agents randomize between the first two bins while $\lambda - 1$ agents report to those bins deterministically. Define type 1, 2, and 3 agents as before. From the perspective of a type 1 or 2, the maximum value $\alpha_t\left(\Pi_{-m}^*\right)$ takes for $t = 1, 2$ is λ while $\alpha_s\left(\Pi_{-m}^*\right) = \lambda$ for all $s > 2$. In addition, the distributions of $\alpha_1\left(\Pi_{-m}^*\right)$ and $\alpha_2\left(\Pi_{-m}^*\right)$ are identical. For a type 3 agent reporting to bin $t > 2$, $\alpha_t\left(\Pi_{-m}^*\right) = \lambda - 1$, which is the smallest possible value of $\alpha_s\left(\Pi_{-m}^*\right)$ for $s = 1, 2$. Hence, this is an equilibrium for any W.

The above example shows that even if a mixed strategy equilibrium is independent of delay cost functions it is not necessarily easier to implement than a pure strategy equilibrium. In this case, we must divide the agents into three groups and within groups 1 and 3 we must assign individual agents to particular bins. It would be much simpler to have a symmetric equilibrium that is independent of the agents' congestion cost function. In a symmetric equilibrium, $\pi_m^* = \pi_n^*$ for all agents m and n. We note the following.

Lemma 3 *If Π^* is a symmetric equilibrium, all elements of π_m^* must be strictly positive.*

Proof: If $\pi_m^{t*} = 0$ for some t, any agent could deviate to t and have no delay. ■

Let $\pi_m^U = \{1/T, ..., 1/T\}$ and $\Pi^U = \left(\pi_1^U, ..., \pi_M^U\right)$. It is straightforward to verify that Π^U is a mixed strategy equilibrium.

Theorem 5 Π^U *is the only symmetric equilibrium. It is independent of the agents' delay cost functions.*

Proof: Any other symmetric equilibrium must have $\pi_m^t > \pi_m^s$ for some bins t and s. $\alpha_t(\Pi_{-m})$ would be stochastically larger than $\alpha_s(\Pi_{-m})$ for all m. Agents would prefer s to t so the equilibrium fails. It is easy to verify that Π^U satisfies the requirements of Theorem 4. ■

The theorem establishes that equilibrium Π^U is an attractive alternative since it requires minimal coordination between the agents. It is independent of their delay cost functions and is even independent of the number of agents. Given Lemma 1, an alternative interpretation of the equilibrium is that each agent randomly selects a pure strategy Π from the set $\bar{\Pi}$ and then reports to the bin to which she is assigned under Π.

In some sense, Π^U is an obvious outcome. If an agent is completely uniformed – unsure of how many others will seek service or of their delay cost – how could she do better than uniformly picking among the bins? It is important to recognize that such reasoning depends on strategically anticipating the actions of others. Uniformly randomizing only makes sense if others are doing it as well. One might reasonably conjecture that the other agents are "early birds" or are prone to procrastination, but uniformly picking an arrival time is then no longer optimal. Thus, the arrival pattern generated by Π^U depends on the strategic interaction between the agents.

It is straightforward to see that under Π^U, the distribution of arrivals to bin t is now a binomial distribution with mean λ. The arrival pattern, $\alpha\left(\Pi^U\right)$, it generates thus has some similarity to a Poisson process. Its increments are stationary. If one is told that over some subset of bins there have been, say, k arrivals, then those arrivals are uniformly distributed across the bins. Where the similarity fails is independence. Since there are a finite number of agents, the arrivals in distinct bins are not independent. Indeed, it is easy to show that the joint distribution of arrivals is a multinomial distribution. However, we can relax this as the number of agents and time periods gets large.

Theorem 6 *Consider a sequence of system (M_n, T_n) such that M_n and T_n are integers for all n, $M_{n+1} > M_n$, $\lim_{n\to\infty} M_n = \infty$, and $\frac{M_n}{T_n} = \lambda$ for all n. Let Π_n^U denote the equilibrium in which all M_n agents play $\{1/T_n, ..., 1/T_n\}$.*

1. $\lim_{n\to\infty} P\left(\alpha_t\left(\Pi_n^U\right) = k\right) = \frac{e^{-\lambda}\lambda^k}{k!}$.
2. Let Ω be a set of bins such that $\|\Omega\| = \omega < T_1 - 1$. For all $s \in \Omega$, one knows the realized value of α_s and $\sum_{s\in\Omega} \alpha_s = A < M_1$. Let $P\left(\alpha_t\left(\Pi_n^U\right) = k|\Omega\right)$ denote the probability that $\alpha_t\left(\Pi_n^U\right) = k$ given Ω for $t \notin \Omega$. Then $\lim_{n\to\infty} P\left(\alpha_t\left(\Pi_n^U\right) = k|\Omega\right) = \frac{e^{-\lambda}\lambda^k}{k!}$.

Proof: The first part is a standard result on the limit of a binomial distribution (Ross, 1983). More details are required for the second part since $E\left(\alpha_t\left(\Pi_n^U\right)=k|\Omega\right)$ does not necessarily equal λ. First, it is possible to show that the joint distribution of arrivals for states outside of Ω given Ω is a multinomial distribution that depends on Ω only through ω and A. The marginal distribution of $\alpha_t\left(\Pi_n^U\right)$ is then $P\left(\alpha_t\left(\Pi_n^U\right)=k|\Omega\right)=\binom{M_n-A}{k}\left(1-\frac{1}{T_n-\omega}\right)^{M_n-A-k}\left(\frac{1}{T_n-\omega}\right)^k$. Since $\frac{M_n}{T_n}=\lambda$, we have

$$P\left(\alpha_t\left(\Pi_n^U\right)=k|\Omega\right)=\left(\frac{(M_n-A)!}{(M_n-A-k)!\,(M_n-\omega\lambda)^k}\right)\left(1-\frac{\lambda}{M_n-\omega\lambda}\right)^{M-A-k}\frac{\lambda^k}{k!}.$$

The first term goes to 1 as n gets large while the second goes to $e^{-\lambda}$. ■

Thus agents under equilibrium Π^U generate an arrival pattern that approaches a Poisson process, i.e., customers show up according to a renewal process. This result complements earlier work on queuing systems with strategic customers that assume renewal arrivals. We have shown that if one allows customers to also act strategically in choosing their arrival time a plausible equilibrium leads to a renewal process.

We emphasize that this result depends critically on the strategic interactions between customers. A given agent chooses an arrival point in such a way that total arrivals have a Poisson distribution because of how the other agents choose to play. Not all mixed strategies lead to Poisson outcomes. Consider Example 2. If one scales up that example, the equilibrium continues to hold but does not lead to Poisson arrivals. Thus our result also goes beyond the usual interpretation that a Poisson process arises from having a large number of agents acting independently. Yes, equilibrium Π^U assumes that each agent picks an arrival bin independently of all others but the same is true for any mixed strategy equilibria and not all mixed strategy equilibria lead to Poisson arrivals. Not only does equilibrium Π^U lead to Poisson arrivals it has a number of other appealing properties. It is the only symmetric equilibrium and is independent of the delay cost functions and the number of agents. Hence, it requires very little coordination to implement.

Theorem 6 is a limiting result, but assuming Poisson arrivals can be a reasonable approximation for Π^U at finite values. In Table 1, we consider how quickly $P\left(\alpha_t\left(\Pi_n^U\right)=k\right)$ converges to a Poisson distribution for two examples. Both begin with $T_1=25$ but differ in the arrival intensity. In the first $\lambda=2$ and in the second $\lambda=0.6$. We then increase the number of bins and agents. We report the maximum absolute deviation in the probability mass function (PMF) and in the cumulative distribution function (CDF). We see that the lower arrival rate converges more slowly but is still is closely approximated by a Poisson process. With just 50 bins and 30 agents, the maximum deviation is less than half a percent.

Table 2 examines how quickly the dependence between bins diminishes. We focus on a setting with 400 agents and 200 bins and examine the distribution conditional on knowing the arrivals for the first 50, 100, and 150 bins. As the conditional distribution depends only on the cumulative number of arrivals, we consider having arrivals run 10 and 20 agents above or below their expected value. We report the maximum absolute deviation between the PMF and the CDF of $\alpha_t\left(\Pi_n^U\right)$ given Ω and a Poisson distribution with $\lambda = 2$.[5] As one might expect, the fit is not as close as for the unconditional distribution, but it remains a reasonable approximation as long as one does not know "too much." When one has observed 150 bins (75% of the horizon) and the arrivals deviate significantly from the average, the Poisson is not a close fit (the maximum deviation is over 5%), but when observing fewer bins and having smaller deviations from the mean, the fit is much tighter.

2.3 Extensions

We briefly consider some extensions for which Π^U continues to be an equilibrium.

2.3.1 A finite waiting room

We have thus far assumed that an agent arriving to bin t can enter and be served (although possibly incurring a significant delay cost). We now suppose that the system can only hold K customers at one time. Customers are randomly sequenced upon arrival, and the first K are admitted and served according to the established sequence. Any remaining customers are denied service. If agent m is admitted, she incurs a cost $C_m(k)$ if she is the k^{th} person processed. Assume that $V_m \geq C_m(K)$ so that she always values service if admitted. If she is denied admission, she incurs a cost $\omega_m \geq 0$ and does not receive V_m. Her expected net benefit given α_t arrivals is then:

$$U_m(\alpha_t) = \begin{cases} \frac{1}{\alpha_t}\sum_{k=1}^{K}\left[V_m - C_m(k)\right] & if\ \alpha_t \leq K \\ \frac{K}{\alpha_t}\left(\frac{1}{K}\sum_{k=1}^{K}\left[V_m - C_m(k)\right]\right) - \frac{\alpha_t - K}{\alpha_t}\omega_m & if\ \alpha_t > K \end{cases}.$$

This net benefit function models a system with a finite waiting room and first in first out (FIFO) service. An agent denied entry receives no net benefit from service and may incur

[5] Thus if one observes 100 bins and sees total arrivals of 210, we compare a Poisson distribution with $\lambda = 2$ to a binomial with parameters $(400 - 210)$ and $1/(200 - 100)$.

a loss. One could specify a similar cost function for the batch model of (4). As $U_m(\alpha_t)$ decreases in the α_t, our analysis is unchanged, and Π^U is still an equilibrium.

2.3.2 Exogenous arrivals

Suppose that in each bin there is a stochastic shock in addition to the arrival of the agents. Let $X = (X_1, ..., X_T)$ denote the vector of shocks. Such a shock could be additional customers arriving from some other source. Agents choose their bins before the shocks are observed. If the realized number of agents is α_t, then agent m's congestion costs are $W_m(\alpha_t + x_t)$ where x_t is the realized value of X_t. If the marginal distribution of the shocks is the same for all bins, our analysis goes through. For pure strategies, agents spread themselves out as much as possible. For mixed strategies, Π^U is again viable. Note that we require the marginal distributions to be the same but do not require independence. Thus we could have the arrivals in bins being positively or negatively correlated.

2.3.3 Priorities

Suppose there are two classes of agents. There are M_1 type 1 agents and M_2 type 2 agents. Let α_t^i be the number of type i agents arriving to bin t. Suppose m_1 is a type 1 agent with congestion costs $W_{m_1}(\alpha_t^1, \alpha_t^2)$. We assume that $\frac{\partial W_{m_1}(\alpha_t^1,\alpha_t^2)}{\alpha_t^1} > \frac{\partial W_{m_1}(\alpha_t^1,\alpha_t^2)}{\alpha_t^2} = 0$. Let m_2 be a type 2 agent with congestion costs $W_{m_2}(\alpha_t^1, \alpha_t^2)$. We assume that $\frac{\partial W_{m_2}(\alpha_t^1,\alpha_t^2)}{\alpha_t^1} \geq \frac{\partial W_{m_2}(\alpha_t^1,\alpha_t^2)}{\alpha_t^2} > 0$. These cost functions are consistent with a priority scheme that serves type 1 arrivals before type 2 arrivals. A type 1's net utility is unaffected by the arrival of a type 2 customer, but type 1 arrivals always lowers the utility of a type 2 customer.

Consider first the behavior of type 1 agents. Since their waits are independent of the actions of the second type, they face the same problem we analyzed above. Now consider type 2 customers. If type 1 agents are playing a pure strategy, suppose that the second type cannot verify which strategy (i.e., they do not know which bins have $\underline{\lambda_1}$ agents and which have $\underline{\lambda_1} + 1$ agents. Then they can treat the number of arrivals of type one agents as an exogenous shock as discussed above. Alternatively, if type 1 agents follow mixed strategy equilibrium Π^U, we again have an exogenous shock. In either setting, type 2 agents are willing to employ equilibrium Π^U.

3 Equilibria with limited capacity

We now suppose that the system can only serve a limited number of customers in each period and allow customers unserved in period t to carry over to period $t+1$. Our intention is to develop conditions such that Π^U is again an equilibrium. If this is the case, Theorem 6 continues to hold and a Poisson process is again a valid approximation for the arrival process generated by strategic customers.

Let I_t denote the inventory of customers in the system at the start of period t prior to any new arrivals. Let s_t denote the maximum number of customers that can be served in period t. The assumed sequence of events is that I_t agents are carried into period t, α_t new customers arrive, and then s_t is realized enabling $\max\{I_t + \alpha_t, s_t\}$ customers to exit the system. The number of customers carried into period $t+1$ is then:

$$I_{t+1} = [I_t + \alpha_t - s_t]^+ ,$$

where $[x]^+ = \max\{x, 0\}$. We assume s_t is a non-negative random variable that takes only integer values. The draw in each period is identically and independently distributed (IID). $E[s_t] = \mu > \lambda$ and $P(s_t < M) > 0$. Restricting s_t to be integer-valued simplifies the state space by assuring that I_t is always integer-valued. We only need to track the number of agents in the system as opposed to the work in the system. A positive probability that s_t is less than M implies limited capacity. There is some chance that the system cannot process all arrivals. Possible examples of s_t include a Poisson random variable with mean μ or a Bernoulli random variable with probability of success equal to μ. Note that I_t is not a Markov chain since its distribution depends on the entire history of the process.

We assume that the system employs a FIFO discipline. Arrivals in period t must wait for the I_t customers already in the system to exit before any of them may enter service. If $\alpha_t \geq 2$, these new arrivals are randomly ordered and served in that sequence. Delay costs as a function of α_t and I_t thus follow a modified form of (1):

$$W_m(\alpha_t, I_t) = \frac{1}{\alpha_t} \sum_{k=1}^{\alpha_t} C_m(k + I_t) .$$

Let I_{T+1} denote the number of customers who remain unserved at the end of the horizon. We assume the system continues to operate until all customers have been served. Additional service draws $s_{T+1}, s_{T+2}, ...$ are taken until all I_{T+1} customers have exited the system where draw s_{T+k} has the same distribution as s_t for $t = 1, ..., T$. For example, if $s_t = 1$ with certainty, an additional I_{T+1} periods of operation are required. Assuming continued operations removes the end of horizon affect. Together with the FIFO service discipline, it assures that $W_m(\alpha_T, I_T) = W_m(\alpha_t, I_t)$ for $t < T$ as long as $(\alpha_T, I_T) = (\alpha_t, I_t)$.

Let I_1 denote the initial population of customers. To initialize the system, we assume that I_1 is a random variable such that $P(I_1 = k) = f_1(k)$ for $k = 0, 1, ...$ The agents know the distribution of I_1 but do not see its realized value until after they have selected their arrival bin. Allowing I_1 to be a random variable is a generalization over merely assuming the system begins empty since one can always suppose $f_1(0) = 1$.

Agent m's objective is again to minimize her net benefit $U_m(\alpha_t, I_t)$ or equivalently to minimize her expected delay cost $W_m(\alpha_t, I_t)$. Agents still choose their arrival bins simultaneously. The novel aspect of imposing limited capacity is that an agent's cost now depends on both the number of agents that arrive with her as well as the existing inventory of agents. Note that because of the FIFO discipline her costs depends only on the number of agents that arrive before her and with her, not on the number that come after her.

Let Π^* be a candidate equilibrium. $\alpha_t\left(\Pi^*_{-m}\right)$ and $I_t\left(\Pi^*_{-m}\right)$ respectively denote the number of arrivals to bin t and the number of agents already present in bin t under Π^* holding agent m out. Since $I_1\left(\Pi^*_{-m}\right) = I_1$, Π^* depends on the distribution of I_1. We suppress this dependence to simplify the notation. For Π^* to be a Nash equilibrium, we require for $m = 1, ..., M$:

$$V_m - \mathcal{W}_m(\Pi^*) \geq V_m - E\left[W_m\left(\alpha_t\left(\Pi^*_{-m}\right) + 1, I_t\left(\Pi^*_{-m}\right)\right)\right] \quad \text{for } t = 1, ..., T, \tag{7}$$

where

$$\mathcal{W}_m(\Pi^*) = \sum_{t=1}^{T} \pi^t_m E\left[W_m\left(\alpha_t\left(\Pi^*_{-m}\right) + 1, I_t\left(\Pi^*_{-m}\right)\right)\right].$$

Lemma 4 *Given M, T and the distribution of I_1, at least one Nash equilibrium Π^* exists.*

Proof: Every finite strategic form game has a Nash equilibrium. See Theorem 1.1 of Fudenberg and Tirole (1996). ■

The lemma does not guarantee that a pure strategy exists or that Π^U is an equilibrium. It is easy to show that Π^U is not necessarily an equilibrium. Suppose $I_1 \equiv 0$ and that all agents follow Π^U. Then if agent m reports to bin 1, she expects arrivals of $\alpha_t\left(\Pi^U_{-m}\right)$ but no inventory of existing customers. If she reports to bin 2, she again expects arrivals of $\alpha_t\left(\Pi^U_{-m}\right)$

but now $P(I_2 > 0) > 0$ (because $P(s_t < M) > 0$). She consequently strictly prefers the first bin, and Π^U cannot be an equilibrium.

For Π^U to be an equilibrium, we need an alternative initialization. Consider the following:

$$\hat{I}_{t+1} = \left[\hat{I}_t + \hat{\alpha}_t - s_t\right]^+,$$

where s_t is as defined above and $\hat{\alpha}_t$ has a binomial distribution with parameters $M-1$ and $1/T$. Draws of s_t and $\hat{\alpha}_t$ are independent across periods. While the inventory process for our system I_t does not form a Markov chain, $\hat{I}_t$ does. Let $\hat{P}_{ij} = P\left(\hat{I}_{t+1} = j | \hat{I}_t = i\right)$. We assume that $\hat{I}_t$ is ergodic and denote its stationary distribution by $\gamma = (\gamma_0, \gamma_1, ...)$. That is, $\gamma_j = \sum_{i=0}^{\infty} \gamma_i \hat{P}_{ij}$.(See Ross, 1983.) The following theorem shows that if the initialization of the system is given γ, Π^U is an equilibrium.

Theorem 7 *If the distribution of I_1 is such that $f_1(k) = \gamma_k$ for $k = 0, 1, ...$, Π^U is a Nash equilibrium for any set of agent delay cost functions.*

Proof: Given that others play Π^U, agent m perceives the distribution of $\alpha_t\left(\Pi^U_{-m}\right)$ as independent of t. If m deviates from Π^U, it must be because of the evolution of the inventory process $I_t\left(\Pi^U_{-m}\right)$. Suppose the realized value of I_1 equals i. We have

$$I_2\left(\Pi^U_{-m}\right) = \left[i + \alpha_1\left(\Pi^U_{-m}\right) - s_t\right]^+.$$

However, the distribution of $\alpha_1\left(\Pi^U_{-m}\right)$ is a binomial with parameters $M-1$ and $1/T$. Consequently, $P\left(I_2\left(\Pi^U_{-m}\right) = j | I_1 = i\right) = \hat{P}_{ij}$. Unconditioning, one has

$$P\left(I_2\left(\Pi^U_{-m}\right) = j\right) = \sum_{i=0}^{\infty} \gamma_i \hat{P}_{ij} = \gamma_j.$$

Thus the distribution of I_2 is the same as I_1. An induction then extends the result to I_t for $t = 1, ..., T$. We then have $E\left[W_m\left(\alpha_t\left(\Pi^U_{-m}\right) + 1, I_t\left(\Pi^U_{-m}\right)\right)\right]$ is independent of t for any W_m and that Π^U is an equilibrium. ■

It is tempting to interpret Theorem 7 as saying that if the system starts in steady state, the agents play Π^U, but that is not quite correct. First, I_t is not a Markov chain. Second, even if one considers a Markov version of I_t in which arrivals are independent across periods, its transition probabilities would not be $\hat{P}_{ij}$ since its arrivals would have a different distribution than $\hat{\alpha}_t$. That said, as M and T get large, the distribution of $\hat{\alpha}_t$ converges to that of $\alpha_t\left(\Pi^U\right)$, and γ converges to the steady state distribution of I_t. Thus, looking at a limiting system as in Theorem 6, we have that Poisson arrivals see time averages of the state of the system.

A steady-state initialization may seem a limitation to the result, but it is sufficient for our purpose. From the outset, our interest has been whether previous research was limited by assuming renewal arrivals. Most existing research also assumes that arrivals experience steady state waits (e.g., Mendelson, 1985; Mendelson and Whang, 1990). Thus, if one believes that it is sufficient to look at long run average waits, our results suggest that strategic customers will plausibly pick an arrival strategy that approaches a Poisson process.

A second consideration is that strategic customers will take the system to steady state. This is inherent in the definition of an equilibrium as given in (7). If agent m puts a positive probability on multiple bins, it must be the case that she expects the same wait in those bins. Suppose all agents have linear waiting cost, i.e.,

$$W\left(\alpha_t, I_t\right) = \frac{1}{\alpha_t}\sum_{i=1}^{\alpha_t}\left(k - 1 + I_t\right) = \frac{\alpha_t - 1}{2} + I_t,$$

and that for some initialization of I_t (not necessarily γ) there exists an equilibrium Π' such that some agent m puts positive probability on states s and t. It must be:

$$E\left[\frac{\alpha_t\left(\Pi'_{-m}\right) - \alpha_s\left(\Pi'_{-m}\right)}{2}\right] = E\left[I_s\left(\Pi'_{-m}\right) - I_t\left(\Pi'_{-m}\right)\right].$$

Differences in arrival rates compensates for differences in inventory levels, and equilibrium workload levels across bins must be constant. If bin is expected to have a low inventory (as with $I_1 \equiv 0$), it must have a higher arrival rate and vice versa. In general, *any* equilibrium will require that agents expect a (nearly) constant wait across states. If the system does not start in steady state, the agents will choose an equilibrium that moves it to steady state.

4 Conclusion

We have presented a simple timing game. A set of customers seek service over some horizon. All find congestion costly and so try to arrive when the facility is under-utilized. Working in discrete time, we characterize pure strategy equilibria for the case of ample capacity. We show that agents try to spread out their arrivals as much as possible. If all agents have the same well-behaved delay costs, self-interested agents will choose a socially efficient outcome that minimizes total waiting costs.

While potentially efficient, pure strategy equilibria are cumbersome and difficult to implement. We consequently consider mixed strategy equilibria and identify a unique symmetric equilibrium. This equilibrium has several appealing properties: It is independent of both the delay cost functions of the agents and the number of agents. Further as the number of agents and time periods gets large, the number of arrivals in any period goes to a Poisson distribution and the number of arrivals across bins becomes independent. Thus, a large population of strategic customers seeking to avoid congestion generates an arrival pattern that is well approximated by a Poisson process. Our results extend to the case of limited capacity given an appropriate initialization of the system.

Our model lends support to the traditional literature on strategic behavior in queuing systems. This work has generally assumed that customers arrive according to a renewal process but act strategically upon arrival. We show that assuming renewal arrivals is an acceptable assumption given a large population and long horizon.

Our simple setting has limitations. We have worked in a discrete time setting. In many ways, this is an reasonable approximation of human behavior; few people plan their day in, say, five minute increments let alone in continuous time. However, traditional queuing models are in continuous time so it would be worth considering such a setting. This may raise additional technicalities. Considering only a finite number of arrival bins means that we do not have to worry about the existence of an equilibrium (see the proof of Lemma 4). A continuous time formulation may necessitate some additional structure.

We have assumed that all agents have the same horizon and are indifferent to when they are served over that horizon. The latter may be true in some settings (e.g., when to run a simple errand) but not in others (e.g., when to have lunch). Time dependent preferences or overlapping horizons will likely generate non-stationary arrivals. However, an insight gained from our analysis will likely continue to hold: Strategic customers will move the system to steady state by competing away difference in expected waits.

5 References

Dewan, S. and H. Mendelson, 1990, "User Delay Costs and Internal Pricing for a Service Facility," *Management Science*, 36:12, 1502-1517.

Fudenberg, D. and J. Tirole, 1996, *Game Theory*, Cambridge, MA: The MIT Press.

Lippman, S. and S. Stidham, 1977, "Individual versus Social Optimization in Exponential Congestion Systems," *Operations Research*, 25:2, 233-247.

Mendelson, H., 1985, "Pricing Computer Services: Queuing Effects," *Communications of the ACM*, 28:3, 312-321.

Mendelson, H. and S. Whang, 1990, "Optimal Incentive-Compatible Priority Pricing for the M/M/1 Queue," *Operations Research*, 38:5, 870-883.

Naor, P., 1969, "The Regulation of Queue Size by Levying Tolls" *Econometrica*, 37:1, 15-24.

Ostrovsky, M. and M. Schwarz, 2002, "The Adoption of Standards under Uncertainty," working paper, Harvard University.

Ross, S. M., *Stochastic Processes*, New York: John Wiley & Sons, 1983.

Rump, C., and S. Stidham, 1998, "Stability and Chaos in Input Pricing for a Service Facility with Adaptive Customer Response to Congestion," *Management Science*, 44:2, 246-261.

Shaked, M. and J. G. Shanthikumar, 1994, *Stochastic Orders and Their Applications*, San Diego, CA: Academic Press.

Stidham, S., 1992, "Pricing and Capacity Decisions for a Service Facility: stability and Multiple Local Optima," *Management Science*, 38:8, 1121-1139.

Van Mieghem, J., 2000, "Price and Service Discrimination in Queueing Systems: Incentive-Compatibility of G$c\mu$ Scheduling," *Management Science*, 46:9, 1249-1267.

Yechali, U., 1971, "On Optimal Balking Rules and Toll Charges in the GI/M/1 Queuing Process," *Operations Research*, 19:2, 3489-370.

Table 1

(M, T)	*Maximum Deviation* *PMF*	*CDF*
(50, 25)	0.00556	0.00552
(100, 50)	0.00274	0.00273
(200, 100)	0.00136	0.00136
(400, 200)	0.00068	0.00068
(30, 25)	0.00952	0.00673
(60,50)	0.00468	0.00333
(120, 100)	0.00232	0.00165
(240, 200)	0.00116	0.00083

Table 2

Observed Number of Bins	*Deviation from Mean Arrivals*	*Maximum Deviation* *PMF*	*CDF*
50	*-20*	0.01811	0.03616
50	*-10*	0.00960	0.01868
50	*10*	0.00908	0.01810
50	*20*	0.01834	0.03645
100	*-20*	0.02715	0.05423
100	*-10*	0.01417	0.02781
100	*10*	0.01365	0.02718
100	*20*	0.02847	0.05563
150	*-20*	0.05384	0.10802
150	*-10*	0.02741	0.05450
150	*10*	0.02745	0.05450
150	*20*	0.06331	0.11697