STRATEGIC EQUILIBRIUM VERSUS GLOBAL OPTIMUM FOR A PAIR OF COMPETING SERVERS

BENJAMIN AVI-ITZHAK,* *Rutgers University*

BOAZ GOLANY ** *** AND

URIEL G. ROTHBLUM,** **** *Technion, Israel Institute of Technology*

Abstract

Christ and Avi-Itzhak (2002) analyzed a queueing system with two competing servers who determine their service rates so as to optimize their individual utilities. The system is formulated as a two-person game; Christ and Avi-Itzhak proved the existence of a unique Nash equilibrium which is symmetric. In this paper, we explore globally optimal solutions. We prove that the unique Nash equilibrium is generally strictly inferior to a globally optimal solution and that optimal solutions are symmetric and require the servers to adopt service rates that are smaller than those occurring in equilibrium. Furthermore, given a symmetric globally optimal solution, we show how to impose linear penalties on the service rates so that the given optimal solution becomes a unique Nash equilibrium. When service rates are not observable, we show how the same effect is achieved by imposing linear penalties on a corresponding signal.

Keywords: Queueing system; Nash equilibrium; linear reward scheme

2000 Mathematics Subject Classification: Primary 90B22
                    Secondary 68M20; 91A05; 91B52

1. Introduction

Game theory provides tools for studying multi-person systems in which agents' individual outcomes depend on the decisions of others and not only on their own actions. The study of such situations requires consideration of the interests of all the agents that are present in the system; in particular, optimization methodology is not useful as it focuses on a single centralized agent who makes all the decisions. The analysis of game-theoretic perspectives is getting increasing attention in the study of managerial systems, for example in transportation systems [1], queueing systems [6], communications networks [7], and supply chains [2].

One of the most important concepts in the analysis of multi-agent systems is Nash equilibrium. Specifically, a Nash equilibrium is a collection of decisions of the various agents under which the action that each agent selects is optimal for him/her when all other agents do nothing. The extent of the study and use of Nash equilibria in multi-person systems is beyond description; see the aforementioned references for a tiny sample of the exploding

literature. Despite the desirable properties of Nash equilibria, they possess a significant short-coming – because of the focus on individual interests, they are generally not globally optimal. Famous examples of the inefficiency of Nash equilibria include 'the prisoners' dilemma' [10] and 'the tragedy of the commons' [5]. A major challenge to management having global responsibility is to overcome this inefficiency. A natural method of meeting this challenge is the introduction of rewards and penalties that will influence agents' incentives and gear them towards an optimal solution.

Qualitative discussion of the gap between equilibrium and optimal performance in the context of queueing systems started to emerge in the 1960s. Naor [9] was the first to propose a quantitative mechanism that aimed at inducing a globally optimal solution in a queueing system with noncooperative agents, that is, converting the globally optimal solution into a Nash equilibrium. For an extensive study of these issues, see [6]. Recently, Rothblum [11] and Golany and Rothblum [4] introduced a general framework for the use of linear rewards (and penalties) on actions of competing agents under which a given globally optimal solution satisfies first-order conditions of a Nash equlibrium. The 'per-unit-action' rewards they use are determined by the marginal influence of agents' actions on the utilities of the other agents, evaluated at a given optimal solution. In reality, the penalties can be interpreted as a tax which the agents pay to some authority for the right to operate within the system. Indeed, there are real-world systems which follow such a practice. For example, tolls on drivers serve as taxes on their actions of selecting particular roads. Also, in systems where agents bid for products of limited supply (e.g. retailers in supply chains or participants in auctions), a common practice is to require (refundable) deposits. Lost interest for such deposits is a per-unit tax on the agents' actions – their bid size.

The purpose of this paper is to show how linear rewards and penalties can be used to close the gap between Nash equilibrium and global optimality in a queueing system in which servers set their service rates while competing for customers. The specific model we analyze is the one introduced recently by Christ and Avi-Itzhak [3]. It considers a queueing system in which two servers compete over randomly arriving customers; in particular, Christ and Avi-Itzhak generalized a model (the CAI model) introduced in [8] by considering customers' incentives and allowing them to leave the system when the waiting lines are long. Both of these papers focus on proving existence and uniqueness of Nash equilibria which are symmetric; while the inefficiency of these equilibria is observed, its managerial implication is ignored. Our goal is to address this inefficiency and provide management with a tool that will create incentives for the servers so that the (unique) globally optimal solution becomes a Nash equilibrium. In particular, it is shown that optimal performance requires reduction of service rates in comparison to those of the (unique) Nash equilibrium. The linear penalties that are imposed to accomplish this task are set on observable performance measures rather than the explicit service rates that the servers select, which may be unobservable to a central manager. We follow [3] and [8] and focus on two-server systems because of their analytic tractability; we conjecture that the multi-agent systems will facilitate similar analysis to the one we conduct herein.

The specific CAI model we use has customers that arrive according to a Poisson process. Upon arrival, each customer decides whether to join the system's common queue or to leave the system without getting service; the decision is random and the probability that a customer will join the queue is a decreasing function of the queue's length. The servers are to determine their service rates with the goal of improving their expected profits. For each server the revenue is linearly dependent on the number of customers he/she serves and the cost is a convex and increasing function of the service rate. The main result of Christ and Avi-Itzhak [3] is

the existence of a symmetric unique Nash equilibrium in this two-person game and a characterization of this equilibrium.

2. Analysis

We start by analyzing the gap between the equilibrium and the globally optimal solutions for this game. To do so, we use the notation of Christ and Avi-Itzhak [3] as follows. We denote the mean arrival rate of customers by λ. Let μ_j be the service rate for server j, $j = 1, 2$, satisfying $\mu_1 \geq 0$, $\mu_2 \geq 0$, and $\mu_1 + \mu_2 > 0$. We denote the fixed per-customer revenue by R. Let $c(\mu)$ be the cost of server j, $j = 1, 2$, per unit time when serving at rate μ. The function $c(\mu)$ is nonnegative and convex increasing in μ. Let θ_n be the probability that a customer will join a queue whose length is n. It is assumed throughout that $\theta_0 = \theta_1 = 1$, $\theta_i \geq \theta_{i+1}$, $i = 2, 3, \ldots$, and $\lim_{i \to \infty} \theta_i = 0$. Finally, let $\alpha_j(\mu_1, \mu_2)$ be the steady-state proportion of the arriving customers (including those who do not join a queue) served by server j, $j = 1, 2$.

Using the auxiliary variables $\eta = \lambda / (\mu_1 + \mu_2)$, $\beta_i = \prod_{n=0}^{i} \theta_n$, $i = 0, 1, \ldots$, and

$$\sigma = \sum_{i=0}^{\infty} \eta^i \beta_{i+1} \tag{1}$$

(note that $\sigma \geq 1$), Christ and Avi-Itzhak [3] showed that the steady-state proportions are given by

$$\alpha_j(\mu_1, \mu_2) = \frac{\mu_1 \mu_2 + \lambda \mu_j \sigma}{2 \mu_1 \mu_2 + \lambda(\mu_1 + \mu_2) + \lambda^2 \sigma}, \qquad j = 1, 2,$$

and, in addition, they defined $\alpha_j(0, 0) = 0$, $j = 1, 2$.

Now, the individual expected profit function for server j is

$$\Pi_j(\mu_1, \mu_2) = R\lambda \alpha_j(\mu_1, \mu_2) - c(\mu_j), \qquad j = 1, 2,$$

and the global profit function is

$$\Pi(\mu_1, \mu_2) = R\lambda \alpha_1(\mu_1, \mu_2) + R\lambda \alpha_2(\mu_1, \mu_2) - c(\mu_1) - c(\mu_2).$$

Our analysis explores the dependence of the profit function on the selection of the service rates by the two servers with the arrival rate (λ) fixed; hence, it is important to recall that σ depends on these rates and should be written as $\sigma(\mu_1, \mu_2)$. But, to avoid cumbersome notation, we sometimes suppress this dependence and refer simply to σ.

In order to prove our first proposition, we need the following lemma.

Lemma 1. *If $\mu_1, \mu_2 > 0$ then $\eta > (\sigma - 1)/\sigma$.*

Proof. Using (1), $0 < \eta$, $\beta_1 = 1$, $\beta_i \geq \beta_{i+1}$ for all i, and $\lim_{i \to \infty} \beta_i = 0$, we have

$$\frac{\sigma - 1}{\sigma} - \eta = \frac{(\beta_1 + \eta\beta_2 + \eta^2\beta_3 + \cdots) - 1}{\beta_1 + \eta\beta_2 + \eta^2\beta_3 + \cdots} - \eta$$

$$= \eta \left[\frac{\beta_2 + \eta\beta_3 + \cdots}{\beta_1 + \eta\beta_2 + \eta^2\beta_3 + \cdots} - 1 \right]$$

$$= \frac{\sum_{i=0}^{\infty} (\beta_{i+2} - \beta_{i+1}) \eta^i}{\sigma}$$

$$< 0$$

(where we emphasize that the last inequality holds strictly).

We say that a pair (μ_1, μ_2) of service rates is *symmetric* if $\mu_1 = \mu_2$.

Proposition 1. *Every globally optimal solution has symmetric service rates.*

Proof. Consider service rates (μ_1, μ_2) with $\mu_1 \neq \mu_2$ and let $\mu := (\mu_1 + \mu_2)/2$. We will show that $\Pi(\mu, \mu) > \Pi(\mu_1, \mu_2)$, i.e.

$$R\lambda[\alpha_1(\mu, \mu) + \alpha_2(\mu, \mu)] - 2c(\mu) > R\lambda[\alpha_1(\mu_1, \mu_2) + \alpha_2(\mu_1, \mu_2)] - c(\mu_1) - c(\mu_2).$$

First, since the cost function $c(\cdot)$ is convex, $2c(\mu) \leq c(\mu_1) + c(\mu_2)$. Also,

$$\begin{aligned}
&[\alpha_1(\mu, \mu) + \alpha_2(\mu, \mu)] - [\alpha_1(\mu_1, \mu_2) + \alpha_2(\mu_1, \mu_2)] \\
&\quad = \frac{2\mu^2 + 2\lambda\mu\sigma}{2\mu^2 + 2\lambda\mu + \lambda^2\sigma} - \frac{2\mu_1\mu_2 + 2\lambda\mu\sigma}{2\mu_1\mu_2 + 2\lambda\mu + \lambda^2\sigma} \\
&\quad = \frac{2\lambda(\mu^2 - \mu_1\mu_2)(2\mu + \lambda\sigma - 2\sigma\mu)}{(2\mu^2 + 2\lambda\mu + \lambda^2\sigma)^2}.
\end{aligned} \tag{2}$$

Now, in the numerator of (2), the term in the first set of parentheses is positive since $\mu^2 - \mu_1\mu_2 = \frac{1}{4}(\mu_1 - \mu_2)^2$ and the term in the second set of parentheses is positive due to Lemma 1 (noting that $\sigma \geq 1$).

Writing $\bar{\sigma}(\mu) = \sigma(\mu, \mu)$ with derivative $\bar{\sigma}'(\mu) = \mathrm{d}\bar{\sigma}(\mu)/\mathrm{d}\mu$, we have

$$\left. \frac{\partial\sigma(\mu_1, \mu_2)}{\partial\mu_1} \right|_{(\mu_1,\mu_2)=(\mu,\mu)} = \frac{1}{2}\bar{\sigma}'(\mu) = -\sum_{i=1}^{\infty} \frac{i\lambda^i}{(2\mu)^{i+1}}\beta_{i+1} < 0.$$

Christ and Avi-Itzhak proved (see [3, Theorem 1]) that there exists a unique Nash equilibrium solution which is symmetric. We next show that, when positive, the service rates of this solution are strictly larger than the service rates of any globally optimal solution.

Proposition 2. *Whenever the (joint) service rates at the unique Nash equilibrium are positive, they are strictly larger than the (joint) service rates of each globally optimal solution.*

Proof. In [3], Christ and Avi-Itzhak proved that

$$A(\mu) := \left. \frac{\partial\alpha_1(\mu_1, \mu_2)}{\partial\mu_1} \right|_{(\mu_1,\mu_2)=(\mu,\mu)}$$

is strictly decreasing in μ, and that a necessary and sufficient condition for $(\mu, \mu) > (0, 0)$ to be a Nash equilibrium is that it satisfies the first-order condition $R\lambda A(\mu) = c'(\mu)$. Also, a necessary (first-order) condition for $(\mu, \mu) > (0, 0)$ to be a globally optimal solution is $\mathrm{d}\Pi(\mu, \mu)/\mathrm{d}\mu = 0$, i.e. $R\lambda B(\mu) = c'(\mu)$ with

$$B(\mu) := \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\mu}[\alpha_1(\mu, \mu) + \alpha_2(\mu, \mu)].$$

Next, observe that

$$A(\mu) = \frac{\partial \alpha_1(\mu_1, \mu_2)}{\partial \mu_1} \Bigg|_{(\mu_1, \mu_2)=(\mu, \mu)}$$

$$= \frac{1}{[2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)]^2}$$
$$\times \left[[\mu + \lambda\bar{\sigma}(\mu) + \tfrac{1}{2}\lambda\mu\bar{\sigma}'(\mu)][2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)] \right.$$
$$\left. - [\mu^2 + \lambda\mu\bar{\sigma}(\mu)][2\mu + \lambda + \tfrac{1}{2}\lambda^2\bar{\sigma}'(\mu)] \right]$$

and

$$B(\mu) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\mu}[\alpha_1(\mu, \mu) + \alpha_2(\mu, \mu)]$$

$$= \frac{\mathrm{d}}{\mathrm{d}\mu}\left[\frac{\mu^2 + \lambda\mu\bar{\sigma}(\mu)}{2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)} \right]$$

$$= \frac{1}{[2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)]^2}$$
$$\times \left[[2\mu + \lambda\bar{\sigma}(\mu) + \lambda\mu\bar{\sigma}'(\mu)][2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)] \right.$$
$$\left. - [\mu^2 + \lambda\mu\bar{\sigma}(\mu)][4\mu + 2\lambda + \lambda^2\bar{\sigma}'(\mu)] \right].$$

Consequently,

$$B(\mu) - A(\mu) = \frac{\mu[2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)] - [\mu^2 + \lambda\mu\bar{\sigma}(\mu)][2\mu + \lambda]}{[2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)]^2}$$
$$+ \frac{\tfrac{1}{2}\bar{\sigma}'(\mu)\{\lambda\mu[2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)] - [\mu^2 + \lambda\mu\bar{\sigma}(\mu)]\lambda^2\}}{[2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)]^2}$$
$$= \frac{\lambda\mu^2[1 - 2\bar{\sigma}(\mu)] + \tfrac{1}{2}\bar{\sigma}'(\mu)\lambda\mu^2(2\mu + \lambda)}{[2\mu^2 + 2\lambda\mu + \lambda^2\bar{\sigma}(\mu)]^2}$$
$$< 0.$$

The last inequality follows from the fact $\beta_1 = \theta_1\theta_0 = 1$, which implies that $\sigma = \sum_{i=0}^{\infty} \eta^i \beta_{i+1} \geq \beta_1 = 1$, and from $\bar{\sigma}'(\mu) < 0$. Now, let (μ, μ) be a globally optimal solution with $\mu > 0$; in particular, $R\lambda B(\mu) = c'(\mu)$. As $R\lambda A(\mu) > R\lambda B(\mu) = c'(\mu)$, because $A(\mu)$ is strictly decreasing and $c'(\mu)$ is nondecreasing ($c(\mu)$ is convex), we have, for all $\hat{\mu} \leq \mu$,

$$R\lambda A(\hat{\mu}) \geq R\lambda A(\mu) > R\lambda B(\mu) = c'(\mu) \geq c'(\hat{\mu}),$$

ensuring that the unique μ^* satisfying $R\lambda A(\mu^*) = c'(\mu^*)$ has $\mu < \mu^*$. This completes our proof (as the verification of the asserted bound when $\mu = 0$ is trivial).

A unified framework was presented in [4] for using linear rewards and penalties to convert a given globally optimal solution into a Nash equilibrium. In particular, the coefficients of the corresponding linear expressions are the marginal contributions of each agent's actions to the total welfare (expected utility) of the other agents, evaluated at the given optimal solution. Results in [11] show that first-order conditions for optimality imply the Nash equilibrium first-order conditions and, under restrictive conditions, the globally optimal solution becomes a

Nash equilibrium. Broader applicability of the approach was demonstrated in [4], where it was further shown that, for specific models, the resulting Nash equilibrium is unique. We next demonstrate that these linear rewards (in fact, penalties) can be used to convert any given globally optimal solution of the systems considered herein into a *unique* Nash equilibrium.

Proposition 3. *Let $(\hat{\mu}, \hat{\mu}) > (0, 0)$ be a globally optimal solution. Then $(\hat{\mu}, \hat{\mu})$ is a unique Nash equilibrium of the system in which the cost function $c(\mu)$ is modified by adding the linear penalty $-K\mu$, where*

$$K := R\lambda \frac{\lambda \hat{\mu}^2 [1 - 2\bar{\sigma}(\hat{\mu})] + \frac{1}{2}\bar{\sigma}'(\hat{\mu})\lambda\hat{\mu}^2(2\hat{\mu} + \lambda)}{[2\hat{\mu}^2 + 2\lambda\hat{\mu} + \lambda^2\bar{\sigma}(\hat{\mu})]^2}; \tag{3}$$

furthermore,

$$K = \left. \frac{\partial \Pi_1(\mu_1, \mu_2)}{\mu_2} \right|_{(\mu_1, \mu_2) = (\hat{\mu}, \hat{\mu})} < 0.$$

Proof. From Proposition 2, we see that the K selected in (3) satisfies $K = R\lambda[B(\hat{\mu}) - A(\hat{\mu})] < 0$. Adding any linear term $L\mu$ to the cost function $c(\mu)$ preserves its convexity and at the same time does not change the evolution of the queueing process for any prescribed service rates. Consequently, [3, Theorem 1] ensures the existence of a unique Nash equilibrium which is symmetric and is characterized by the (modified) first-order conditions $R\lambda A(\mu) = c'(\mu) + L$.

As $(\hat{\mu}, \hat{\mu}) > (0, 0)$ is optimal, it satisfies the first-order condition $R\lambda B(\hat{\mu}) = c'(\hat{\mu})$, ensuring that $R\lambda A(\hat{\mu}) = c'(\hat{\mu}) - R\lambda[B(\hat{\mu}) - A(\hat{\mu})] = c'(\hat{\mu}) - K$. So, $\hat{\mu}$ satisfies the Nash equilibrium first-order conditions for the system in which a linear penalty $-K\mu$ is added to the cost function $c(\mu)$, implying that $\hat{\mu}$ is the unique Nash equilibrium for this system.

To complete the proof, recall that

$$\left. \frac{\partial \sigma(\mu_1, \mu_2)}{\partial \mu_1} \right|_{(\mu_1, \mu_2) = (\mu, \mu)} = \frac{1}{2}\bar{\sigma}'(\mu)$$

and observe that

$$\begin{aligned}
\left. \frac{\partial \Pi_1}{\partial \mu_2} \right|_{(\mu_1, \mu_2) = (\hat{\mu}, \hat{\mu})} &= R\lambda \left. \frac{\partial \alpha_1}{\partial \mu_2} \right|_{(\mu_1, \mu_2) = (\hat{\mu}, \hat{\mu})} \\
&= R\lambda \frac{1}{[2\hat{\mu}^2 + 2\lambda\hat{\mu} + \lambda^2\bar{\sigma}(\hat{\mu})]^2} \\
&\quad \times \Big[\big[\hat{\mu} + \tfrac{1}{2}\lambda\hat{\mu}\bar{\sigma}'(\hat{\mu})\big][2\hat{\mu}^2 + 2\lambda\hat{\mu} + \lambda^2\bar{\sigma}(\hat{\mu})] \\
&\qquad - [\hat{\mu}^2 + \lambda\hat{\mu}\bar{\sigma}(\hat{\mu})]\big[2\hat{\mu} + \lambda + \tfrac{1}{2}\lambda^2\bar{\sigma}'(\hat{\mu})\big]\Big] \\
&= R\lambda \frac{\lambda\hat{\mu}^2(1 - 2\bar{\sigma}) + \tfrac{1}{2}\bar{\sigma}'(\hat{\mu})\lambda\hat{\mu}^2(2\hat{\mu} + \lambda)}{[2\hat{\mu}^2 + 2\lambda\hat{\mu} + \lambda^2\bar{\sigma}(\hat{\mu})]^2}.
\end{aligned}$$

An important issue in imposing linear rewards/penalties is the observability of the agents' actions. When actions are observed only by the agents who select them, it is still possible to achieve the desired effect by imposing linear rewards/penalties on observable random signals whose expected values equal, respectively, the agents' actions. The expected value of the corresponding rewards/penalties will then coincide with linear rewards/penalties that are imposed on the agents' actions, and the agents will respond to the two mechanisms in the same way. In our queueing system, the agents' actions are the service rates, which may, or may not, be

observed by an external entity. An observable (random) signal whose expectation is equal to the service rate of a server is the number of service completions per unit time during the periods in which the server is busy.

3. Multi-server systems

We conjecture that results similar to those which we have obtained for two-server systems also hold for multi-server systems. However, the more general systems are difficult to analyze. Specifically, determining the proportion of customers that are served by each server as a function of the service rates requires the solution of a system of 2^r nonlinear equations, where r is the number of servers. The methods that were used so far to solve the system with $r = 2$ fail for larger r, and solving the resulting systems is an open problem. In fact, it is not even known if the simplest multi-server queueing systems have a symmetric Nash equilibrium.

The difficulty of explicitly determining the queue length distribution and the split of customers among servers diminishes in situations where the traffic is heavy, i.e. the arrival rate is high in comparison to the cumulative service rate. (We credit an anonymous referee for suggesting that we consider multi-server systems in which a heavy traffic approximation applies and that, in such cases, the analysis becomes trivial.) Specifically, consider a system with $r > 2$ servers where the individual service rates that can be selected are bounded by some $\mu^\diamond > 0$. This can be the case because of resource limitations or because $\lim_{t \to \infty} c'(\mu) > R$, in which case servers have no incentive to select a service rate

$$\mu > \mu^\diamond := \min\{t : c'(t) \le R\}.$$

Now, if the arrival rate λ is much larger than $r\mu^\diamond$, then the servers will be busy almost continuously, and the expected individual profit function $\Pi_j(\mu_1, \dots, \mu_r)$ of player j can be approximated by $R\mu_j - c(\mu_j)$. But, when the profit function of each server depends only on his/her selected action and not on the actions selected by the others, every globally optimal solution is a Nash equilibrium; in fact, selecting the service rate

$$\hat{\mu} = \min\{t : c'(t) \le R\}$$

(which maximizes the individual expected profit function) will result in a Nash equilibrium that is globally optimal.

4. Summary

When each server determines its service rate so as to maximize its individual profit, both end up selecting values that are larger than those that will bring the performance of the system as a whole to a global optimum. This is a consequence of the convexity of the cost functions. When both servers implement the reduced service rates prescribed by the globally optimal solution, the system loses some revenue as the queue length increases and some customers leave. But, this loss is more than compensated by the savings in the cost of maintaining the smaller service rates. Since both servers are locked in an equilibrium solution, neither of them will make the first move out of it. By introducing a linear penalty simultaneously on both servers, we induce them to switch from the equilibrium service rates to a new equilibrium point, which corresponds to globally optimal performance. A promising line of future research is to explore dynamic rewards/penalties; these are imposed periodically and the servers respond by setting service rates, taking into account the service rate set by their counterparts in the previous period. The

policy of setting the rewards and penalties will be aimed at providing individual incentives that induce trajectories that approach global optima.

References

[1] BECKMANN, M., MCGUIRE, C. B. AND WINSTEN, C. B. (1956). *Studies in the Economics of Transportation*. Yale University Press.

[2] CACHON, G. P. AND NETESSINE, S. (2004). Game theoretic applications in supply chain analysis. In *Handbook of Quantitative Supply Chain Analysis: Modeling in the eBusiness Era*, eds S. D. Wu, D. Simchi-Levi and Z. Shen, Kluwer, Dordrecht, pp. 13–66.

[3] CHRIST, D. AND AVI-ITZHAK, B. (2002). Strategic equilibrium for a pair of competing servers with convex cost and balking. *Manag. Sci.* **48,** 813–820.

[4] GOLANY, B. AND ROTHBLUM, U. G. (2006). Inducing coordination in supply chains through linear reward schemes. *Naval Res. Logistics* **53,** 1–15.

[5] HARDIN, G. (1968). The tragedy of the commons. *Science* **162,** 1243–1248.

[6] HASSIN, R. AND HAVIV, M. (2003). *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*. Kluwer, Norwell, MA.

[7] JOHARI, R. AND TSITSIKLIS, J. N. (2004). Efficiency loss in a network resource allocation game. *Math. Operat. Res.* **29,** 407–435.

[8] KALAI, E., KAMIEN, M. I. AND RUBINOVITZ, M. (1992). Optimal service speeds in a competitive environment. *Manag. Sci.* **38,** 1154–1163.

[9] NAOR, P. (1969). The regulation of queue size by levying tolls. *Econometrica* **37,** 15–24.

[10] RAPOPORT, A. AND CHAMMAH, A. M. (1965). *Prisoner's Dilemma*. University of Michigan Press.

[11] ROTHBLUM, U. G. (2005). Optimality vs. equilibrium: inducing stability by linear rewards and penalties. Unpublished manuscript.