

Incentives and Pricing in Communications Networks

Asuman Ozdaglar and R. Srikant

Abstract

In this chapter, we study two types of pricing mechanisms: one where the goal of the pricing scheme is to achieve some socially beneficial objective for the network and the other where prices are set by multiple competing service providers to maximize their revenues. For both cases, we present an overview of the mathematical models involved and the relevant optimization and game-theoretic techniques needed to study these models. We study the impact of different degrees of strategic interactions among users and between users and service providers on the network performance. We also relate our models and solutions to practical resource allocation mechanisms used in communication networks such as congestion control, routing, and scheduling. We conclude the chapter with a brief introduction to other game-theoretic topics in emerging networks.

This chapter studies the problem of decentralized resource allocation among competing users in communication networks. The growth in the scale of communication networks and the newly emerging interactions between administrative domains and end users with different needs and quality of service requirements necessitate new approaches to the modeling and control of communication networks that recognize the difficulty of formulating and implementing centralized control protocols for resource allocation. The current research in this area has developed a range of such approaches. Central to most of these approaches is the modeling of end users and sometimes also of service providers as self-interested agents that make decentralized and selfish decisions. This research has two important implications:

- (i) The modeling of communication networks consisting of multiple selfish agents requires tools from game theory.
- (ii) In the absence of centralized control, the interaction of multiple selfish agents may lead to suboptimal resource allocation.

This chapter will survey and develop existing work focusing on the role of prices, both used as control parameters in the network and set by service providers to increase their revenues. We will identify the different roles that prices may play in

communication networks depending on the degree of strategic interactions among users and between users and service providers, and explore their impact on network performance under different scenarios. We will also highlight how the study of large-scale communication networks raises new modeling challenges and develop the mathematical tools that are commonly used in this analysis.

The chapter is organized into three sections: the first two sections correspond to two conceptually different strategic settings, one where pricing is used to achieve some socially beneficial objective, and the other where prices are set by multiple service providers to maximize their revenues. The last section places the material in this chapter in the context of the broader literature, discusses some emerging applications of game theory to communication networks, and suggests a number of areas for future research.

22.1 Large Networks – Competitive Models

In this section, we present a brief overview of the literature on pricing to maximize system utility in a network with a large number of users. This line of research has had a tremendous impact on communication networks, having contributed both to a deeper understanding of network architectures and to the development of new protocols for more efficient use of resources in the Internet. We will end the section with some extensions to wireless networks.

Consider a large network shared by many users, where the goal is to share the network resources in an optimal manner. It may be useful to think of the network as a graph with nodes and links. Each end user in the network is interested in transferring data between a source node and a destination node along a fixed route (or connection). We will use the terms “user,” “source,” and “connection” interchangeably. The nodes are interconnected by links. The network resources that we consider here are the link bandwidths. The bandwidth of a link is the maximum rate at which it can transmit data between the two nodes at either end of the link. We associate a utility function with each user in the network, and we will refer to a resource allocation scheme as being *socially optimal* if it maximizes the sum of utilities of all users in the network.¹

A network is modeled as a set of resources indexed by l , called links, with finite capacities c_l . It is shared by a set of sources, indexed by r . Let $U_r(x_r)$ be the utility of source r as a function of its rate x_r (measured in packets per unit time). The utility function U_r is assumed to be a strictly increasing, strictly concave function. Associated with each source is a route that is a collection of links in the network. Let R be a *routing matrix* whose (l, r) entry is 1 if source r 's route includes link l and is 0 otherwise. Since there is a one-to-one mapping between users and routes, we will use the same index to denote both a user and its route. For example, an index r can represent both user r and its route. Thus, the notation $l \in r$ indicates that link l is in the route of user r .

¹ In the networking literature, social optimality and fairness are often used interchangeably. For other notions of fairness, see Cho and Goel (2006).

The resource allocation problem can be formulated as the following nonlinear optimization problem (Kelly, 1997):

$$\max_{x \geq 0} \sum_r U_r(x_r), \quad Rx \leq c, \quad (22.1)$$

where x is the vector of source rates and c is the vector of link capacities. The constraint says that, at each link l , the aggregate source rate $\sum_r R_{lr} x_r$ does not exceed the capacity c_l . If the utility functions are strictly concave, then the above optimization problem has a unique optimal solution, which we refer to as the socially optimal allocation.

To solve this problem directly, we have to know the utility functions and routes of all the sources in the network. In a large network such as the Internet, this information is not available centrally. One solution to this problem is to devise a mechanism such as the celebrated Vickrey–Clarke–Groves (VCG) mechanism to encourage users to reveal their utilities truthfully (see Chapters 5 and 9). However, such a mechanism is computationally complex to implement and would also require a central authority to solve an optimization problem to compute the prices. Instead, Kelly devised a simple mechanism capable of achieving the optimal allocation of resources in the presence of selfish users (see also Chapter 21). We will describe this scheme in the rest of this section and also show how the pricing motivation also leads to protocols for managing the Internet. Such a scheme was originally proposed in Kelly (1997), Kelly et al. (1998) and variations have been considered in Low and Lapsley (1999), Yaiche et al. (2000), and Kunniyur and Srikant (2002); for a more exhaustive survey of the work in this area, see Srikant (2004).

Given the convexity of (22.1), a vector of rates \hat{x} is optimal if there exists a vector of Lagrange multipliers \hat{p} satisfying the following Karush–Kuhn–Tucker (KKT) conditions:

$$U'_r(\hat{x}_r) = \sum_{l:l \in r} \hat{p}_l, \quad \forall r, \quad (22.2)$$

$$\hat{p}_l \left(\sum_{r:l \in r} \hat{x}_r - c_l \right) = 0, \quad \forall l, \quad (22.3)$$

$$\sum_{r:l \in r} \hat{x}_r \leq c_l, \quad \forall l, \quad (22.4)$$

$$\hat{p}, \hat{x} \geq 0. \quad (22.5)$$

Now, suppose that the network can compute \hat{p} and charges each user r a price per bit of \hat{q}_r where \hat{q}_r is given by

$$\hat{q}_r = \sum_{l:l \in r} \hat{p}_l. \quad (22.6)$$

In vector form, the above relationship can be written as $\hat{q} = R^T \hat{p}$.

If the contribution of each user's flow to the aggregate is negligible, we expect them to take aggregate quantities, in particular prices, as given in their decisions. In this case, we refer to the users as *price takers*. Under this assumption, user r 's optimization

problem can be expressed as

$$\max_{x_r \geq 0} U_r(x_r) - \hat{q}_r x_r. \quad (22.7)$$

This expression is intuitive since it implies that each user is maximizing his utility minus the marginal cost of his flow, which consists of the sum of the Lagrange multiplier of each link traversed on its route. Clearly the solution to this problem is given by \hat{x}_r in (22.2). The equilibrium under this pricing scheme where each user is charged the sum of the Lagrange multipliers on its path coincides with the socially optimum outcome. There are two key assumptions for this implication: (1) Users are price takers, which is reasonable in the case of a large network such as the Internet and (2) prices are set equal to the Lagrange multipliers to implement the socially optimal allocation. This assumption is reasonable when prices are set by a network controller interested in the overall performance. We will discuss how the situation is different when prices are set by profit-maximizing service providers in the next section.

For the above pricing scheme to work, the network has to be able to compute the Lagrange multipliers. There are two problems associated with this computation:

P1 The network does not know the utility functions of the users.

P2 Even if all the utility functions are known, there is no central authority that knows all the link capacities and the network topology to be able to solve (22.2)–(22.5).

To address (P1)–(P2), we consider the following two-step mechanism. First, each user r announces a bid w_r , which is the price per unit time that it is willing to pay. Then, the network decides to allocate rates to users according to the solution of the following optimization problem:

$$\max_{x \geq 0} \sum_r w_r \log(x_r), \quad Rx \leq c. \quad (22.8)$$

The solution to the above optimization problem is called a weighted proportionally fair rate allocation. The KKT conditions for the optimization problem (22.8) are given by

$$\frac{w_r}{x_r^*} = \sum_{r:l \in r} p_l^*, \quad \forall r, \quad (22.9)$$

$$p_l^* \left(\sum_{r:l \in r} x_r^* - c_l \right) = 0, \quad \forall l, \quad (22.10)$$

$$\sum_{r:l \in r} x_r^* \leq c_l, \quad \forall l, \quad (22.11)$$

$$p^*, x^* \geq 0, \quad (22.12)$$

where x^* is the solution to (22.8) and p^* is the associated vector of Lagrange multipliers. Furthermore, if the user can be induced to select $w_r = x_r^* U'_r(x_r^*)$, then $x^* = \hat{x}$ and the network problem coincides with the social welfare maximization problem.

To implement the mechanism described above, we have to first design a distributed algorithm to solve (22.8). The algorithm that we design is a dynamic algorithm where each link computes a price as a function of time according to a differential equation. The differential equation is designed so that, in steady state, the price of each link converges

to the Lagrange multiplier corresponding to the link’s resource constraint. To this end, suppose that each link computes a price according to the differential equation

$$\dot{p}_l = (y_l - c_l)_{p_l}^+, \tag{22.13}$$

where $p_l(t)$ is the instantaneous link price at time t , $y_l = \sum_{r:l \in r} x_r$ is the total arrival rate at link l , and $(a)_b^+$ is equal to $\max(a, 0)$ when $b = 0$ and is equal to a if $b > 0$. Note that the equilibrium of this differential equation is either $y_l = c_l$ or $p_l = 0$ which satisfy one of the KKT conditions (22.10). Each user’s computer is hardwired with a program that computes rates according to the equation

$$x_r = \frac{w_r}{q_r}, \tag{22.14}$$

where q_r is the price of route r and is given by $q_r = \sum_{l:l \in r} p_l$.

To implement the above set of equations, it is assumed that the user r ’s computer is equipped with a protocol to collect q_r , the price of its path, from the network. In networking parlance, equation (22.14) is called a congestion control algorithm since the user reacts to congestion indication in the form of q_r . It is easy to see that if equations (22.13)–(22.14) converge, then their steady-state values satisfy (22.9)–(22.12) and thus, solve the optimization problem (22.8). Indeed the above set of equations converge under some mild assumptions. Let us suppose that the routing matrix R has full row rank, i.e., given a vector q of route prices, the vector of link prices p is uniquely determined by the equation $q = R^T p$. Since x^* is unique, this assumption ensures that p^* is unique. The following identity is useful:

$$q^T x = p^T R x = p^T y.$$

Now, consider the Lyapunov function

$$V(p) = \frac{1}{2}(p - p^*)^T(p - p^*).$$

Differentiating the Lyapunov function, we get

$$\begin{aligned} \frac{dV}{dt} &= \sum_l (p_l - p_l^*)(y_l - c_l)_{p_l}^+ \\ &\stackrel{(a)}{\leq} \sum_l (p_l - p_l^*)(y_l - c_l) \\ &\leq (p - p^*)^T (y - c) \\ &= (p - p^*)^T (y - y^*) + (p - p^*)^T (y^* - c) \\ &\stackrel{(b)}{\leq} (p - p^*)^T (y - y^*) \\ &= (p - p^*)^T R(x - x^*) = (q - q^*)^T (x - x^*) \\ &= \sum_r \left(\frac{w_r}{x_r} - \frac{w_r}{x_r^*} \right) (x_r - x_r^*) \\ &\stackrel{(c)}{\leq} 0, \end{aligned}$$

where (a) follows from the fact if the projection $(\cdot)_{p_l}^+$ is not active, then the inequality holds as an equality and if the projection is active, the right-hand side of (a) is positive while the right-hand side of the equation above (a) is zero. Inequality (b) follows from the fact that either $y_l^* = c_l$ or $y_l^* < c_l$ and $p_l^* = 0$. Finally, inequality (c) follows from the fact that $1/x_r$ is a decreasing function. Thus, for a fixed set of bids $\{w_r\}$, the system of equations (22.13)–(22.14) converges to the point (x^*, p^*) .

The above Lyapunov argument indicates that the congestion control algorithm is stable if w_r is fixed. However, since the price that a user pays is a function of its bid w_r , it is in the interest of the user to vary w_r . How might the user vary w_r ? In general, we may expect users to act strategically and take into account the impact of their current bid on the future prices they will face. However, for our purposes here, let us suppose that they ignore these strategic aspects and behave myopically. In this case, they will simply maximize instantaneous net utility, the user's optimization problem to choose w_r is given by

$$\max_{w_r} U_r \left(\frac{w_r}{q_r} \right) - w_r.$$

Thus, the user chooses w_r to satisfy

$$U_r' \left(\frac{w_r}{q_r} \right) = q_r,$$

or equivalently as

$$w_r = x_r U_r'(x_r).$$

The congestion control algorithm then becomes

$$U_r'(x_r) = q_r. \quad (22.15)$$

The equilibrium point of the differential equation (22.13) is then given by (22.9)–(22.12) with w_r replaced by $x_r^* U_r'(x^*)$. In this case, the $x^* = \hat{x}$ where we recall that \hat{x} is the optimal solution of (22.1) and satisfies (22.2)–(22.5). Thus, if the user is price-taking and myopic, then the users' selfish objectives coincide with the social welfare objective of the system. To prove the convergence of (22.13)–(22.15), one can use the same Lyapunov function $V(p)$ as before and proceed along the same lines.

An interesting side benefit of the pricing scheme above is that it provides a natural decomposition of the network functionalities that is useful in designing the architecture of a communication network. The pricing model suggests that the resource allocation functionality should be decomposed into pieces implemented in different parts of the network:

- (i) *Congestion control at the end users:* The end users should be equipped with a protocol to adapt their rates in response to congestion feedback (route price) from the network.
- (ii) *Congestion indication at the routers:* The routers (the nodes in the graph) in the network should be equipped with a protocol to compute the price of each link that originates from the router. The price is an indicator of congestion on the link.

(iii) *Congestion feedback from the network to the users:* There must be a protocol that allows an end user to collect congestion information from the network. For example, each data packet could contain a field to collect the congestion information. This congestion field could be set to zero at the source and each router on the path can add its price to this field. When the data packet reaches the destination, the congestion field will contain the price of the route. The destination can then send a packet to the source to convey the route price information.

The pricing framework introduced in this section can also be extended to incorporate other functionalities such as scheduling in a wireless network. We will briefly illustrate the extension to wireless networks, using a simple model; for a more general treatment, please see the survey (Lin et al., 2006) and the references within.

In a wireline network, packets can be transferred on all links simultaneously. However, in a wireless network, due to interference and collision, if a packet is scheduled on a link, other links in a neighborhood should be silent to avoid collisions and the resulting packet loss. We refer to a set of links that can be scheduled simultaneously as a *schedule*. Let M_1, M_2, \dots, M_n be the set of possible schedules in a network. Let f_i be the fraction of time that the network uses schedule M_i . The resource constraints in the network can now be expressed as

$$\sum_{r:l \in r} x_r \leq \sum_{i:l \in M_i} f_i c_l, \tag{22.16}$$

$$\sum_{i=1}^n f_i \leq 1, \tag{22.17}$$

$$f, x \geq 0, \tag{22.18}$$

where c_l is the number of packets that can be served by link l if it is scheduled. The goal is to find $\{x_r\}$ and $\{f_i\}$ to maximize $\sum_r U_r(x_r)$. The dual of the problem of maximizing $\sum_r U_r(x_r)$ subject to the constraints (22.16)–(22.18) is

$$\max_{p, \lambda \geq 0} D(p, \lambda),$$

where

$$\begin{aligned} D(p, \lambda) &= \max_{x \geq 0, f \geq 0} \sum_r U_r(x_r) - \sum_l p_l \left(\sum_{r:l \in r} x_r - \sum_{i:l \in M_i} f_i c_l \right) \\ &\quad - \lambda \left(\sum_{i=1}^n f_i - 1 \right) \\ &= \max_{x \geq 0} \sum_r U_r(x_r) - \sum_l p_l \sum_{r:l \in r} x_r \end{aligned} \tag{22.19}$$

$$+ \max_{f \geq 0} \sum_l p_l \sum_{i:l \in M_i} f_i c_l - \lambda \left(\sum_{i=1}^n f_i - 1 \right). \tag{22.20}$$

It is not difficult to see that the dual objective for the wireline problem would also contain the term (22.19), while (22.20) is unique to the wireless problem. This suggests that the algorithm to compute x and p would be quite similar to the wireline case, but

additional computation is necessary to find the optimal value of f . Without using the Lagrange multiplier λ , note that (22.20) can be equivalently written as

$$\max_{\sum_{i=1}^n f_i \leq 1, f_i \geq 0} \sum_l p_l \sum_{i:l \in M_i} f_i c_l = \max_{\sum_{i=1}^n f_i \leq 1, f_i \geq 0} \sum_i f_i \sum_{l \in M_i} p_l c_l = \max_i \sum_{l \in M_i} p_l c_l,$$

where the first equality is a simple interchange of the sums and the second equality follows from the fact that the optimization is a linear program and hence the solution will occur at a corner point. The last maximization problem can be interpreted as follows: pick the schedule that has the largest weighted price where the weights are the link capacities. The update equation at the source remains the same as before and is given by (22.15). It should be noted that while the network picks one of the schedules M_1, M_2, \dots, M_n to solve (22.15) at each time instant, it turns out that the long-run fraction of time that each schedule is the optimal solution to the utility maximization problem; the interested reader is referred to Lin et al. (2006) and references within.

The price updates at the links are given by

$$\dot{p}_l = \left(y_l - \sum_{i:l \in M_i} f_i c_l \right)_{p_l}^+. \quad (22.21)$$

Note that the above equation does not have to be explicitly implemented; it is simply the queue length at link l , which will be automatically maintained by each link. Thus, the only additional implementation required in a wireless network is the computation of the maximum weighted price schedule. This is a computationally hard problem and, in practice, also requires a distributed implementation to be feasible. The problem of low complexity, distributed algorithms to approximate the maximum weighted price schedule is currently open. Assuming that such an algorithm exists, the stability of equations (22.15)–(22.21) can be established using a Lyapunov function approach similar to the wireline case.

22.2 Pricing and Resource Allocation – Game Theoretic Models

The previous section explored how prices can be used as control parameters for allocating resources in communication networks. The analysis was non-game theoretic since users were assumed to be price takers and prices were set as control parameters to achieve the socially optimal allocation. While the framework with prices as control parameters is a useful starting point, it ignores a number of issues that are important for the analysis of resource allocation in large-scale communication networks. First, in a number of settings, where centralized control signals may be impractical or impossible, end users may not face explicit prices. It is therefore important to understand the implications of selfish end-user behavior when the congestion they create and their use of scarce resources are not priced. Second, prices are often set by multiple service providers in control of their administrative domains with the objective of maximizing their (long-run) revenues. In this section, we investigate the implications of

profit-maximizing pricing by multiple decentralized service providers. We turn to a discussion of other possible generalizations in the next section.

22.2.1 Pricing and Efficiency with Congestion Externalities

We now construct a model of resource allocation in a network with competing self-ish users and profit-maximizing service providers. The central question is whether the equilibrium prices that emerge in such a framework will approximate the prices implementing the socially optimal allocation discussed in the previous section. The class of models incorporating strategic behavior by service providers introduces new modeling and mathematical challenges. These models translate into game-theoretic competition models with negative congestion externalities,² whereby the pricing decision of a service provider affects the level of traffic and thus the extent of congestion in other parts of the network. Nevertheless, tractable analysis of pricing decisions and routing patterns are possible under many network topologies.

Models incorporating for-profit service providers have been previously investigated in Basar and Srikant (2002a, 2002b) and Acemoglu and Ozdaglar (2004). Here, we develop a general framework for the analysis of price competition among providers in a congested (and potentially capacitated) network building on Acemoglu and Ozdaglar (2006a, 2006b). We will see that despite its conceptual simplicity, this framework has rich implications. We illustrate some of these, for example, by showing the counterintuitive result that increasing competition among providers can reduce efficiency, which is different from the results of the most common models of competition in economics. Most importantly, we also show that it is possible to quantify the extent to which prices set by competing service providers approximate control role of prices discussed in the previous section. While generally service provider competition does not lead to an equilibrium replicating the system optimum, the extent of inefficiency resulting from price competition among service providers can often be bounded.

We start with a simple example that shows the efficiency implications of competition between two for-profit service providers.

Example 22.1 One unit of traffic will travel from an origin to a destination using either route 1 or route 2 (cf. Figure 22.1). The latency functions of the links, which represent the delay costs as a function of the total link flow, are given by

$$l_1(x) = \frac{x^2}{3}, \quad l_2(x) = \frac{2}{3}x.$$

It is straightforward to see that the efficient allocation [i.e., one that minimizes the total delay cost $\sum_i l_i(x_i)x_i$] is $x_1^S = 2/3$ and $x_2^S = 1/3$, while the (Wardrop) equilibrium allocation that equates delay on the two paths is $x_1^{WE} \approx .73 > x_1^S$ and $x_2^{WE} \approx .27 < x_2^S$. The source of the inefficiency is that each unit of traffic does not internalize the greater increase in delay from travel on route 1, so there is too much use of this route relative to the efficient allocation.

² An externality arises when the actions of the player in a game affects the payoff of other players.

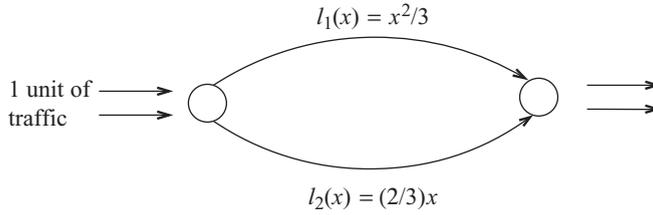


Figure 22.1. A two link network with congestion-dependent latency functions.

Now consider a monopolist controlling both routes and setting prices for travel to maximize its profits. We show below that in this case, the monopolist will set a price including a markup, which exactly internalizes the congestion externality. In other words, this markup is equivalent to the Pigovian tax that a social planner would set in order to induce decentralized traffic to choose the efficient allocation. Consequently, in this simple example, monopoly prices will be $p_1^{\text{ME}} = (2/3)^3 + k$ and $p_2^{\text{ME}} = (2/3)^2 + k$, for some constant k . The resulting traffic in the Wardrop equilibrium will be identical to the efficient allocation, i.e., $x_1^{\text{ME}} = 2/3$ and $x_2^{\text{ME}} = 1/3$.

Finally, consider a duopoly situation, where each route is controlled by a different profit-maximizing provider. In this case, it can be shown that equilibrium prices will take the form $p_i^{\text{OE}} = x_i^{\text{OE}}(l'_1 + l'_2)$ [see Eq. (22.27) in Section 22.2.4], or more specifically, $p_1^{\text{OE}} \approx 0.61$ and $p_2^{\text{OE}} \approx 0.44$. The resulting equilibrium traffic is $x_1^{\text{OE}} \approx .58 < x_1^{\text{S}}$ and $x_2^{\text{OE}} \approx .42 > x_2^{\text{S}}$, which also differs from the efficient allocation. It is noteworthy that although the duopoly equilibrium is inefficient relative to the monopoly equilibrium, in the monopoly equilibrium k is chosen such that all of the consumer surplus is captured by the monopolist, while in the oligopoly equilibrium users may have positive consumer surplus.³

The intuition for the inefficiency of the duopoly relative to the monopoly is related to a new source of (differential) monopoly power for each duopolist, which they exploit by distorting the pattern of traffic: when provider 1, controlling route 1, charges a higher price, it realizes that this will push some traffic from route 1 to route 2, raising congestion on route 2. But this makes the traffic using route 1 become more “locked-in,” because their outside option, travel on route 2, has become worse. As a result, the optimal price that each duopolist charges will include an additional markup over the Pigovian markup. Since the two markups are generally different, they will distort the pattern of traffic away from the efficient allocation.

22.2.2 Model

We consider a network with I parallel links. Let $\mathcal{I} = \{1, \dots, I\}$ denote the set of links. Let x_i denote the total flow on link i , and $x = [x_1, \dots, x_I]$ denote the vector of link

³ Consumer surplus is the difference between users’ willingness to pay (reservation price) and effective costs, $p_i + l_i(x_i)$, and is thus different from the social surplus (which is the difference between users’ willingness to pay and latency cost, $l_i(x_i)$), thus also takes into account producer surplus/profits.

flows. Each link in the network has a flow-dependent latency function $l_i(x_i)$, which measures the delay as a function of the total flow on link i . We assume that the latency function l_i is convex, nondecreasing, and continuously differentiable. The analysis can be extended to the case when the links are capacity-constrained as in the previous section; see Acemoglu and Ozdaglar (2006b). We also assume that $l_i(0) = 0$ for all i .⁴ We denote the price per unit flow (bandwidth) of link i by p_i . Let $p = [p_1, \dots, p_I]$ denote the vector of prices.

We are interested in the problem of routing d units of flow across the I links. We assume that this is the aggregate flow of many “small” users and thus adopt the Wardrop’s principle (see Wardrop, 1952) in characterizing the flow distribution in the network; i.e., the flows are routed along paths with minimum effective cost, defined as the sum of the latency at the given flow and the price of that path. We also assume that the users have a homogeneous *reservation utility* R and decide not to send their flow if the effective cost exceeds the reservation utility.

More formally, for a given price vector $p \geq 0$, a vector $x^{\text{WE}} \in \mathbb{R}_+^I$ is a *Wardrop equilibrium* (WE) if

$$\begin{aligned} l_i(x_i^{\text{WE}}) + p_i &= \min_j \{l_j(x_j^{\text{WE}}) + p_j\}, & \forall i \text{ with } x_i^{\text{WE}} > 0, & \quad (22.22) \\ l_i(x_i^{\text{WE}}) + p_i &\leq R, & \forall i \text{ with } x_i^{\text{WE}} > 0, & \\ \sum_{i \in \mathcal{I}} x_i^{\text{WE}} &\leq d, & & \end{aligned}$$

with $\sum_{i \in \mathcal{I}} x_i^{\text{WE}} = d$ if $\min_j \{l_j(x_j^{\text{WE}}) + p_j\} < R$. We denote the set of WE at a given p by $W(p)$.⁵

We next define the social problem and the social optimum, which is the routing (flow allocation) that would be chosen by a planner that has full information and full control over the network. A flow vector x^S is a *social optimum* if it is an optimal solution of the *social problem*

$$\max_{\substack{x \geq 0 \\ \sum_{i \in \mathcal{I}} x_i \leq d}} \sum_{i \in \mathcal{I}} (R - l_i(x_i))x_i. \quad (22.23)$$

Hence, the social optimum is the flow allocation that maximizes the social surplus, i.e., the difference between users’ willingness to pay and total latency. For two links, let x^S be a social optimum with $x_i^S > 0$ for $i = 1, 2$. Then it follows from the definition that

$$l_1(x_1^S) + x_1^S l'_1(x_1^S) = l_2(x_2^S) + x_2^S l'_2(x_2^S). \quad (22.24)$$

This implies that the prices $x_i^S l'_i(x_i^S)$, i.e., the marginal congestion prices, can be used to decentralize the system optimum [cf. Eq. (22.22)].

⁴ This assumption is a good approximation to communication networks where queueing delays are more substantial than propagation delays. We will talk about the efficiency implications of relaxing this assumption in different models.

⁵ It is possible to account for additional constraints, such as capacity constraints on the links, by using a variational inequality formulation (see Acemoglu and Ozdaglar, 2006b; Correa et al., 2005).

For a given vector $x \geq 0$, we define the value of the objective function in the social problem,

$$\mathbb{S}(x) = \sum_{i \in \mathcal{I}} (R - l_i(x_i)) x_i, \quad (22.25)$$

as the *social surplus*, i.e., the difference between users' willingness to pay and the total latency.

22.2.3 Monopoly Pricing and Equilibrium

We first assume that a monopolist service provider owns the I links and charges a price of p_i per unit bandwidth on link i . The monopolist sets the prices to maximize his profit given by

$$\Pi(p, x) = \sum_{i \in \mathcal{I}} p_i x_i,$$

where $x \in W(p)$. This defines a two-stage dynamic *pricing-congestion game*, where the monopolist sets prices anticipating the demand of users, and given the prices (i.e., in each subgame), users choose their flow vectors according to the WE. We define a vector $(p^{\text{ME}}, x^{\text{ME}}) \geq 0$ to be a *Monopoly Equilibrium* (ME) if $x^{\text{ME}} \in W(p^{\text{ME}})$ and

$$\Pi(p^{\text{ME}}, x^{\text{ME}}) \geq \Pi(p, x), \quad \forall p \geq 0, \forall x \in W(p).^6$$

In Acemoglu and Ozdaglar (2006b), it was shown that price-setting by a monopolist internalizes the negative externality and achieves efficiency. In particular, a vector x is the flow vector at an ME if and only if it is a social optimum. This result was extended to a model that incorporates a general network topology in Huang et al. (2006). This is a significant departure from the existing performance results of selfish routing in the literature that assert that the efficiency losses with general latency functions can be arbitrarily bad.

22.2.4 Oligopoly Pricing and Equilibrium

We next assume that there are S service providers, denote the set of service providers by \mathcal{S} , and assume that each service provider $s \in \mathcal{S}$ owns a different subset \mathcal{I}_s of the links. Service provider s charges a price p_i per unit bandwidth on link $i \in \mathcal{I}_s$. Given the vector of prices of links owned by other service providers, $p_{-s} = [p_i]_{i \notin \mathcal{I}_s}$, the profit of service provider s is

$$\Pi_s(p_s, p_{-s}, x) = \sum_{i \in \mathcal{I}_s} p_i x_i,$$

for $x \in W(p_s, p_{-s})$, where $p_s = [p_i]_{i \in \mathcal{I}_s}$.

The objective of each service provider, like the monopolist in the previous section, is to maximize profits. Because their profits depend on the prices set by other service

⁶ Our definition of the ME is stronger than the standard subgame perfect Nash equilibrium concept for dynamic games. In Acemoglu and Ozdaglar (2006b), we show that the two solution concepts coincide for this game.

providers, each service provider forms conjectures about the actions of other service providers, as well as the behavior of users, which, we assume, they do according to the notion of (subgame perfect) Nash equilibrium. We refer to the game among service providers as the *price competition game*. We define a vector $(p^{\text{OE}}, x^{\text{OE}}) \geq 0$ to be a (pure strategy) *Oligopoly Equilibrium* (OE) if $x^{\text{OE}} \in W(p_s^{\text{OE}}, p_{-s}^{\text{OE}})$ and for all $s \in \mathcal{S}$,

$$\Pi_s(p_s^{\text{OE}}, p_{-s}^{\text{OE}}, x^{\text{OE}}) \geq \Pi_s(p_s, p_{-s}^{\text{OE}}, x), \quad \forall p_s \geq 0, \forall x \in W(p_s, p_{-s}^{\text{OE}}). \quad (22.26)$$

We refer to p^{OE} as the *OE price*.

Analysis of the optimality conditions for the oligopoly problem [cf. (22.26)] allows us to characterize the OE prices (see Acemoglu and Ozdaglar, 2006b). In particular, let $(p^{\text{OE}}, x^{\text{OE}})$ be an OE such that $p_i^{\text{OE}} x_i^{\text{OE}} > 0$ for some $i \in \mathcal{I}$. Then, for all $s \in \mathcal{S}$ and $i \in \mathcal{I}_s$,

$$p_i^{\text{OE}} = \begin{cases} x_i^{\text{OE}} l'_i(x_i^{\text{OE}}), & \text{if } l'_j(x_j^{\text{OE}}) = 0 \text{ for some } j \notin \mathcal{I}_s, \\ \min \left\{ R - l_i(x_i^{\text{OE}}), x_i^{\text{OE}} l'_i(x_i^{\text{OE}}) + \frac{\sum_{j \in \mathcal{I}_s} x_j^{\text{OE}}}{\sum_{j \notin \mathcal{I}_s} l'_j(x_j^{\text{OE}})} \right\}, & \text{otherwise.} \end{cases}$$

The preceding characterization implies that in the two link case with minimum effective cost less than R , the OE prices satisfy

$$p_i^{\text{OE}} = x_i^{\text{OE}} (l'_1(x_1^{\text{OE}}) + l'_2(x_2^{\text{OE}})) \quad (22.27)$$

as claimed before. Intuitively, the price charged by an oligopolist consists of two terms: the first, $x_i^{\text{OE}} l'_i(x_i^{\text{OE}})$, is equal to the marginal congestion price that a social planner would set [cf. Eq. (22.24)] because the service provider internalizes the further congestion caused by additional traffic. The second, $x_i^{\text{OE}} l'_j(x_j^{\text{OE}})$, reflects the markup that each service provider can charge users because of the negative congestion externality (as users leave its network, they increase congestion in the competitor network).

22.2.5 Efficiency Analysis

We investigate the efficiency properties of price competition games that have pure strategy equilibria.⁷ Given a price competition game with latency functions $\{l_i\}_{i \in \mathcal{I}}$, we define the efficiency metric at some oligopoly equilibrium flow x^{OE} as the ratio of the social surplus in the oligopoly equilibrium to the surplus in the social optimum [cf. Eq. 22.25 for the definition of the social surplus], i.e., the efficiency metric is given by

$$r_I(\{l_i\}, x^{\text{OE}}) = \frac{\mathbb{S}(x^{\text{OE}})}{\mathbb{S}(x^{\mathcal{S}})}, \quad (22.28)$$

where $x^{\mathcal{S}}$ is a social optimum given the latency functions $\{l_i\}_{i \in \mathcal{I}}$ and R is the reservation utility. In other words, the efficiency metric is the ratio of the social surplus in an equilibrium relative to the surplus in the social optimum. Following the literature on the “price of anarchy,” in particular Koutsoupias and Papadimitriou (1999), we are interested in the worst-case performance of an oligopoly equilibrium, so we look for

⁷ This set includes, but is substantially larger than, games with linear latency functions, see Acemoglu and Ozdaglar (2006a).

a lower bound on $r_I(\{l_i\}, x^{\text{OE}})$ over all price competition games and all oligopoly equilibria.

We next give an example of an I link network that has positive flows on all links at the OE and an efficiency metric of $5/6$.

Example 22.2 Consider an I link network where each link is owned by a different provider. Let the total flow be $d = 1$ and the reservation utility be $R = 1$. The latency functions are given by

$$l_1(x) = 0, \quad l_i(x) = \frac{3}{2}(I - 1)x, \quad i = 2, \dots, I.$$

The unique social optimum for this example is $x^S = [1, 0, \dots, 0]$. It can be seen that the flow allocation at the unique OE is $x^{\text{OE}} = [\frac{2}{3}, \frac{1}{3(I-1)}, \dots, \frac{1}{3(I-1)}]$. Hence, the efficiency metric for this example is $r_I(\{l_i\}, x^{\text{OE}}) = \frac{5}{6}$.

The next theorem establishes the main efficiency result.

Theorem 22.3 Consider a general parallel link network with $I \geq 2$ links and S service providers, where provider s owns a set of links $\mathcal{I}_s \subset \mathcal{I}$. Then, for all price competition games with pure strategy OE flow x^{OE} , we have

$$r_I(\{l_i\}, x^{\text{OE}}) \geq \frac{5}{6},$$

and the bound is tight.

A notable feature of Example 22.2 and this theorem is that the (tight) lower bound on inefficiency is independent of the number of links I and how these links are distributed across different oligopolists (i.e., of market structure). Thus arbitrarily large networks can feature as much inefficiency as small networks.⁸

22.2.6 Extensions

In this subsection, we extend the preceding analysis in two directions: First, we consider elastic traffic, which models applications that are tolerant of delay and can take advantage of even the minimal amounts of bandwidth (e.g., e-mail). We next focus on more general network topologies.

Elastic Traffic

To model elastic traffic, we assume that user preferences can be represented by an increasing, concave, and twice continuously differentiable *aggregate* utility function $u(\sum_{i \in \mathcal{I}} x_i)$, which represents the amount of utility gained from sending a total amount of flow $\sum_{i \in \mathcal{I}} x_i$ through the network.

⁸ This result superficially contrasts with theorems in the economics literature that large oligopolistic markets approach competitive behavior. These theorems do not consider arbitrary large markets, but replicas of a given market structure.

We assume that at a price vector, the amount of flow and the distribution of flow across the links is given by the Wardrop’s principle (Wardrop, 1952). In particular, for a given price vector $p \geq 0$, a vector $x^* \in \mathbb{R}_+^I$ is a Wardrop equilibrium if

$$l_i(x_i^*) + p_i = u' \left(\sum_{j \in \mathcal{I}} x_j^* \right), \quad \forall i \text{ with } x_i^* > 0,$$

$$l_i(x_i^*) + p_i \geq u' \left(\sum_{j \in \mathcal{I}} x_j^* \right), \quad \forall i \in \mathcal{I}.$$

We define the social optimum and the efficiency metric as in Eqs. (22.23) and (22.28), replacing $R \sum_{i \in \mathcal{I}} x_i$ (i.e., users’ willingness to pay) by $u(\sum_{i \in \mathcal{I}} x_i)$.

It can be shown that for elastic traffic with a general concave utility function, the efficiency metric can be arbitrarily close to 0 (see Ozdaglar, 2006). The two-stage game with multiple service providers and elastic traffic with a single user class was first analyzed by Hayrapetyan, Tardos and Wexler (2005). Using an additional assumption on the utility function (i.e., the utility function has a concave first derivative), their analysis provides nontight bounds on the efficiency loss.⁹ Using mathematical tools similar to the analysis in Acemoglu and Ozdaglar (2006b), the recent work (Ozdaglar, 2006) provides a tight bound on the efficiency loss of this game, as established in the following theorem.

Theorem 22.4 *Consider a parallel link network with $I \geq 1$ links, where each link is owned by a different provider. Assume that the derivative of the utility function, u' is a concave function. Then, for all price competition games with elastic traffic and pure strategy OE flow x^{OE} , we have*

$$r_I(u, \{l_i\}, x^{\text{OE}}) \geq \frac{2}{3},$$

and the bound is tight.

Parallel-Serial Topologies

Most communication networks cannot be represented by parallel link topologies, however. A given source-destination pair will typically transmit through multiple interconnected subnetworks (or links), potentially operated by different service providers. Existing results on the parallel-link topology do not address how the cooperation and competition between service providers will impact efficiency in such general networks.

Here, we take a step in this direction by considering the simplest network topology that allows for serial interconnection of multiple links/subnetworks, which is the parallel-serial topology (see Figure 22.2). It was shown in Acemoglu and Ozdaglar (2006a) that the efficiency losses resulting from competition are considerably higher with this topology. When a particular provider charges a higher price, it creates a negative externality on other providers along the same path, because this higher price reduces

⁹ For example, they provide the nontight bound of 1/5.064 in general, and the bound of 1/3.125 for the case when latency without congestion is 0.

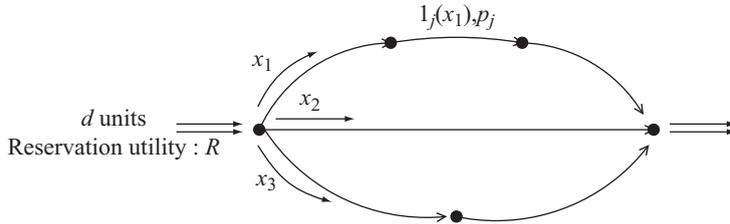


Figure 22.2. A network with serial and parallel links.

the transmission that all the providers along this path receive. This is the equivalent of the *double marginalization* problem in economic models with multiple monopolies and is the source of the significant degradation in the efficiency performance of the network.

In its most extreme form, the double marginalization problem leads to a type of “coordination failure,” whereby all providers, expecting others to charge high prices, also charge prohibitively high prices, effectively killing all data transmission on a given path. We may expect such a pathological situation not to arise since firms should not coordinate on such an equilibrium (especially when other equilibria exist). For this reason, we focus on a stronger concept of equilibrium introduced by Harsanyi, the *strict equilibrium*. In strict OE, each service provider must play a strict best response to the pricing strategies of other service providers. We also focus our attention on equilibria in which all traffic is transmitted (otherwise, it can be shown that the double marginalization problem may cause entirely shutting down transmission, resulting in arbitrarily low efficiency, see Acemoglu and Ozdaglar, 2006a).

The next theorem establishes the main efficiency result for this topology.

Theorem 22.5 Consider a general $I \geq 2$ path network, with serial links on each path, where each link is owned by a different provider. Then, for all price competition games with strict OE flow x^{OE} , we have

$$r_I(x^{\text{OE}}) \geq \frac{1}{2},$$

and the bound is tight.

Despite this positive result, it was shown in Acemoglu and Ozdaglar (2006a) that when the assumption $l_i(0) = 0$ is relaxed, the efficiency loss of strict OE relative to the social optimum can be arbitrarily large. This suggests that unregulated competition in general communication networks may have considerable costs in terms of the efficiency of resource allocation and certain types of regulation may be necessary to make sure that service provider competition does not lead to significant degradation of network performance.

22.3 Alternative Pricing and Incentive Approaches

The two approaches we have presented so far incorporate many of the important ideas in the role of prices and incentives in communication networks. Nevertheless, a variety of different approaches have also been developed in the literature, and the models presented in the previous two sections leave out several interesting aspects, which can be studied in future work. In this section, we first discuss the previous work on pricing in networks. We then mention several alternative approaches pursued in ongoing work. We conclude with a number of areas for future research.

22.3.1 Previous Work on Pricing

Despite the fact that current Internet access is based on a flat access charge, it has been recognized that the future of the Internet will involve multiple service classes, their use regulated by differentiated prices. The most natural approach to this problem involves the modeling of profit-maximizing service providers as developed in the previous section. Here we discuss some other aspects involved in the use of such prices.

Pricing for Differentiated Services: Service differentiation brings in a clear need for offering incentives to users to encourage them to choose the service appropriate for their needs, hence preventing overutilization of network resources. Pricing mechanisms provide an efficient way to ensure QoS guarantees and regulate system usage. One of the key debates in network pricing area is whether charges should be based on fixed access prices or usage-based prices. While usage-based pricing has the potential to fulfill at least partially the role of a congestion control mechanism, there were criticisms in view of the apparent disadvantages of billing overheads and the resulting uncertainties in networking expenses (see DaSilva, 2000).

A variety of pricing mechanisms have been proposed over the last decade. A well-known usage-based pricing proposal is by Mackie-Mason and Varian (1995), who proposed a “smart market” for resource allocation over a single link. In this scheme, users bid for transmission of each individual packet while the network provides service to packets whose bid exceeds a cutoff level determined by the marginal willingness-to-pay and marginal congestion costs. Users do not pay the price they bid, but rather the market-clearing price which is lower than the bids of all admitted packets. This mechanism resembles the Vickrey auction, and therefore provides users the correct incentives to reveal their true values in their bids. Odlyzko, in his seminal Paris Metro Pricing proposal (1990), suggested partitioning the network into several logical subnetworks. Users choose one of these logical networks for the transmission of their traffic, and this implicitly defines the service level; i.e., higher-priced networks will experience lower utilizations, and therefore will be able to provide a higher service level. Other proposed pricing schemes include *edge-pricing*, which focuses on locally computed charges based on expected values of congestion levels and routes; *expected capacity pricing*, in which users are charged according to the expected capacity the network provisions; and *effective bandwidth pricing*, which proposes the pricing of real-time traffic with QoS requirements, in terms of its “effective bandwidth”; see DaSilva (2000) for an overview of various pricing mechanisms.

First-Best Pricing: There is also a large theoretical literature in both communication networks and transportation networks area that study control mechanisms to induce efficient allocation of resources among competing users. The main focus is to use prices (or tolls) to induce flow patterns that optimize an overall system objective (also referred to as *first-best pricing*). It is well-known that *marginal cost pricing*, i.e., charging individual users for the negative (congestion) externality they impose on other users, achieves the system optimal flows. A number of studies have also characterized the “toll set,” i.e., the set of all tolls that induce optimal flows, with the goal of choosing tolls from this set according to secondary criteria, e.g., minimizing the total amount of tolls or the number of tolled routes; see Hearn and Ramana (1998). Other related work focuses on models with heterogeneous users (i.e., users with different congestion-price sensitivities) and studies tolls that induce system optimal flows (see Cole et al., 2003; Fleischer et al., 2004).

22.3.2 Current Research on Pricing and Incentive Models

Many other game-theoretic models are useful in studying communication networks. Instead of providing a comprehensive survey, we now discuss a few models that are of significant practical relevance.

Fixed Pricing and the Marginal User Principle: As mentioned in the previous subsection, for various practical reasons (some of which are perhaps simply legacy reasons), consumers are accustomed to paying a flat-fee (e.g., monthly) for their service. In markets with a flat fee, typically a service provider has some idea of the distribution of the user’s utility functions but not the utility function of each individual user.

An important problem therefore is to determine the fixed flat fee that maximizes the service provider revenue and to understand the impact of such a pricing scheme on the allocation of resources. In Acemoglu et al. (2004), we show that in a wireless network the profit-maximizing fixed price is equal to the utility of the marginal user in the network, where the marginal user is defined as a user who is indifferent to joining the network. Since the price and the resource allocation scheme determine the marginal user, they have to be chosen jointly to maximize the network revenue and it has been shown in Acemoglu et al. (2004) that such a resource allocation algorithm and price can be computed by the service provider under certain assumptions on the utility functions.

Incentives for Cooperation in P2P Networks: It is estimated that nearly half the traffic in today’s Internet is due to peer-to-peer (P2P) networks. P2P networks are used to typically share large files among users. Some well-known examples of P2P networks are BitTorrent, Gnutella, KaZaa, etc. A P2P network is a collection of a large number of users who contribute some resources (typically, bandwidth, and memory) to not only download files of interest to themselves but to also store and transmit files that may be of interest to others. A P2P network has remarkable scaling properties compared to a Web server that stores many files that can be downloaded by users. A Web server has finite upload bandwidth and therefore, as more users join the network, the bandwidth per user has to decrease. On the other hand, in a P2P network since each user is a potential user as well as a server, as the number of users in the network increases, the capacity of the network also increases to keep up with the demand. In fact, simple

analytical models suggest that there is no loss of performance as the number of users increases in a BitTorrent-type network (Qiu and Srikant, 2004). However, such scaling benefits can be achieved only if users cooperate. For example, if all users are only willing to download but refuse to upload files, then the network capacity will not scale with the number of users. Networks such as BitTorrent have some simple built-in incentive mechanisms to combat such problems and these have been studied in Qiu and Srikant (2004). As P2P networks continue to proliferate, it becomes quite important to study incentive mechanisms for such networks. Such issues are studied elsewhere in this book.

Incentives for Cooperation in Wireless Networks: Another form of networking that is expected to see tremendous growth in the near future is multihop wireless networks. In such networks, laptop computer or other mobile radio devices will communicate with each other in a multihop fashion without any infrastructure such as an access point or a base station. For such communication to be feasible, each radio must be willing to forward packets for other users in the network. While on the face of it, the problem appears to be similar to the case of P2P networks, there are some key differences. In a wireless network, since the communication medium is shared, it is possible for a wireless node (say node A) to hear whether a neighbor (call it node B) is being selfish or not. For example, if node A forwards a packet (destined for another node) to node B , then A can listen to see if B forwarded the packet or not. However, if another neighbor of A (say, node C) transmits at the same time as node B , then A will not hear B 's transmission and thus, may erroneously assume that B is a selfish user. This is similar to a prisoner's dilemma model with noisy observations of the players' true actions (Piccione, 2002) and has been studied in He et al. (2004) and Mahajan et al. (2005) in a non-game-theoretic setting and in Milan et al. (2006) using game theory. However, the models used for the analysis of cooperation in multihop radio networks are currently quite simplistic and ignore the topological structure of the network. It is an open problem to develop more detailed models of the network and medium-access protocols, and to study the game-theoretic interactions for these more realistic models.

22.3.3 Areas for Future Research

The models presented so far highlight a number of fruitful areas for future research. These include but are not limited to the following topics.

Incentive-compatible Differentiated Pricing: As discussed above, a key role of prices in networks will be in allocating users with different requirements to differentiated services. If the service requirements and other characteristics of users were known by a central controller or service providers, this problem would be similar to those studied above. In practice, however, such information is not available and the market mechanism (i.e., the pricing scheme) has to ensure that individuals choose the services designed for them. This problem can be analyzed as a combination of the competition models developed above and the classical mechanism design approach. In particular, the celebrated Revelation Principle in the mechanism design theory (see Mas-Colell et al., 1995) implies that we can think of direct mechanisms in which individuals truthfully report their types, and are allocated services and charged prices accordingly.

The mathematical formulation then necessitates that a set of incentive-compatibility constraints that make truthful reporting optimal for each user is satisfied. The modeling challenge in this approach lies in combining the competition among service providers and the incentive-compatibility constraints.

Capacity Investments: While the focus of the current literature has been in ensuring the efficiency of the allocation of existing network resources, an arguably more important problem is to ensure that the right amount and type of infrastructure investment and capacity are installed in newly emerging networks. The analysis of this set of problems requires (multi-stage) models in which service providers choose not only prices but also investment levels and capacities.

Simple Pricing Rules: One potential criticism of economic approaches for resource allocation in networks is whether the complicated pricing schemes necessary for achieving socially optimal or profit-maximizing allocations can be computed and implemented in real time. The question of whether simple pricing rules can approximate these objectives and the quantification of the extent of efficiency or profits from such simple rules constitute another area for future research.

Bibliography

- D. Acemoglu and A. Ozdaglar. Competition in Parallel-Serial Networks. To appear in *IEEE J. Selected Areas in Commun.*, special issue on Non-cooperative Behavior in Networking, 2006a.
- D. Acemoglu and A. Ozdaglar. Competition and efficiency in congested markets. To appear in *Math. Operat. Res.*, 2006b.
- D. Acemoglu and A. Ozdaglar. Flow control, routing, and performance from service provider viewpoint. *LIDS report*, WP-1696, May 2004.
- D. Acemoglu, A. Ozdaglar and R. Srikant. The marginal user principle for resource allocation in wireless networks. *Proc. of CDC*, 2004.
- T. Basar and R. Srikant. A Stackelberg network game with a large number of followers. *J. Optimization Theory Appl.*, 115(3):479–490, December 2002a.
- T. Basar and R. Srikant. Revenue-maximizing pricing and capacity expansion in a many-users regime. *Proc. of INFOCOM*, 2002b.
- S. Cho and A. Goel. Pricing for fairness: distributed resource allocation for multiple objectives. To appear in *ACM Symp. Theory of Computing*, 2006.
- R. Cole and Y. Dodis and T. Roughgarden. Pricing network edges for heterogeneous selfish users. *Proc. of STOC*, 2003.
- J.R. Correa, A.S. Schulz and N.S. Moses. On the inefficiency of equilibria in congestion games. *Proc. of IPCO*, pp. 167–181, 2005.
- L.A. DaSilva. Pricing for QoS-enabled networks: a survey. *IEEE Communication Surveys and Tutorials*, 3(2):2–8, 2000.
- L. Fleischer, K. Jain and M. Mahdian. Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games. *Proc. of FOCS*, pp. 277–285, 2004.
- A. Hayrapetyan, E. Tardos and T. Wexler. A network pricing game for selfish traffic. *Proc. of ACM SIGACT-SIGOPS Symp. Princ. of Distributed Computing*, pp. 284–291, 2005
- Q. He, D.Wu and P. Khosla, SORI: A secure and objective reputation based incentive scheme for ad-hoc networks. In *Proc. of IEEE Wireless Communications and Networking Conference (WCNC2004)*, Atlanta, GA, pp. 825–830, 2004.

- D.W. Hearn and M.V. Ramana. Solving congestion toll pricing models. In P. Marcotte and S. Nguyen, editors, *Proc. of the Equilibrium and Advanced Transportation Modelling Colloquium*, pp. 109–124, 1998.
- X. Huang, A. Ozdaglar and D. Acemoglu. Efficiency and Braess' Paradox under pricing in general networks. *IEEE J. Selected Areas Commun.*, 24(5):977–991, 2006.
- F.P. Kelly. Charging and rate control for elastic traffic. *Euro. Trans. on Telecommun.*, 8:33–37, 1997.
- F.P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *J. Operational Research Society*, 49:237–252, 1998.
- E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In: *Proc. 16th Symp. on Theoretical Aspects of Computer Science*, pp. 404–413, 1999.
- S. Kunniyur and R. Srikant. A time-scale decomposition approach to adaptive ECN marking. *IEEE Trans. on Automatic Control*, June 2002.
- X. Lin, N.B. Shroff and R. Srikant. Cross-layer design in wireless networks: A tutorial. To appear in *IEEE J. Selected Areas Commun.*, June 2006.
- S.H. Low and D.E. Lapsley. Optimization flow control—I: basic algorithm and convergence. *IEEE/ACM Trans. on Networking*, 7(6):861–874, December 1999.
- J.K. Mackie-Mason and H. Varian. Pricing congestible network resources. *IEEE J. Selected Areas Commun.*, 13(7):1141–1149, 1995.
- R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, Sustaining cooperation in multi-hop wireless networks. In *Proc. Second USENIX Symp. on Networked System Design and Implementation (NSDI 05)*, Boston, MA, May 2005.
- A. Mas-Colell, M.D. Whinston, and J.R. Green. *Microeconomic Theory*, Oxford University Press, NY, 1995.
- F. Milan, J.J. Jaramillo and R. Srikant. Sustaining cooperation in a multi-hop wireless network with selfish nodes. To appear *Proc. of Workshop on Game Theory for Networks (GameNets '06)*, Pisa, Italy, October 2006.
- A.M. Odlyzko. Paris Metro Pricing for the Internet. In *Proc. of the 1st ACM Conf. Electronic Commerce*, pp. 140–147, 1999.
- A. Ozdaglar. Price competition with elastic traffic. LIDS report, 2006.
- M. Piccione. The repeated prisoner's dilemma with imperfect private monitoring. *J. Econ. Theory*, 70–83, 2002.
- D. Qiu and R. Srikant, Modeling and performance analysis of BitTorrent-like peer-to-peer networks. *Computer Communications Review: Proc. ACM SIGCOMM*, Portland, OR, Sept. 2004.
- R. Srikant *The Mathematics of Internet Congestion Control*, Birkhauser, 2004.
- J.G. Wardrop. Some theoretical aspects of road traffic research. In: *Proc. of the Institute of Civil Engineers*, II, 1:325–378, 1952.
- H. Yaiche, R. Mazumdar, and C. Rosenberg. A game theoretic framework for bandwidth allocation and pricing in broadband networks. *IEEE/ACM Trans. on Networking*, 8(5):667–678, Oct. 2000.