November 2004

Flow Control, Routing, and Performance with a For-profit Service Provider

by

Daron Acemoglu [1] **and Asuman Ozdaglar** [2]

Abstract

We consider a game theoretic framework to analyze traffic in a congested network, where a profit-maximizing monopolist sets prices for different routes. Each link in the network is associated with a flow-dependent latency function which specifies the time needed to traverse the link given its congestion. Users have utility functions defined over the amount of data flow transmitted, the delays they incur in transmission, and the expenditure they make for using the bandwidth. Given the prices of the links, each user chooses the amount of flow to send and the routes to maximize the utility he receives according to the notion of Wardrop equilibrium. We define a monopoly equilibrium (ME) as the equilibrium prices set by the monopolist and the corresponding Wardrop equilibrium. We use this framework to study the performance of the ME relative to the Wardrop equilibrium without prices and the social optimum, which would result from the choice of a network planner with full information and full control over the flow and routing choices of users. We illustrate that ME can improve the performance over the Wardrop equilibrium without prices because the monopolist internalizes the effects of increased congestion on different paths due to lack of coordination among users. We next consider a model for the routing problem, where each user has a fixed amount of data to transmit, under monopoly pricing. Despite nonconvexities in the model, we show that there exists a Wardrop equilibrium under profit-maximizing prices and that the ME always achieves the social optimum.

Keywords: Flow control, routing, Wardrop equilibrium, monopoly pricing, externalities, efficiency.

[1] Dept. of Economics, M.I.T., Cambridge, Mass., 02142.
[2] Dept. of Electrical Engineering and Computer Science, M.I.T., Cambridge, Mass., 02139.

1. INTRODUCTION

A fundamental problem in communication and data networks is the management of congestion, both to ensure timely transmission of information and to prevent loss of data in transmission. The standard approach is to use optimization methods to achieve the best potential network performance by adjusting the input flow rates of users and routing the resulting traffic. However, in many scenarios, it is impossible or impractical to regulate the traffic in such a centralized manner. Moreover, this approach requires considerable knowledge about the needs (preferences) of all the users in the network, an increasingly unrealistic assumption in the major networks of today such as the Internet.

The recognition of this problem has motivated a recent theoretical literature to consider the selfish flow choice and routing behavior of users in the absence of central planning (see, among others, [ORS93], [Kel97], [KMT98], [LoL99], [ABS02], [BaS02], [RoT02], [JoT03]). In these models, individuals choose their input flow rates and the routes to optimize their own objective, and are assumed to form conjectures about the behavior of other users consistent with the game theoretic notion of Nash Equilibrium, or the notion of Wardrop Equilibrium first introduced in the analysis of congestion in transportation networks (see [War52]). Not surprisingly, the resulting allocation differs markedly from the full-information social optimum, which would be chosen by a network manager with full information. For example, a recent paper by Roughgarden and Tardos [RoT03] studies a model where selfish agents decide the routing of a given traffic rate and finds that generally the performance of such "selfish" routing can be much worse than the full-information optimum, but they also provide a bound on the performance gap in a specific case (when the latency functions are linear).

With a few notable exceptions (e.g., [BaS02], [MRW01], [HeW03]), however, this literature considers only situations in which users face no monetary costs of sending information. In contrast, many real-world data networks are for-profit entities that charge prices for transmission of information. In this paper, we construct a model to analyze traffic in a congested network where there are prices associated with different routes and a profit-maximizing monopolist setting prices. Our objective is to investigate the consequences of profit-maximizing pricing on the performance of a network with selfish flow control and routing.

Our model consists of a parallel-link network topology, a large number of heterogeneous users, and a profit-maximizing monopolist. Each link in the network is associated with a flow-dependent latency function specifying the time needed to traverse this link given the total flow through this link. Users have potentially different utility functions defined over the amount of data flow transmitted, the delays they incur in transmission, and the expenditure they make for using the bandwidth. The monopolist sets prices per unit bandwidth for each link, and each user pays a price proportional to the amount of bandwidth he uses over the links. Given prices of the links, each user chooses the amount of flow to send and the routes. Formally, this corresponds to a two-stage game, where the monopolist is the Stackelberg leader setting prices anticipating the subsequent behavior of users. Each price vector defines a different subgame, and users play this (sub)game taking prices as given. It is important to note that, even though users are price-takers, they have to anticipate the amount of congestion on all routes, which they do according to the notion of Nash Equilibrium (NE) or Wardrop Equilibrium (WE). In the NE, users take the effect of their own flow on congestion into account, while in the WE, they ignore their own effect. Therefore, in terms of the general equilibrium analysis in the economics literature, the WE also corresponds to a generalized competitive equilibrium where agents are "latency takers" (i.e., they take latency on each route as given).[1] We define a Monopoly Equilibrium (ME) as the equilibrium prices set by the monopolist and the corresponding WE. An important result of our paper is that despite the potential non-convexities in the corresponding optimization problem, it is possible to provide an explicit characterization of equilibrium prices.

Characterization of equilibrium prices enables us to study the performance of the ME relative to the full-information social optimum and the WE without prices (i.e., at zero prices). We show that the WE without prices always generates too much flow for each user relative to the social optimum. This is because each user creates an externality on other users by making the routes that he or she uses more congested.[2] Monopoly pricing may improve the performance of the system because the monopolist internalizes this externality (or put differently, monopoly prices induce users to internalize this externality). In particular, a key insight of our analysis is that

[1] See, for example, Debreu [Deb59], or Arrow and Hahn [ArH71] on general competitive analysis. The major difference between the analysis here and the standard models of competitive analysis is the presence of congestion externalities. See the next footnote for a definition of an externality.

[2] An externality arises when an individual's actions affects the utility of others through non-market means. For example, in our framework when latency functions are not constant, by using a particular link, an individual creates a negative externality on all other users transmitting data through that link. The negative externality implies that individuals tend to transmit too much data and create too much congestion.

the monopolist realizes that a higher price for a particular route may not reduce the attraction of this route to users by much, because, with any reduction in traffic, there will be a corresponding decrease in congestion. We show that for the routing problem (i.e., where each user has a given traffic rate), the ME achieves the full-information social optimum despite the selfish behavior of both users and the monopolist for all strictly increasing latency functions. This results illustrates a significant difference in performance of selfish routing compared to the case where there is no service provider (see [RoT02]).

Our paper is most closely related to [BaS02] and [MRW01]. [BaS02] studies the asymptotic user behavior of a model with a single service provider, but it considers only specific utility and latency functions and does not investigate the efficiency properties of the ME. [MRW01] also considers the profit maximization objective and studies pricing and routing in a model with constant elasticity demand. Our paper is different from this work since it derives demand from user objectives and focuses on the implications of profit maximization for network performance. Our paper is also related to research on various strategies to cope with the inefficiency created by the selfish behavior of the agents (e.g., taxing or artificial Stackelberg games, see [KLO97], [CDR03]). But, except for special scenarios, these can be viewed as outside interventions in the system (e.g., by a government regulator), whereas the presence of a service provider in the system automatically creates a natural market mechanism and in some important cases, the right incentives to price the effects of congestion. To the best of our knowledge, this is a new insight in the literature.

The paper is organized as follows. In Section 2, we introduce our model in detail. For a given price vector, we define an equilibrium among the users of the network and explore the properties of this equilibrium. In Section 3, we define and characterize the Monopoly Equilibrium (ME). In Section 4, we compare the ME to the social optimum. In Section 5, we present a model for the routing problem under monopoly pricing and analyze the performance of the monopolized system. We present our conclusions and summarize future directions in Section 6.

Regarding notation, all vectors are viewed as column vectors, and a prime denotes transposition, so $x'y$ denotes the inner product of the vectors x and y.

2. MODEL

We consider a network with I parallel links accessed by J users. Let $\mathcal{I} = \{1, \ldots, I\}$ denote the set of links and $\mathcal{J} = \{1, \ldots, J\}$ denote the set of users. Let x_j^i denote the flow of user j on link i and $x_j = [x_j^1, \ldots, x_j^I]'$ denote the vector of flows of user j. We assume that user j receives a utility of

$u_j(\Gamma_j)$, where

$$\Gamma_j = \sum_{i=1}^{I} x_j^i$$

denotes the total flow rate of user j on all links. Each link in the network has a flow-dependent latency function $l^i(\gamma^i)$, where

$$\gamma^i = \sum_{j=1}^{J} x_j^i \qquad (2.1)$$

denotes the total flow (link load) on link i. The latency function represents the delays in transmission as a function of the link load. We denote the price per unit bandwidth of link i by p^i, and we will later allow a monopolist service provider to control prices.

An important aspect of our model is our assumption that each user is "small" in the sense that when he switches his flows from one path to another, there is no considerable change in the link latencies. This is known as the "Wardrop's Principle", due to a paper by Wardrop [War52]. Wardrop's Principle is used extensively in characterizing user behavior in transportation networks (see [BMW56], [DaN83]) and communication networks with inelastic traffic; i.e., users have fixed rates to transmit and only their routing choices are subject to optimization (see [RoT02], [ScS03], [CSS03]). Here, we use this idea for the problem of combined rate control and routing. [HaM85] show that the Wardrop Equilibrium can be obtained as the limit of a sequence of Nash Equilibria of games as the number of users goes to infinity, a result generalized to this setup in [AcO04]. Hence, with a large number of users, the focus on WE in this context is natural.

We denote the vector of total flows on the links [cf. Eq. (2.1)] by $\gamma = [\gamma^1, \dots, \gamma^I]$, and the vector of prices of the links by $p = [p^1, \dots, p^I]$. Given γ and p, each user chooses $x_j \geq 0$ to maximize his payoff function $v_j(x_j; \gamma, p)$, given by

$$v_j(x_j; \gamma, p) = u_j\Big(\sum_{i=1}^{I} x_j^i\Big) - \sum_{i=1}^{I} l^i(\gamma^i) x_j^i - \sum_{i=1}^{I} p^i x_j^i. \qquad (2.2)$$

The fact that user utility is additively-separable between total flow, Γ_j, and delay on link i, $l^i(\gamma^i)$, is a useful simplification in line with the rest of the literature. The important feature is that the utility of each user (or alternately, the total flow allocated to each user) depends on the flows of all the users. The notation where the utility function is conditioned on the price vector, p, and the link load vector, γ also emphasizes that, consistent with the Wardrop's Principle, each user acts not only as a "price taker", but also as a "link load taker".

An equilibrium of this game is defined as follows.

Definition 2.1: Let $x = [x'_1, \ldots, x'_J]$ denote the vector of flows of all the users in the network. For a given price vector $p \geq 0$, a flow vector x is a *Wardrop Equilibrium* (WE) of the game where the payoff functions of the users are given by Eq. (2.2) if

$$x_j \in \arg \max_{0 \leq y_j^i \leq C^i, \ \forall \ i} v_j(y_j; \gamma, p), \qquad \forall \ j \in \mathcal{J},$$

where C^i denotes the capacity of link i, and

$$\gamma^i = \sum_{j=1}^{J} x_j^i, \qquad \forall \ i \in \mathcal{I}.$$

Note that the WE is a function of the price vector p. Whenever there is a possibility of confusion, we write it as $x(p)$ to make the dependence on price explicit. We also use the notation Γ_j to denote the flow rate of user j, and γ^i to denote the load of link i at this equilibrium. We will adopt the following assumption on the utility and latency functions throughout the paper except Section 5.

Assumption 2.1: Assume that for each j, the utility function $u_j : [0, \infty) \mapsto [0, \infty)$ satisfies the following conditions:

(a) u_j is strictly concave, nondecreasing, and continuously differentiable.

(b) $0 < u'_j(0) < \infty$.

Also assume that for each i, the link latency function $l^i : [0, C^i) \mapsto [0, \infty)$ satisfies the following conditions:

(a') l^i is continuous and strictly increasing.

(b') $l^i(0) = 0$.

(c') $l^i(x) \to \infty$ as $x \to C^i$.

Assumption 2.1 is natural in this context. Concavity of the utility function implies that we are considering elastic traffic (i.e., traditional data applications like file transfer and e-mail, which are tolerant of delays, see [She95]), for which transmitting more data has diminishing returns for users. In Section 5, we will also consider non-concave utility functions. Assumption 2.1(b') imposes that the cost of transmitting information without any congestion is equal to 0 on all links.[1] Finally,

[1] We will see that this assumption is essential to guarantee that all link loads are positive at the ME. The main equilibrium price characterization (cf. Proposition 3.7) and the results of Sections 4 and 5 remain valid, even when this assumption is relaxed.

Assumption 2.1(c') captures the capacity constraints on the links.[1] In view of this assumption and Assumption 2.1(b), it can be seen that at any WE, $x(p)$, we have $\sum_{j\in\mathcal{J}} x_j^i(p) < C^i$. Therefore, the capacity constraint in the definition of the WE can be neglected.

In the next proposition, we show that under Assumption 2.1, a WE always exists. Although the existence of an equilibrium (WE) is a useful starting point, in many general equilibrium environments there can be multiple equilibria, making comparisons between equilibria and system optimum difficult. Therefore the next proposition establishes the essential uniqueness of the equilibrium. By essential uniqueness, we mean that, even though the x's may not be unique, the flow rate of each user and the load of each link are uniquely defined. The proof uses a classical technique that was used for showing related uniqueness results for the Wardrop equilibrium (see [BMW56], [Kel91]), and is therefore omitted.

Proposition 2.1: (Existence-Essential Uniqueness) Let Assumption 2.1 hold. For a given $p \geq 0$, consider the J-player game where the payoff functions of the users are given by Eq. (2.2). Then, there exists a WE, and the flow rates and the link loads at any WE are unique.

Sensitivity to Prices

The essential uniqueness of the equilibrium enables us to examine the continuity and monotonicity properties of the link loads and flow rates of users as functions of the link prices. These results will be key in showing the properties of the ME in the next section. The proofs of these results are omitted for brevity and can be found in [AcO04].

Proposition 2.2: (Continuity) Let Assumption 2.1 hold. For a given $p \geq 0$, let $x(p)$ be a WE of the J-player game where the payoff functions of the users are given by Eq. (2.2). Denote the corresponding link loads and flow rates by $\gamma^i(p)$, $i \in \mathcal{I}$, and $\Gamma_j(p)$, $j \in \mathcal{J}$. Then, for all i and j, $\gamma^i(p)$ and $\Gamma_j(p)$ are continuous at all $p \geq 0$.

The next proposition establishes the monotonicity of flow rates in prices, and captures the intuitive notion that when prices are higher, users choose lower (no higher) flow rates.

Proposition 2.3: (Monotonicity of Flow Rates) Let Assumption 2.1 hold. Let p and $\tilde{p}$ be two price vectors such that $p \geq \tilde{p}$. Then, we have

$$\Gamma_j(p) \leq \Gamma_j(\tilde{p}), \qquad \forall\, j \in \mathcal{J},$$

[1] This assumption guarantees that no individual has an infinite demand or infinite willingness to pay for data transmission. An alternative assumption could be that, for each j, there exists a nonzero scalar B_j such that $u_j'(B_j) = 0$.

where $\Gamma_j(p)$ and $\Gamma_j(\tilde{p})$ are the flow rates of user j at a WE given prices p and $\tilde{p}$, respectively.

The next proposition shows the monotonicity of link loads in prices. It establishes that link loads are non-increasing in their own prices and non-decreasing in the prices of other links. Both of these are intuitive; the first implies that a higher price reduces the traffic on the link, while the second implies that, from the point of view of users, different links are substitutes.

Proposition 2.4: (Monotonicity of Link Loads) Let Assumption 2.1 hold. For some scalar $\epsilon > 0$ and some $t = 1, \dots, I$, let e_t denote the vector whose t^{th} component is ϵ and all the remaining components are equal to 0. Then, we have

$$\gamma^t(p + e_t) \leq \gamma^t(p), \tag{2.3}$$

$$\gamma^s(p + e_t) \geq \gamma^s(p), \qquad \forall\ s \neq t. \tag{2.4}$$

We next establish two lemmas which enable a sharper characterization of the WE. This characterization will be useful in the analysis of subsequent sections. The following lemma is immediate using the optimality conditions.

Lemma 2.1: Let Assumption 2.1 hold. For a given $p \geq 0$, let γ^i be the load of link i at a WE. Then, for all i with $\gamma^i > 0$, we have

$$p^i + l^i(\gamma^i) = \min_{m \in \mathcal{I}} \big\{ p^m + l^m(\gamma^m) \big\}.$$

Let us define the effective cost of link i as $p^i + l^i(\gamma^i)$, which is the monetary cost plus delay cost of using the link. The lemma states that the effective costs of all links with positive flow must be equal.

Lemma 2.2: Let Assumption 2.1 hold. For a given $p \geq 0$, let x be a WE. Define the sets $\overline{\mathcal{I}} = \{i \mid \gamma^i > 0\}$ and $\overline{\mathcal{J}} = \{j \mid \Gamma_j > 0\}$.

(a) Then for all $i \in \overline{\mathcal{I}}$ and $j \in \overline{\mathcal{J}}$, we have

$$u_j'(\Gamma_j) - l^i(\gamma^i) - p^i = 0.$$

(b) There exists a WE $\tilde{x}$ at this price vector that satisfies $\tilde{x}_j^i > 0$ for all $i \in \overline{\mathcal{I}}$ and $j \in \overline{\mathcal{J}}$.

Proof:

(a) Let $i \in \overline{\mathcal{I}}$ and $j \in \overline{\mathcal{J}}$. Since $\Gamma_j > 0$ by assumption, it follows that there exists some link s such that $x_j^s > 0$, which implies by the first order conditions that

$$u_j'(\Gamma_j) - l^s(\gamma^s) - p^s = 0.$$

Since $\gamma^s > 0$ and $\gamma^i > 0$, we have by Lemma 2.1 that

$$l^i(\gamma^i) + p^i = l^s(\gamma^s) + p^s.$$

Substituting this in the previous equation yields the desired result.

(b) Consider the vector $\tilde{x} = [\tilde{x}^i_j]$ generated in the following way. Set $\tilde{x}^i_j = 0$ if $i \notin \overline{\mathcal{I}}$ or $j \notin \overline{\mathcal{J}}$. Let $i \in \overline{\mathcal{I}}$ and $j \in \overline{\mathcal{J}}$. If $x^i_j > 0$, then set $\tilde{x}^i_j = x^i_j$. Assume $x^i_j = 0$. Then, since $i \in \overline{\mathcal{I}}$ and $j \in \overline{\mathcal{J}}$, there exists some $k \neq j$ and $s \neq i$ such that $x^i_k > 0$ and $x^s_j > 0$. We set

$$\tilde{x}^i_j = \epsilon, \qquad \tilde{x}^s_j = x^s_j - \epsilon,$$

$$\tilde{x}^i_k = x^i_k - \epsilon, \qquad \tilde{x}^s_k = x^s_k + \epsilon,$$

where $\epsilon > 0$ is small enough such that all of the above terms are positive. It can be seen that the link loads and the input flow rates of users are kept constant by this transformation (to see this, add vertically to see the change in link loads and horizontally to see the change in flow rates). Also, since $k \in \overline{\mathcal{J}}$ and $s \in \overline{\mathcal{I}}$, it follows from part (a) that

$$u'_k(\Gamma_k) - l^s(\gamma^s) - p^s = 0.$$

This implies that $\tilde{x}$ satisfies the first order necessary and sufficient optimality conditions, and therefore is a WE at the price vector p. **Q.E.D.**

The second part of this lemma exploits the fact that the allocation of the unique WE flow rates and link loads across individuals is indeterminate, and shows that starting from any WE, we can always construct an alternative WE, with the same individual flow rates and link loads, in which all individual flows to all links with minimum effective cost are positive.

3. PRICE DETERMINATION BY PROFIT MAXIMIZATION

In Section 2, we characterized the flow rates and the link loads at a WE for a given price vector. In this section, we show the existence of profit-maximizing prices for the monopolist service provider and provide a characterization of these prices. The monopolist sets the prices as the optimal solution of the problem, [1]

$$\begin{aligned} \text{maximize } & \sum_{i \in \mathcal{I}} p^i \gamma^i(p) \\ \text{subject to } & p \geq 0, \end{aligned} \tag{3.1}$$

[1] This formulation ignores the costs incurred by the monopolist in data transmission. It is straightforward to add a constant for marginal cost of transmission without affecting the results.

where $\gamma^i(p)$ is the load of link i at a WE given price vector p.

It can be seen that, under Assumption 2.1, problem (3.1) has an optimal solution, denoted by p^*. We will refer to p^* as the *monopoly equilibrium price*. Let $x(p^*)$ be a WE given price vector p^*. Hence, $\big(p^*, x(p^*)\big)$ is a subgame perfect equilibirium of the two-stage game, where the monopolist service provider (Stackelberg leader) sets prices and each price defines a subgame among the users. We will refer to $\big(p^*, x(p^*)\big)$, or more simply to $\big(p^*, x^*\big)$, as the *monopoly equilibrium* (ME) of the overall game.

Proposition 3.5: Let Assumption 2.1 hold. Let p be a monopoly equilibrium price vector. Then

$$p^i > 0, \qquad \forall\ i \in \mathcal{I}.$$

Proof: Assume, to arrive at a contradiction, that $p^s = 0$ for some s. We first show that this implies that $\gamma^s > 0$. Assume the contrary, i.e., $\gamma^s = 0$. This implies by the first order optimality conditions and Assumption 2.1 (u_j is nondecreasing and $l^i(0) = 0$), that $u_j'(\Gamma_j) = 0$, for all j. But this implies that $\gamma^i = 0$ for all i, and therefore, $\Gamma_j = 0$ for all j, which is a contradiction and shows that we must have $\gamma^s > 0$.

We next consider the price vector $\tilde{p} = p + e_s$, where e_s is a vector whose components are all equal to 0, except the s^{th} component, which is equal to some $\epsilon > 0$. By Proposition 2.4, we have

$$\gamma^s(\tilde{p}) \le \gamma^s(p),$$

$$\gamma^i(\tilde{p}) \ge \gamma^i(p), \qquad \forall\ i \ne s.$$

We choose ϵ small enough such that $\gamma^s(\tilde{p}) > 0$. [This can be done in view of the continuity of γ^i in p, cf. Proposition 2.2]. Hence, we have

$$\sum_{i\in\mathcal{I}} p^i\gamma^i(p) < \sum_{i\in\mathcal{I}} \tilde{p}^i\gamma^i(\tilde{p}),$$

contradicting the fact that (p, x) is an ME, and showing that $p^i > 0$ for all i. **Q.E.D.**

This proposition shows that the monopolist will charge positive prices for all links. This result is intuitive. A zero price for link i would hurt the monopolist in two ways; first, directly by not generating any revenues from link i, and second, indirectly, by reducing traffic and profits on other links, in view of our result in Proposition 2.4.

Proposition 3.6: Let Assumption 2.1 hold. Let (p, x) be an ME. Then, the corresponding link loads γ^i satisfy

$$\gamma^i > 0, \qquad \forall\, i \in \mathcal{I}.$$

Proof: Define $\overline{\mathcal{I}} = \{i \mid \gamma^i > 0\}$. Assume, to arrive at a contradiction, that $\gamma^s = 0$ for some s. Using Lemma 2.1 and Assumption 2.1(b'), this implies that

$$K = p^i + l^i(\gamma^i) \leq p^s + l^s(0) = p^s, \qquad \forall\, i \in \overline{\mathcal{I}},$$

for some $K > 0$. Since $l^i(\gamma^i) > 0$ for all $i \in \overline{\mathcal{I}}$, we can choose some $\epsilon > 0$ such that

$$p^i < p^s - \epsilon < p^i + l^i(\gamma^i), \qquad \forall\, i \in \overline{\mathcal{I}}. \tag{3.2}$$

We next consider the price vector $\tilde{p} = p - e_s$, where e_s is a vector whose components are all equal to 0, except the s^{th} component, which is equal to ϵ. It follows from Proposition 2.4 that if $\gamma^i(p) = 0$ for some $i \neq s$, then $\gamma^i(\tilde{p}) = 0$. By Proposition 2.3, we also have

$$\Gamma_j(\tilde{p}) \geq \Gamma_j(p), \qquad \forall\, j, \tag{3.3}$$

which, after summing over all j, yields

$$\sum_{m \in \mathcal{I}} \gamma^m(\tilde{p}) \geq \sum_{m \in \mathcal{I}} \gamma^m(p). \tag{3.4}$$

There are now two cases to consider:

Case 1: $\gamma^i(\tilde{p}) = \gamma^i(p)$, for all $i \in \overline{\mathcal{I}}$. In this case, we have that $\gamma^s(\tilde{p}) > 0$. To see this, note that, if $\gamma^s(\tilde{p}) = 0$, we would have by Lemma 2.1 that

$$p^s - \epsilon \geq p^i + l^i\big(\gamma^i(\tilde{p})\big), \qquad \forall\, i \in \overline{\mathcal{I}},$$

which contradicts Eq. (3.2). But this implies that

$$\sum_{m \in \mathcal{I}} p^m \gamma^m(p) < \sum_{m \in \mathcal{I}} \tilde{p}^m \gamma^m(\tilde{p})$$

thus contradicting the fact that (p, x) is an ME.

Case 2: If $\gamma^i(\tilde{p}) < \gamma^i(p)$ for some $i \in \overline{\mathcal{I}}$, then it follows from Eq. (3.4) that $\gamma^s(\tilde{p}) > 0$. Then the change in the profit can be written as

$$\begin{aligned}
\sum_{m \in \mathcal{I}} \tilde{p}^m \gamma^m(\tilde{p}) - \sum_{m \in \mathcal{I}} p^m \gamma^m(p) &= (p^s - \epsilon)\gamma^s(\tilde{p}) + \sum_{m \neq s} p^m \big(\gamma^m(\tilde{p}) - \gamma^m(p)\big) \\
&> (p^s - \epsilon)\gamma^s(\tilde{p}) + (p^s - \epsilon) \sum_{m \neq s} \big(\gamma^m(\tilde{p}) - \gamma^m(p)\big) \\
&= (p^s - \epsilon) \sum_{j} \big(\Gamma_j(\tilde{p}) - \Gamma_j(p)\big) \\
&\geq 0,
\end{aligned}$$

where the strict inequality follows from Eq. (3.2) and Proposition 2.4, and the last inequality follows from Eq. (3.3). But this again contradicts the fact that (p, x) is an ME.

This shows that for all i, we must have $\gamma^i > 0$, and completes the proof. **Q.E.D.**

This proposition establishes that the monopolist would not choose to have zero link load on any link. Intuitively, allowing some positive traffic on a link would provide positive profits on that link. The proof is somewhat more complicated than this observation, however, because it also creates the indirect effect of reducing the loads on other links, thus reducing the rest of the monopolist's profits. The proof amounts to showing that this indirect effect is always dominated by the direct effect.

Note also that Assumption 2.1(b'), which imposes that $l^i(0) = 0$ for all i is important for this result. If $l^i(0) > 0$ for some i, it would be possible for the monopolist to have zero link load on some link.

We are now ready to provide a characterization of equilibrium prices. Before doing so, however, it is convenient to establish a lemma which will be useful in deriving this characterization.

Lemma 3.3: Let Assumption 2.1 hold. Let (p, x) be an ME, and $\overline{\mathcal{J}} = \{j \in \mathcal{J} \mid \Gamma_j > 0\}$. Assume without loss of generality that $1 \in \overline{\mathcal{J}}$. Then $\left(p, \gamma, [\Gamma_j]_{j \in \overline{\mathcal{J}}}\right)$ is an optimal solution of the following problem,

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in \mathcal{I}} p^i \gamma^i \\ \text{subject to} \quad & u_1'(\Gamma_1) - l^i(\gamma^i) - p^i = 0, \qquad \forall\, i \in \mathcal{I}, \\ & u_j'(\Gamma_j) - l^1(\gamma^1) - p^1 = 0, \qquad \forall\, j \in \overline{\mathcal{J}} - \{1\}, \\ & \sum_{i \in \mathcal{I}} \gamma^i = \sum_{j \in \overline{\mathcal{J}}} \Gamma_j, \\ & \gamma^i \geq 0,\ \Gamma_j \geq 0, \qquad \forall\, i \in \mathcal{I},\ j \in \overline{\mathcal{J}}. \end{aligned} \tag{3.5}$$

Proof: Assume to arrive at a contradiction that there exists some vector $\left(\overline{p}, \overline{\gamma}, [\overline{\Gamma}_j]_{j \in \overline{\mathcal{J}}}\right)$ feasible for problem (3.5) such that

$$\sum_{i \in \mathcal{I}} \overline{p}^i \overline{\gamma}^i > \sum_{i \in \mathcal{I}} p^i \gamma^i. \tag{3.6}$$

We first show that this implies

$$\overline{\gamma}^i \leq \gamma^i(\overline{p}), \qquad \forall\, i \in \mathcal{I}, \tag{3.7}$$

where $\gamma^i(\overline{p})$ is the load of link i at a WE given $\overline{p}$. Suppose, to arrive at a contradiction, that the preceding relation does not hold for some s, i.e., $\overline{\gamma}^s > \gamma^s(\overline{p})$. Since $\left(\overline{p}, \overline{\gamma}, [\overline{\Gamma}_j]_{j \in \overline{\mathcal{J}}}\right)$ is a feasible solution for problem (3.5), it can be seen that we have

$$u_k'(\overline{\Gamma}_k) - l^s(\overline{\gamma}^s) - \overline{p}^s = 0, \qquad \forall\ k \in \overline{\mathcal{J}}.$$

We also have using the first order optimality conditions for a WE that

$$u_k'\big(\Gamma_k(\overline{p})\big) - l^s\big(\gamma^s(\overline{p})\big) - \overline{p}^s \le 0, \qquad \forall\ k \in \mathcal{J},$$

where $\Gamma_k(\overline{p})$ is the flow rate of user k at a WE given $\overline{p}$. Since $\overline{\gamma}^s > \gamma^s(\overline{p})$, we obtain

$$\Gamma_k(\overline{p}) > \overline{\Gamma}_k, \qquad \forall\ k \in \overline{\mathcal{J}}. \tag{3.8}$$

Moreover, we have from the feasibility of $\big(\overline{p}, \overline{\gamma}, [\overline{\Gamma}_j]_{j\in\overline{\mathcal{J}}}\big)$ that

$$l^i(\overline{\gamma}^i) + \overline{p}^i = l^s(\overline{\gamma}^s) + \overline{p}^s, \qquad \forall\ i \in \mathcal{I}.$$

It also follows from Lemma 2.1 that for all i with $\gamma^i(\overline{p}) > 0$, we have

$$l^i\big(\gamma^i(\overline{p})\big) + \overline{p}^i \le l^s\big(\gamma^s(\overline{p})\big) + \overline{p}^s.$$

Since $\overline{\gamma}^s > \gamma^s(\overline{p})$, the preceding relations imply that

$$\gamma^i(\overline{p}) < \overline{\gamma}^i, \qquad \forall\ i \text{ with } \gamma^i(\overline{p}) > 0,$$

and therefore that

$$\gamma^i(\overline{p}) \le \overline{\gamma}^i, \qquad \forall\ i \in \mathcal{I}.$$

Summing both sides of this relation over all $i \in \mathcal{I}$, we obtain

$$\sum_{k\in\overline{\mathcal{J}}} \overline{\Gamma}_k = \sum_{i\in\mathcal{I}} \overline{\gamma}^i \ge \sum_{i\in\mathcal{I}} \gamma^i(\overline{p}) = \sum_{k\in\mathcal{J}} \Gamma_k(\overline{p}) \ge \sum_{k\in\overline{\mathcal{J}}} \Gamma_k(\overline{p}) > \sum_{k\in\overline{\mathcal{J}}} \overline{\Gamma}_k,$$

which is a contradiction [the last strict inequality uses Eq. (3.8)]. This implies that Eq. (3.7) holds for all $i \in \mathcal{I}$. Combining Eqs. (3.6) and (3.7), we obtain

$$\sum_{i\in\mathcal{I}} \overline{p}^i \gamma^i(\overline{p}) > \sum_{i\in\mathcal{I}} p^i \gamma^i,$$

which contradicts the fact that (p, x) is an ME, hence proves the desired result. **Q.E.D.**

Using this lemma, we can prove the following result.

Proposition 3.7 (Equilibrium Price Characterization): Let Assumption 2.1 hold. Assume further that u_j is twice continuously differentiable for each j, and l^i is continuously differentiable for each i. Let (p, x) be an ME, and $\overline{\mathcal{J}} = \{j \mid \Gamma_j > 0\}$. Then, for all $i \in \mathcal{I}$, we have

$$p^i = (l^i)'(\gamma^i)\gamma^i + \frac{\sum_{m \in \mathcal{I}} \gamma^m}{-\sum_{j \in \overline{\mathcal{J}}} \frac{1}{u_j''(\Gamma_j)}}.$$

Proof: By Lemma 3.3, it follows that $\left(p, \gamma, [\Gamma_j]_{j \in \overline{\mathcal{J}}}\right)$ is an optimal solution of problem (3.5). To simplify the notation, let us denote $\Gamma = [\Gamma_j]_{j \in \overline{\mathcal{J}}}$. One can check that, in view of the assumptions that u_j is strictly concave for all j and l^i is strictly increasing for all i, the constraint gradients of this problem are linearly independent at (p, γ, Γ), i.e., (p, γ, Γ) is a *regular point* (for the proof, see the Appendix). Hence, problem (3.5) admits Lagrange multipliers (see [BNO03], Chapter 5). The Lagrangian function $L(p, \gamma, \Gamma)$ for this problem is obtained by assigning the multipliers λ^i, $i \in \mathcal{I}$, to the first set of equality constraints in problem (3.5), the multipliers μ_j , $j \in \overline{\mathcal{J}} - \{1\}$, to the second set of equality constraints, and μ_1 to the last equality constraint; i.e.,

$$L(p, x, \lambda) = \sum_{i \in \mathcal{I}} p^i \gamma^i + \sum_{i \in \mathcal{I}} \lambda^i \left[u_1'(\Gamma_1) - l^i(\gamma^i) - p^i\right] + \sum_{j \in \overline{\mathcal{J}} - \{1\}} \mu_j \left[u_j'(\Gamma_j) - l^1(\gamma^1) - p^1\right] + \mu_1 \Big[\sum_{i \in \mathcal{I}} \gamma^i - \sum_{j \in \overline{\mathcal{J}}} \Gamma_j\Big].$$

Using the first order necessary optimality conditions at the optimal solution (p, γ, Γ) for problem (3.5), together with Proposition 3.5 ($p^i > 0$ for all i), Proposition 3.6 ($\gamma^i > 0$ for all i), and the fact that $\Gamma_j > 0$ for all $j \in \overline{\mathcal{J}}$, we obtain

$$\gamma^1 = \lambda^1 + \sum_{j \in \overline{\mathcal{J}} - \{1\}} \mu_j, \tag{3.9}$$

$$\gamma^i = \lambda^i, \qquad \forall\ i \neq 1, \tag{3.10}$$

and

$$p^1 - \lambda^1 (l^1)'(\gamma^1) - \Big(\sum_{j \in \overline{\mathcal{J}} - \{1\}} \mu_j\Big)(l^1)'(\gamma^1) + \mu_1 = 0,$$

$$p^i - \lambda^i (l^i)'(\gamma^i) + \mu_1 = 0, \qquad \forall\ i \neq 1.$$

Using Eqs. (3.9) and (3.10), the preceding two relations can be rewritten as

$$p^i - (l^i)'(\gamma^i)\gamma^i + \mu_1 = 0, \qquad \forall\ i \in \mathcal{I}. \tag{3.11}$$

Taking the partial derivatives of $L(p, \gamma, \Gamma)$ with respect to Γ_j also yields

$$u_1''(\Gamma_1) \sum_{i \in \mathcal{I}} \lambda^i - \mu_1 = 0,$$

$$\mu_j u_j''(\Gamma_j) - \mu_1 = 0, \qquad \forall\ j \in \overline{\mathcal{J}} - \{1\}.$$

Adding Eqs. (3.9) and (3.10) over all i, and using the preceding, we obtain

$$\sum_{m \in \mathcal{I}} \gamma^m = \mu_1 \sum_{j \in \overline{\mathcal{J}}} \frac{1}{u_j''(\Gamma_j)}.$$

Substituting the previous relation for μ_1 in (3.11) yields the desired result. **Q.E.D.**

This proposition is the central result of this section. An important implication is that the profit-maximizing price for the monopolist service provider consists of two markups (above the marginal cost of the monopolist, which is equal to zero here):[1] the first is $(l^i)'(\gamma^i)\gamma^i$. We will see in the next section that this term essentially internalizes the congestion externality ignored by the users. The second term is a further markup, similar to the monopoly markup in the standard economic models. We will see below that this term causes a further reduction in the flow rates, and typically creates a distortion relative to the social optimum.

4. PERFORMANCE COMPARISON WITH THE SOCIAL OPTIMUM

A network planner with full information and centralized control of the system allocates the resources as the optimal solution of the following problem:

$$\begin{aligned} \text{maximize} \ & \sum_{j \in \mathcal{J}} u_j(\Gamma_j) - \sum_{i \in \mathcal{I}} l^i(\gamma^i)\gamma^i \\ \text{subject to} \ & \Gamma_j = \sum_{i \in \mathcal{I}} x_j^i, \qquad \forall\ j \in \mathcal{J} \\ & \gamma^i = \sum_{j \in \mathcal{J}} x_j^i, \qquad \forall\ i \in \mathcal{I} \\ & x_j^i \geq 0, \qquad \forall\ i \in \mathcal{I},\ j \in \mathcal{J}. \end{aligned} \tag{4.1}$$

We call this problem the *social problem* and the optimal solution of this problem (which exists under Assumption 2.1) the *social optimum.* Note that the social problem is defined in terms of link load and flow allocations, and the prices do not appear in the problem.

[1] The markup is defined as the difference between price and marginal cost.

Assuming that the objective function of the social problem is concave, an equivalent characterization of the social optimum, denoted by $\tilde{x}$ is given by the first order conditions,

$$u_j'(\tilde{\Gamma}_j) - l^i(\tilde{\gamma}^i) - (l^i)'(\tilde{\gamma}^i)\tilde{\gamma}^i \begin{cases} = 0 & \text{if } \tilde{x}_j^i > 0, \\ \leq 0 & \text{if } \tilde{x}_j^i = 0. \end{cases}$$

It can be seen from these conditions that if each user is charged $(l^i)'(\tilde{\gamma}^i)\tilde{\gamma}^i$, the allocation resulting from the corresponding WE will be the same as the allocation of the social optimum. This amount is called the *marginal congestion cost* (also referred to as a *Pigovian tax* after the analysis of Alfred Pigou, see [Pig20]). In a decentralized implementation, this price may be charged to the users to internalize the congestion externalities that they ignore.

From the ME price characterization given in Proposition 3.7, we see that the ME price is the sum of the marginal congestion cost and a further markup. Hence, even though the monopolist is a selfish agent maximizing profits, its pricing internalizes the congestion externalities. This is because the monopolist chooses to charge higher prices when the demand is more inelastic (i.e., when demand changes little in response to a change in price) and a higher marginal congestion cost corresponds to more inelastic demand (with higher congestion costs, demand is less responsive to price because of the countervailing effect of increased congestion).

Comparison

Proposition 4.8: Let Assumption 2.1 hold. Assume further that u_j is twice continuously differentiable for each j, l^i is continuously differentiable for each i, and the function $\tilde{l}^i$, defined by $\tilde{l}^i(x) = l^i(x)x$ for all $x \geq 0$, is convex for each i. Let $\overline{x}$ denote the WE at 0 prices, $\tilde{x}$ denote the social optimum, and x denote the ME (and $\overline{\Gamma}_j$, $\tilde{\Gamma}_j$, and Γ_j the corresponding flow rates of user j, respectively). For all $j \in \mathcal{J}$, we have,

$$\Gamma_j \leq \tilde{\Gamma}_j \leq \overline{\Gamma}_j.$$

Proof: See the Appendix.

This result shows that when the prices are equal to 0, the users generate too much flow relative to the social optimum since they ignore the congestion externalities. Monopoly pricing improves this behavior because of the presence of the $(l^i)'(\gamma^i)\gamma^i$ term in the price expression. At the same time, monopoly pricing may introduce potential distortion relative to the social optimum, because the further markup induces users to reduce their flows below their socially optimal levels. This comparison intuitively suggests that the presence of a monopolist should improve performance in a network relative to the allocation without prices when congestion externalities are important. The next example illustrates this by comparing the performance of monopoly pricing to the performance of WE at zero prices for different utility and latency functions.

Example 4.1

Consider a network with J identical users and a single link. Assume that the utility function of the users $u(x)$ and the latency function of the link $l(x)$ are given by

$$u(x) = x^a, \qquad \text{for some } 0 < a \leq 1,$$

$$l(x) = \frac{1}{J^b} x^b, \qquad \text{for some } b \geq 0.$$

By symmetry, it can be seen that the flow rate of each user is the same at any WE. Let x_W denote the flow rate of a single user at the WE with prices set equal to 0. Similarly let x_M denote the flow rate of a single user at the ME. Using the optimality conditions, we compute x_W and x_M as

$$x_W = a^{\frac{1}{1+b-a}},$$

$$x_M = \left(\frac{a^2}{1+b}\right)^{\frac{1}{1+b-a}}.$$

Let U_W and U_M denote the total utility that the system gets at the WE and ME, respectively, i.e.,

$$\begin{aligned} U_W &= Ju(x_W) - l(Jx_W)Jx_W \\ &= Ja^{\frac{a}{1+b-a}}(1-a), \end{aligned}$$

$$\begin{aligned} U_M &= Ju(x_M) - l(Jx_M)Jx_M \\ &= J\left(\frac{a^2}{1+b}\right)^{\frac{a}{1+b-a}}\left(1 - \frac{a^2}{1+b}\right). \end{aligned}$$

We denote the ratio of U_M to U_W by E,

$$E = \frac{U_M}{U_W} = \left(\frac{a}{1+b}\right)^{\frac{a}{1+b-a}} \frac{1 - \frac{a^2}{1+b}}{1-a}.$$

Figure 4.1 illustrates the values of E as a function of a for different values of b. As can be seen from the figure, at fixed values of a, increasing b improves the performance of the ME, i.e., the more convex the latency function of the link, the better performance we have under monopoly pricing. Moreover, at a fixed value of b, ME performs better at higher values of a as expected, i.e., the less concave the utility function is, the better performance we have in the ME.

This example illustrates an important point in the performance of ME. As our discussion of pricing illustrates, monopoly pricing internalizes the congestion externality, which is the source of inefficiency without prices. The more pronounced this externality, the more useful is monopoly

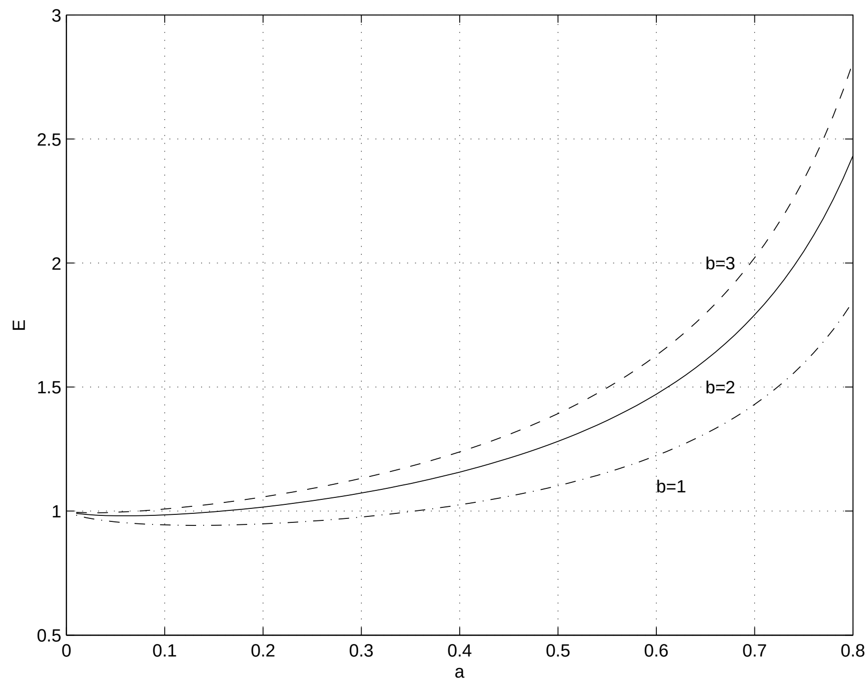

Figure 4.1: Performance of the ME relative to the performance of the WE with zero prices.

pricing. The example illustrates this by showing that the performance of monopoly pricing improves relative to WE with 0 prices as b increases, thus the latency function becomes more convex and the externalities become more important.

We illustrated in this section that monopoly pricing introduces a natural market mechanism to price congestion externalities. However, as can be seen from Proposition 3.7, monopoly pricing also introduces a distortion related to the markup term. It can be inferred from this price characterization that as the utility functions converge to linear functions, the markup term vanishes and the ME achieves the full information social optimum. In the next section, we analyze other utility functions for which the ME results in the socially optimal allocation.

5. ROUTING WITH PARTICIPATION CONTROL

In this section, we assume that the flow rate of each user is fixed, i.e., user j has t_j units of traffic. We are interested in analyzing the routing choices of individual users to send their traffic. In a system with the service provider setting the prices, the users should have the option of not sending any traffic if the prices are set too high. Otherwise, the service provider would charge infinite prices on the links. We refer to this problem as the *routing problem with participation control.* It can be modelled by using the following utility function u_j for user j,

$$u_j(x) = \begin{cases} 0 & \text{if } 0 \leq x < t_j, \\ t_j & \text{if } x \geq t_j, \end{cases} \tag{5.1}$$

together with binary participation decision variables z_j; i.e., $z_j = 1$ if user j chooses to send his t_j units of traffic and $z_j = 0$ if user j chooses not to send any traffic.

An equilibrium of this problem can be defined as follows.

Definition 5.2: For a given price vector $p \geq 0$, a vector $(x^*, z^*) = (x_j^*, z_j^*)_{j \in \mathcal{J}}$, where $x_j^* = \left[(x_j^1)^*, \ldots, (x_j^I)^*\right]$ is a nonnegative vector and $z_j^* \in \{0,1\}$, is a WE of the routing problem with participation control, if for all $j \in \mathcal{J}$,

$$(x_j^*, z_j^*) \in \arg \max_{x_j \geq 0,\ z_j \in \{0,1\}} \left\{ u_j(\Gamma_j z_j) - \sum_{i \in \mathcal{I}} \big(l^i(\gamma^i) + p^i\big) x_j^i \right\}, \tag{5.2}$$

where u_j is given by Eq. (5.1), and

$$\gamma^i = \sum_{j \in \mathcal{J}} x_j^{*i}, \qquad \forall\ i \in \mathcal{I}.$$

Using Eq. (5.1), we can rewrite condition (5.2) equivalently as

$$(x_j^*, z_j^*) \in \arg \max_{\substack{x_j \geq 0,\ z_j \in \{0,1\} \\ \sum_{i \in \mathcal{I}} x_j^i = t_j, \text{ if } z_j = 1}} \left\{ z_j t_j - \sum_{i \in \mathcal{I}} \big(l^i(\gamma^i) + p^i\big) x_j^i \right\}. \tag{5.3}$$

Note that, due to non-concavity of the utility function [see Eq. (5.1)], best response correspondences are not convex-valued, therefore one cannot use fixed point arguments to show the existence of a WE for a given p. In fact, it can be seen through simple examples that there does not exist a WE for every price vector p.

We next provide a different analysis of this case, establishing that there always exists a profit-maximizing price for the monopolist, and that at this price vector there exists a WE. More significantly, we show that in this case the ME always achieves the social optimum.

Using the linear structure of problem (5.3), we obtain the following characterization of a WE (proof omitted for brevity).

Lemma 5.4: For a given price vector $p \geq 0$, a vector $(x_j, z_j)_{j \in \mathcal{J}}$ is a WE if and only if it satisfies the following conditions:

(1) $\sum_{j \in \mathcal{J}} x_j^i = \gamma^i$ for all $i \in \mathcal{I}$.

(2) If $z_j = 1$, $\sum_{i \in \mathcal{I}} x_j^i = t_j$.

(3) If $z_j = 0$, $x_j^i = 0$ for all $i \in \mathcal{I}$.
Define the set $\overline{\mathcal{I}} = \Big\{ i \mid l^i(\gamma^i) + p^i \leq \min\big\{1, \min_{m \in \mathcal{I}} \{l^m(\gamma^m) + p^m\}\big\}\Big\}$.

(4) If $i \notin \overline{\mathcal{I}}$, then $\gamma^i = 0$.

(5) If $\min_{m \in \mathcal{I}}\{l^m(\gamma^m) + p^m\} < 1$, then $z_j = 1$ for all $j \in \mathcal{J}$.

We next show the following lemma related to the uniqueness of link loads at a WE of the routing problem, which is of independent interest.

Lemma 5.5 Assume that l^i is a continuous strictly increasing function. For a given $p \geq 0$, assume that a WE exists. Then, the resulting link loads, $\gamma = [\gamma^1, \ldots, \gamma^I]$, at any WE are unique.

Proof: Let γ and $\tilde{\gamma}$ be two distinct link load vectors at the WE (x, z) and $(\tilde{x}, \tilde{z})$, respectively. By Lemma 5.4 and condition (4), we have

$$p^i + l^i(\gamma^i) \begin{cases} = K & \text{if } i \in \overline{\mathcal{I}}, \\ > K & \text{if } i \notin \overline{\mathcal{I}}. \end{cases} \tag{5.4}$$

for some constant K, where $\overline{\mathcal{I}}$ is the set of links as defined in Lemma 5.4. We also have

$$p^i + l^i(\tilde{\gamma}^i) \begin{cases} = \tilde{K} & \text{if } i \in \tilde{\mathcal{I}}, \\ > \tilde{K} & \text{if } i \notin \tilde{\mathcal{I}}. \end{cases} \tag{5.5}$$

for some constant $\tilde{K}$, where $\tilde{\mathcal{I}}$ is the set of links as defined in Lemma 5.4, with γ^i replaced by $\tilde{\gamma}^i$.

Assume without loss of generality that there exists some s such that $\tilde{\gamma}^s > \gamma^s$. This implies that

$$\tilde{K} = p^s + l^s(\tilde{\gamma}^s) > p^s + l^s(\gamma^s) \geq K,$$

from which we get $\tilde{K} > K$. This implies that $\tilde{\gamma}^i \geq \gamma^i$ for all $i \in \mathcal{I}$. To see this, assume that $\tilde{\gamma}^i < \gamma^i$ for some i. We see from (5.4) that

$$p^i = K - l^i(\gamma^i),$$

and from (5.5) and $\tilde{K} > K$ that

$$p^i \geq \tilde{K} - l^i(\tilde{\gamma}^i) > K - l^i(\gamma^i),$$

yielding a contradiction. Hence, we have

$$\sum_{i \in \mathcal{I}} \tilde{\gamma}^i > \sum_{i \in \mathcal{I}} \gamma^i. \tag{5.6}$$

Since $\tilde{K} > K$, we also have $\tilde{z}_j \leq z_j$ for all $j \in \mathcal{J}$, from which we get

$$\sum_{i\in\mathcal{I}} \tilde{\gamma}^i = \sum_{j\in\mathcal{J}} \tilde{z}_j t_j \leq \sum_{j\in\mathcal{J}} z_j t_j = \sum_{i\in\mathcal{I}} \gamma^i,$$

contradicting Eq. (5.6), and showing that the link loads at a WE for a given p are unique. **Q.E.D.**

Consider next the monopoly problem for the routing problem with participation constraint,

$$\begin{aligned} \text{maximize} \ & \sum_{i\in\mathcal{I}} p^i \gamma^i \\ \text{subject to} \ & \gamma^i = \sum_{j\in\mathcal{J}} x_j^i, \qquad \forall\, i \in \mathcal{I}, \\ & (x,z) \in G(p),\ p \geq 0, \end{aligned} \tag{5.7}$$

where $G(p)$ denotes the set of (x, z) that satisfies conditions (1)-(5) stated in Lemma 5.4. We call the optimal solution of problem (5.7) $\big(p,(x,z)\big)$ the *monopoly equilibrium* (ME) and p as the *monopoly equilibrium price.*

Lemma 5.6: Let $\big(p,(x,z)\big)$ be an ME. Let $\gamma^i = \sum_{j\in\mathcal{J}} x_j^i$. Then for all i with $\gamma^i > 0$, we have

$$p^i = 1 - l^i(\gamma^i).$$

Proof: Since $\gamma^i > 0$, it follows by condition (4) in Lemma 5.4 that $i \in \overline{\mathcal{I}}$, which implies that

$$p^i + l^i(\gamma^i) \leq 1.$$

We next show that $p^i + l^i(\gamma^i) = 1$. Assume, to arrive at a contradiction, that $p^i + l^i(\gamma^i) < 1$. This implies that

$$p^k + l^k(\gamma^k) < \min\big\{1, \min_{m\notin\overline{\mathcal{I}}} p^m\big\}, \qquad \forall\, k \in \overline{\mathcal{I}}.$$

Hence, there exists some $\epsilon > 0$ such that

$$p^k + \epsilon + l^k(\gamma^k) < \min\big\{1, \min_{m\notin\overline{\mathcal{I}}} p^m\big\}, \qquad \forall\, k \in \overline{\mathcal{I}}.$$

Hence, (x, z) satisfies conditions (1)-(5) at the price vector $\tilde{p} = p + e_i$, where e_i is an I-dimensional vector, whose i^{th} component is ϵ if $i \in \overline{\mathcal{I}}$, and 0 otherwise. But this shows that $(x, z) \in G(\tilde{p})$, i.e., $\big(\tilde{p},(x,z)\big)$ is feasible for problem (5.7), and has strictly higher objective function value. This contradicts the fact the $\big(p,(x,z)\big)$ is an ME, showing that $p^i + l^i(\gamma^i) = 1$. **Q.E.D.**

Using the preceding proposition, we can rewrite problem (5.7) as

$$\begin{aligned}
\text{maximize}_{\ x_j^i \geq 0,\ z_j \in \{0,1\}} \ & \sum_{i \in \mathcal{I}} \big(1 - l^i(\gamma^i)\big)\gamma^i \\
\text{subject to} \ & \gamma^i = \sum_{j \in \mathcal{J}} x_j^i, \qquad \forall\ i \in \mathcal{I}, \\
& \sum_{i \in \mathcal{I}} x_j^i = t_j, \ \text{if } z_j = 1, \\
& x_j^i = 0, \ \forall\ i \in \mathcal{I}, \ \text{if } z_j = 0,
\end{aligned}$$

which can be equivalently written as

$$\begin{aligned}
\text{maximize}_{\ x_j^i \geq 0,\ z_j \in \{0,1\}} \ & \sum_{j \in \mathcal{J}} \left\{ z_j t_j - \sum_{i \in \mathcal{I}} l^i(\gamma^i) x_j^i \right\} \\
\text{subject to} \ & \gamma^i = \sum_{j \in \mathcal{J}} x_j^i, \qquad \forall\ i \in \mathcal{I}, \\
& \sum_{i \in \mathcal{I}} x_j^i = t_j, \ \text{if } z_j = 1.
\end{aligned} \tag{5.8}$$

This problem has an optimal solution; i.e., at each $z \in \{0,1\}^I$, there exists an optimal solution x (since the objective function is continuous and the constraint set is compact). This shows the existence of an ME of the routing problem with participation control, ensuring that at the profit maximizing price vector, there exists a WE.

Moreover, note that the social problem for the routing problem with participation control is the following optimization problem that maximizes the aggregate social surplus,

$$\begin{aligned}
\text{maximize}_{\ x_j^i \geq 0,\ z_j \in \{0,1\}} \ & \sum_{j \in \mathcal{J}} \left\{ z_j t_j - \sum_{i \in \mathcal{I}} l^i(\gamma^i) x_j^i \right\} \\
\text{subject to} \ & \gamma^i = \sum_{j \in \mathcal{J}} x_j^i, \qquad \forall\ i \in \mathcal{I}, \\
& \sum_{i \in \mathcal{I}} x_j^i = t_j, \ \text{if } z_j = 1,
\end{aligned}$$

[cf. Eq. (5.3)]. This problem is identical to problem (5.8). Hence, we establish the following result.

Proposition 5.9: A vector (x, z) is a social optimum if and only if there exists a price vector p such that $\big(p, (x, z)\big)$ is an ME.

Although the ME with pure routing achieves the same load as the social optimum for each link, it provides different levels of utilities to users than the choice of a fully-informed network planner who would implement the same allocation without charging users. In fact, in the ME, users pay a considerable amount to the monopolist (all the consumer surplus is taken by the monopolist).

In ongoing work, we study a similar model with multiple providers controlling various links, and show that, under certain circumstances, the corresponding Oligopoly Equilibrium also achieves the social optimum, but in addition, transfers the consumer surplus to the users.

6. CONCLUSIONS

A central objective of data network analysis is to characterize and implement relatively efficient flows of data in large networks. Although much of the literature approaches this problem as a network optimization problem, the practice is often different. First, most networks lack centralized regulation: individual users often have control over their flow and routing decisions, while network planners typically lack detailed knowledge about the needs and preferences of users. Second, most networks are for-profit entities, whose objectives are not efficiency or user welfare, but profit maximization. While a recent literature has investigated implications of the unregulated nature of modern data networks, particularly emphasizing the potential inefficiencies that result from this feature, there has been no systematic attempt to incorporate the second feature into models of data networks.

This paper provides a systematic analysis of decentralized data networks with many users that control their own flow and routing decisions, and service provider(s) charging prices for bandwidth and data transmission to maximize their own profits. Such an analysis first necessitates a unified approach to flow control and routing in the presence of prices, and a characterization of the response of link loads to changes in prices. After developing such a unified approach, this paper provides a characterization of prices, flow rates and link loads in the presence of a profit-maximizing monopolist service provider.

The most important feature of the monopoly equilibrium is that, despite its profit-maximizing objective, the monopolist may induce an allocation that is socially optimal. In particular, in the absence of the monopoly pricing, flow and routing decisions are inefficient, because users ignore the congestion externality that they create on others by their data transmission. The monopolist recognizes that, in the presence of the congestion externality, higher prices do not diminish the attractiveness of its product by as much, because the reduction in data transmission reduces congestion. Consequently, monopoly price for each link consists of two terms; the first exactly internalizing the congestion externality, and the second a further markup to increase profits. We show that in some important special cases, such as a model with only routing decisions, the monopoly equilibrium achieves the social optimum that the network planner with full information and complete control over flow and routing decisions would have implemented.

7. APPENDIX

Proof of Proposition 3.7: We prove that the constraint gradients of problem (3.5) are linearly independent at (p, γ, Γ), where (p, x) is the ME, $\gamma^i = \sum_{j \in \overline{\mathcal{J}}} x_j^i$, for each $i \in \mathcal{I}$, and $\Gamma_j = \sum_{i \in \mathcal{I}} x_j^i$, for each $j \in \overline{\mathcal{J}}$. Let $\overline{J} = |\overline{\mathcal{J}}|$ and assume without loss of generality that $\overline{\mathcal{J}} = \{1, \ldots, \overline{J}\}$. Denote the set of $I + \overline{J}$ equality constraints of problem (3.5) by

$$f(p, \gamma, \Gamma) = 0.$$

We show that $\nabla_{(\gamma,\Gamma)} f(p, \gamma, \Gamma)$ is nonsingular under Assumption 2.1. If it were not, there would be some nonzero $I + \overline{J}$-dimensional vector $(y', t')'$ that belongs to the nullspace of $\nabla_{(\gamma,\Gamma)} f(p, \gamma, \Gamma)$, i.e.,

$$\nabla_{\gamma,\Gamma} f(p, \gamma, \Gamma)(y', t')' = 0.$$

Multiplying the preceding out, we obtain the following set of equations,

$$t_{\overline{J}} = (l^1)'(\gamma^1) \left(\sum_{k=1}^{\overline{J}-1} t_k + y_1 \right),$$

$$t_{\overline{J}} = (l^m)'(\gamma^m) y_m, \qquad \forall\, m = 2, \ldots, I,$$

$$t_{\overline{J}} = u_1''(\Gamma_1) \sum_{m=1}^{I} y_m,$$

$$t_{\overline{J}} = u_k''(\Gamma_k) t_{k-1}, \qquad \forall\, k = 2 \ldots, \overline{J}.$$

Combining the previous relations, we obtain

$$t_{\overline{J}} \left(\sum_{j \in \overline{\mathcal{J}}} \frac{1}{u_j''(\Gamma_j)} - \sum_{i=1}^{I} \frac{1}{(l_i)'(\gamma^i)} \right) = 0.$$

Since $u_j''(\Gamma_j) < 0$ for all j and $(l^i)'(\gamma^i) > 0$ for all i, the term in the paranthesis is negative, implying that $t_{\overline{J}} = 0$. From this, we get that $(y', t')' = 0$, hence proving that $\nabla_{(\gamma,\Gamma)} f(p, \gamma, \Gamma)$ is nonsingular.

Proof of Proposition 4.8: First we show that $\tilde{\Gamma}_j \leq \overline{\Gamma}_j$ for all $j \in \mathcal{J}$. Partition the set of users into two sets R and S as

$$R = \left\{ r \in \mathcal{J} \mid \tilde{\Gamma}_r > \overline{\Gamma}_r \right\},$$

$$S = \left\{ s \in \mathcal{J} \mid \tilde{\Gamma}_s \leq \overline{\Gamma}_s \right\}.$$

We show that the set R is empty. Assume to arrive at a contradiction that the set R is nonempty. Define a subset of links as

$$I_{act} = \left\{ i \in \mathcal{I} \mid \tilde{x}_j^i > 0, \text{ for some } j \in R \right\}.$$

We show that, for all $i \in I_{act}$, we have

$$\tilde{\gamma}^i < \overline{\gamma}^i, \tag{7.1}$$

and

$$\overline{x}_s^i = 0, \qquad \forall\ s \in S. \tag{7.2}$$

Let $i \in I_{act}$. This implies that $\tilde{x}_j^i > 0$ for some $j \in R$. The first order optimality conditions for the social optimum and for the WE at 0 prices imply that

$$\begin{aligned} u_j'(\tilde{\Gamma}_j) - l^i(\tilde{\gamma}^i) - (l^i)'(\tilde{\gamma}^i)\tilde{\gamma}^i &\leq 0, \qquad \text{if } \tilde{x}_j^i \geq 0, \\ &= 0, \qquad \text{if } \tilde{x}_j^i > 0, \end{aligned} \tag{7.3}$$

and

$$\begin{aligned} u_j'(\overline{\Gamma}_j) - l^i(\overline{\gamma}^i) &\leq 0, \qquad \text{if } \overline{x}_j^i \geq 0, \\ &= 0, \qquad \text{if } \overline{x}_j^i > 0. \end{aligned} \tag{7.4}$$

Using the preceding, we obtain

$$u_j'(\tilde{\Gamma}_j) - l^i(\tilde{\gamma}^i) - (l^i)'(\tilde{\gamma}^i)\tilde{\gamma}^i \geq u_j'(\overline{\Gamma}_j) - l^i(\overline{\gamma}^i).$$

Since $j \in R$ and u_j is strictly concave, we have $u_j'(\tilde{\Gamma}_j) < u_j'(\overline{\Gamma}_j)$, which implies that

$$l^i(\tilde{\gamma}^i) + (l^i)'(\tilde{\gamma}^i)\tilde{\gamma}^i < l^i(\overline{\gamma}^i). \tag{7.5}$$

Since l^i is strictly increasing, it follows that

$$\tilde{\gamma}^i < \overline{\gamma}^i,$$

hence proving claim (7.1). To show (7.2), suppose to arrive at a contradiction, that $\overline{x}_s^i > 0$ for some $s \in S$. This implies by the first order optimality conditions [cf. Eqs. (7.3) and (7.4)] that

$$u_s'(\overline{\Gamma}_s) - l^i(\overline{\gamma}^i) \geq u_s'(\tilde{\Gamma}_s) - l^i(\tilde{\gamma}^i) - (l^i)'(\tilde{\gamma}^i)\tilde{\gamma}^i.$$

Since $s \in S$ and u_j is concave, we have $u_s'(\overline{\Gamma}_s) \leq u_s'(\tilde{\Gamma}_s)$. Combining this with the preceding equation and Eq. (7.5), we obtain

$$\begin{aligned} l^i(\overline{\gamma}^i) &\leq l^i(\tilde{\gamma}^i) + (l^i)'(\tilde{\gamma}^i)\tilde{\gamma}^i \\ &< l^i(\overline{\gamma}^i) \end{aligned},$$

thus yielding a contradiction and showing (7.2).

We next use Eqs. (7.1) and (7.2) to obtain

$$\sum_{i\in I_{act}} \overline{\gamma}_i = \sum_{i\in I_{act}} \sum_{j=1}^{J} \overline{x}_j^i = \sum_{i\in I_{act}} \sum_{j\in R} \overline{x}_j^i \leq \sum_{j\in R} \sum_{i=1}^{I} \overline{x}_j^i = \sum_{j\in R} \overline{\Gamma}_j.$$

We also have

$$\sum_{i\in I_{act}} \tilde{\gamma}_i \geq \sum_{i\in I_{act}} \sum_{j\in R} \tilde{x}_j^i = \sum_{j\in R} \sum_{i=1}^{I} \tilde{x}_j^i = \sum_{j\in R} \tilde{\Gamma}_j.$$

The preceding sets of equations together with the definition of set R imply that

$$\sum_{i\in I_{act}} \tilde{\gamma}^i \geq \sum_{i\in I_{act}} \overline{\gamma}^i.$$

Summing Eq. (7.1) over all $i \in I_{act}$ yields a contradiction, thus proving that the set R is empty.

We next show that $\Gamma_j \leq \tilde{\Gamma}_j$ for all $j \in \mathcal{J}$. Partition the set of users into two sets R and S as

$$R = \{r \in \mathcal{J} \mid \Gamma_r > \tilde{\Gamma}_r\},$$

$$S = \{s \in \mathcal{J} \mid \Gamma_s \leq \tilde{\Gamma}_s\}.$$

We show that the set R is empty. Assume to arrive at a contradiction that the set R is nonempty. Define a subset of links as

$$I_{act} = \{i \mid x_j^i > 0, \text{ for some } j \in R\}.$$

We show that, for all $i \in I_{act}$, we have

$$\gamma^i < \tilde{\gamma}^i, \tag{7.6}$$

and

$$\tilde{x}_s^i = 0, \qquad \forall\, s \in S. \tag{7.7}$$

Let $i \in I_{act}$. This implies that $x_j^i > 0$ for some $j \in R$. The first order optimality conditions for the ME imply that

$$\begin{aligned} u_j'(\Gamma_j) - l^i(\gamma^i) - (l^i)'(\gamma^i)\gamma^i + \frac{\sum_{m=1}^{I} \gamma^m}{\sum_{\{j\in \overline{J}\}} \frac{1}{u_j''(\Gamma_j)}} &\leq 0, \qquad \text{if } x_j^i \geq 0, \\ &= 0, \qquad \text{if } x_j^i > 0, \end{aligned} \tag{7.8}$$

[cf. Proposition 3.7]. Together with Eq. (7.3), we obtain

$$u_j'(\Gamma_j) - \overline{l}^i(\gamma^i) + \frac{\sum_{m=1}^{I} \gamma^m}{\sum_{\{j\in \overline{J}\}} \frac{1}{u_j''(\Gamma_j)}} \geq u_j'(\tilde{\Gamma}_j) - \overline{l}^i(\tilde{\gamma}^i),$$

where

$$\bar{l}^i(x) = l^i(x) + (l^i)'(x)x,$$
$$= \big(l^i(x)x\big)'.$$

Since $j \in R$ and u_j is strictly concave, we have $u_j'(\Gamma_j) < u_j'(\tilde{\Gamma}_j)$, which implies that

$$\bar{l}^i(\gamma^i) - \frac{\sum_{m=1}^I \gamma^m}{\sum_{\{j\in \overline{J}\}} \frac{1}{u_j''(\Gamma_j)}} < \bar{l}^i(\tilde{\gamma}^i). \tag{7.9}$$

Since $\bar{l}^i$ is nondecreasing [in view of the convexity of $l^i(x)x$] and u_k is strictly concave for all k, this implies that

$$\gamma^i < \tilde{\gamma}^i,$$

hence proving claim (7.6). To show (7.7), assume to arrive at a contradiction, that $\tilde{x}_s^i > 0$ for some $s \in S$. This implies by the first order optimality conditions [cf. Eqs. (7.3) and (7.8)] that

$$u_s'(\tilde{\Gamma}_s) - \bar{l}^i(\tilde{\gamma}^i) \geq u_s'(\Gamma_s) - \bar{l}^i(\gamma^i) + \frac{\sum_{m=1}^I \gamma^m}{\sum_{\{j\in \overline{J}\}} \frac{1}{u_j''(\Gamma_j)}}.$$

Since $s \in S$ and u_j is concave, we have

$$u_s'(\Gamma_s) \geq u_s'(\tilde{\Gamma}_s),$$

which together with Eq. (7.9) implies that

$$\bar{l}^i(\tilde{\gamma}^i) \leq \bar{l}^i(\gamma^i) - \frac{\sum_{m=1}^I \gamma^m}{\sum_{\{j\in \overline{J}\}} \frac{1}{u_j''(\Gamma_j)}},$$
$$< \bar{l}^i(\tilde{\gamma}^i)$$

thus yielding a contradiction and showing (7.7).

The rest is a similar argument as before and proves that the set R is empty. **Q.E.D.**

8. ACKNOWLEDGEMENTS

The authors would like to thank seminar participants at MIT Operations Research Center, University of Illinois at Urbana-Champaign, and Caltech, and in particular, Tamer Basar, Ramesh Johari, Tim Roughgarden, and Ivan Werning for useful comments and helpful discussions.

REFERENCES

[ABS02] Altman, E., Basar, T., and Srikant, R., "Nash Equilibria for combined flow control and routing in networks: asymptotic behavior for a large number of users," *IEEE Transactions on Automatic Control*, 2002.

[AcO04] Acemoglu, D. and Ozdaglar, A., "Flow Control, routing and performance from a service provider viewpoint," *LIDS report*, WP-1696, January 2004.

[ArH71] Arrow, K. and Hahn, F., General Competitive Analysis, Holden-Day, San Francisco, 1971.

[BaS02] Basar, T. and Srikant, R., "Revenue-maximizing pricing and capacity expansion in a many-users regime," *IEEE INFOCOM*, 2002.

[BMW56] Beckmann, M., Mcguire, C. B., and Winsten, C. B., Studies in the Economics of Transportation. Yale University Press, 1956.

[BNO03] Bertsekas, D. P., Nedic, A., and Ozdaglar, A. E., Convex Analysis and Optimization, Athena Scientific, Belmont, MA, 2003.

[CDR03] Cole, R., Dodis, Y., and Roughgarden, T., "How much can taxes help selfish routing?," *Proceedings of the 35th Annual ACM Symposium on the Theory of Computing*, 2003.

[CSS03] Correa, J. R., Schulz, A. S., and Stier Moses, N., "Selfish Routing in Capacitated Networks," MIT Sloan School of Management, Working paper.

[DaN83] Dafermos, S. and Nagurney, A., "Sensitivity analysis for the asymmetric network equilibrium problem," *Mathematical Programming,* vol. 28, pp. 174-184, 1983.

[Deb59] Debreu, G., Theory of Value, Wiley, New York, 1959.

[HaM85] Haurie, A. and Marcotte, P., "On the relationship between Nash-Cournot and Wardrop Equilibria," *Networks*, 15, 295-308, 1985.

[HeW03] He, L. and Walrand, J., "Pricing internet services with multiple providers," *Proc. of Allerton Conference*, 2003.

[JoT03] Johari, R. and Tsitsiklis, J., "Network resource Allocation and a congestion game," Working paper, 2003.

[KLO97] Korilis, Y. A., Lazar, A. A., and Orda, A.,"Achieving network optima using Stackelberg routing strategies," *IEEE/ACM Transactions on Networking*, vol. 5(1, pp. 161-173, 1997.)

[KMT98] Kelly, F. P., Maulloo A. K., and Tan D. K.,“Rate control for communication networks: shadow prices, proportional fairness, and stability,” *Journal of the Operational Research Society*, vol. 49, pp. 237-252, 1998.

[Kel91] Kelly, F. P., “ Network routing,” *Phil. Trans. R. Soc. Lond.*, vol. 337, pp. 343-367, 1991.

[Kel97] Kelly, F. P., “Charging and rate control for elastic traffic,” *European Transactions on Telecommunications*, vol. 8, pp. 33-37, 1997.

[LoL99] Low, S. and Lapsley, D. E., “Optimization flow control, I: Basic algorithm and convergence,” *IEEE/ACM Transactions on Networking*, vol. 7(6), pp. 861-874, 1999.

[MWG95] Mascolell, A., Whinston, M. D., and Green, J. R., Microeconomic Theory, Oxford, United Kingdom: Oxford University Press, 1995.

[MRW01] Mitra, D., Ramakrishnan K. G., and Wang, Q., “Combined Economic Modeling and Traffic Engineering: Joint Optimization of Pricing and Routing in Multi-Service Networks,” *Proc. 17th International Teletraffic Congress*, Salvador (Brazil), Elsevier, Amsterdam, pp. 73-85, 2001.

[ORS93] Orda, A., Rom, R., and Shimkin, N., “Competitive routing in multi-user communication networks,” *IEEE/ACM Transactions on Networking*, vol. 1, pp. 510-521, 1993.

[Pig20] Pigou, A. C., *The Economics of Welfare.* Macmillan, 1920.

[RoT02] Roughgarden, T. and Tardos, E., “ How bad is selfish routing?” *Journal of the ACM*, vol. 49, no. 2, pp 236-259, 2002.

[Ros65] Rosen, J. B., “Existence and uniqueness of equilibrium points for concave N-person games,” *Econometrica* 33, 3, pp. 520-534, 1965.

[ScS03] Schulz, A. S. and Stier Moses, N., “On the performance of user equilibria in traffic networks,” MIT Sloan School of Management, Working paper 4274-02, 2002.

[SCEH96] Shenker, S., Clark, D., Estrin, D., and Herzog, S., “Pricing in computer networks: reshaping the research agenda,” *Telecommunications Policy*, vol. 20, no. 3, pp. 183-201, 1996.

[She95] Shenker, S., “Fundamental design issues for the future Internet,” *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1176-1188, 1995.

[Tir90] Tirole, J., The Theory of Industrial Organization, MIT Press, Cambridge, 1990.

[War52] Wardrop, J. G., “Some theoretical aspects of road traffic research,” *Proceedings of the Institute of Civil Engineers*, Pt. II, vol. 1, pp. 325-378.