**MATHEMATICS OF OPERATIONS RESEARCH**
Vol. 32, No. 1, February 2007, pp. 1–31
ISSN 0364-765X | EISSN 1526-5471 | 07 | 3201 | 0001

DOI 10.1287/moor.1060.0231


# Competition and Efficiency in Congested Markets

Daron Acemoglu
Department of Economics, Massachusetts Institute of Technology, E52-380B, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, daron@mit.edu, http://econ-www.mit.edu/faculty/?prof_id=acemoglu

Asuman Ozdaglar
Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 32-D630, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, asuman@mit.edu, http://web.mit.edu/asuman/www

We study the efficiency of oligopoly equilibria (OE) in congested markets. The motivating examples are the allocation of network flows in a communication network or of traffic in a transportation network. We show that increasing competition among oligopolists can reduce efficiency, measured as the difference between users' willingness to pay and delay costs. We characterize a tight bound of $5/6$ on efficiency in pure strategy equilibria when there is zero latency at zero flow and a tight bound of $2\sqrt{2}-2$ with positive latency at zero flow. These bounds are tight even when the numbers of routes and oligopolists are arbitrarily large.

*Key words*: pricing; competition; congestion externalities; Wardrop equilibrium; social optimum; oligopoly equilibrium; efficiency; price of anarchy
*MSC2000 subject classification*: Primary: 90B10, 90B20, 91A40; secondary: 91A10, 91A18, 91A43
*OR/MS subject classification*: Primary: networks/graphs, theory; secondary: games, noncooperative, transportation, models
*History*: Received February 18, 2005; revised August 6, 2005.

**1. Introduction.** We analyze price competition in the presence of congestion costs. Consider the following environment: one unit of traffic can use one of $I$ alternative routes. More traffic on a particular route causes delays, exerting a negative (congestion) externality on existing traffic.[1] Congestion costs are captured by a route-specific nondecreasing convex latency function, $l_i(\cdot)$. Profit-maximizing oligopolists set prices (tolls) for travel on each route denoted by $p_i$. We analyze subgame-perfect Nash equilibria of this environment, where for each price vector, $p$, all traffic chooses the path that has minimum (delay plus toll) cost, $l_i + p_i$, and oligopolists choose prices to maximize profits.

The environment we analyze is of practical importance for a number of settings. These include transportation and communication networks, where additional use of a route (path) generates greater congestion for all users, and markets in which there are *snob* effects, so that goods consumed by fewer other consumers are more valuable (see, for example, Veblen [52]). The key feature of these environments is the negative congestion externality that users exert on others. This externality has been well recognized since the work by Pigou [40] in economics, by Samuelson [45], Wardrop [56], and Beckmann et al. [5] in transportation networks, and by Orda et al. [36], Korilis et al. [24], Kelly et al. [23], and Low and Lapsley [30] in communication networks. More recently, there has been a growing literature that focuses on quantification of efficiency loss (referred to as the *price of anarchy*) that results from externalities and strategic behavior in different classes of problems: selfish routing (e.g., Koutsoupias and Papadimitriou [25], Roughgarden and Tardos [44], Correa et al. [10, 11], Perakis [39], and Friedman [15]); resource allocation by market mechanisms (e.g., Johari and Tsitsiklis [22], Sanghavi and Hajek [46], Maheswaran and Basar [31], Young and Hajek [58]); network design (e.g., Anshelevich et al. [3]); and two-stage competitive facility location without congestion costs and externalities (e.g., Vetta [53]). Nevertheless, the game-theoretic interactions between (multiple) service providers and users, or the effects of competition among the providers on the efficiency loss has not been considered in networks with congestion (externalities). This is an important area for analysis because in most networks congestion is a first-order issue and (competing) profit-maximizing entities charge prices for use. Moreover, we will show that the nature of the analysis changes significantly in the presence of price competition.

We provide a general framework for the analysis of price competition among service providers[2] in a congested (and potentially capacitated) network, study existence of pure strategy and mixed-strategy equilibria, and characterize and quantify the efficiency properties of equilibria. There are four sets of major results from our analysis.

First, though the equilibrium of traffic assignment without prices can be highly inefficient (e.g., Pigou [40], Roughgarden and Tardos [44], Correa et al. [10]), price setting by a monopolist internalizes the negative externality and achieves efficiency.

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

[2] We use oligopolist and service provider interchangeably throughout the paper.

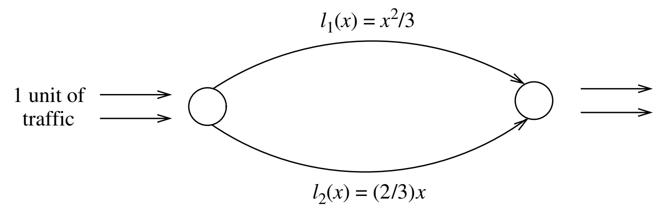


FIGURE 1. A two-link network with congestion-dependent latency functions.

Second, increasing competition can increase inefficiency. In fact, changing the market structure from monopoly to duopoly almost always increases inefficiency. This result contrasts with most existing results in the economics literature where greater competition tends to improve the allocation of resources (e.g., see Tirole [50]). The intuition for this result, which is related to congestion, is illustrated by the example we discuss below.[3]

Third and most important, we provide tight bounds on the extent of inefficiency in the presence of oligopolistic competition. We show that when latency at zero flow (traffic) is equal to zero, social surplus (defined as the difference between users' willingness to pay and the delay cost) in any pure strategy oligopoly equilibrium (OE) is always greater than $5/6$ of the maximum social surplus. When latency at zero flow can be positive, there is a slightly lower bound of $2\sqrt{2}-2 \approx 0.828$. These bounds are independent of both the number of routes, $I$, which could be arbitrarily large, and how these routes are distributed across different oligopolists (i.e., of market structure). Simple examples reach these bounds.

Finally, we also show that pure strategy equilibria may fail to exist. This is not surprising in view of the fact that what we have here is a version of a Bertrand-Edgeworth game where pure strategy equilibria do not exist in the presence of convex costs of production or capacity constraints (e.g., Edgeworth [14], Shubik [48], Benassy [6], Vives [55]). However, in our oligopoly environment when latency functions are linear, a pure strategy equilibrium always exists, essentially because congestion externalities remove the payoff discontinuities inherent in the Bertrand-Edgeworth game. Nonexistence becomes an issue when latency functions are highly convex. In this case, we prove that mixed-strategy equilibria always exist. We also show that mixed-strategy equilibria can lead to arbitrarily inefficient worst-case realizations; in particular, social surplus can become arbitrarily small relative to the maximum social surplus, though the average performance of mixed-strategy equilibria is much better.

The following example illustrates some of our results.

EXAMPLE 1.1. Figure 1 shows a situation similar to the one first analyzed by Pigou [40] to highlight the inefficiency because of congestion externalities. One unit of traffic will travel from origin A to destination B, using either route 1 or route 2. The latency functions are given by

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

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 $x_1^{WE} \approx 0.73 > x_1^S$ and $x_2^{WE} \approx 0.27 < x_2^S$. 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.

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, $x_i l_i'$ (when $l_i$ is differentiable), which exactly internalizes the congestion externality. In other words, this markup is equivalent to the Pigovian tax that a social planner would set to induce decentralized traffic to choose the efficient allocation. Consequently, in this simple example, monopoly prices will be $p_1^{ME} = (2/3)^3 + k$ and $p_2^{ME} = (2/3^2) + k$ for some constant $k$. The resulting traffic in the Wardrop equilibrium (WE) will be identical to the efficient allocation, i.e., $x_1^{ME} = 2/3$ and $x_2^{ME} = 1/3$.

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 $p_i^{OE} = x_i(l_1' + l_2')$ (see Equation (20) in §4), or more specifically, $p_1^{OE} \approx 0.61$ and $p_2^{OE} \approx 0.44$. The resulting equilibrium traffic is $x_1^{OE} \approx 0.58 < x_1^S$ and $x_2^{OE} \approx 0.42 > x_2^S$, which also differs from the efficient allocation. We will show that this is generally the

[3] Because, in our model, users are homogeneous and have a constant reservation utility, in the absence of congestion externalities, all market structures would achieve efficiency, and a change from monopoly to duopoly, for example, would have no efficiency consequence.

case in the OE. Interestingly, while in the WE without prices, there was too much traffic on route 1, now there is too little traffic because of its greater markup. It is also 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 OE users may have positive consumer surplus.[4]

The intuition for the inefficiency of duopoly relative to 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 the route 2, has become worse.[5] As a result, the optimal price that each duopolist charges will include an additional markup over the Pigovian markup. These are $x_1 l_2'$ for route 1 and $x_2 l_1'$ for route 2. Because these two markups are generally different, they will distort the pattern of traffic away from the efficient allocation. Naturally, however, prices are typically lower with duopoly, so even though social surplus declines, users will be better off than in monopoly (i.e., they will command a positive consumer surplus).

There is a large literature on models of congestion both in transportation and communication networks (e.g., Beckmann et al. [5], Patriksson [38], Rosenthal [43], Milchtaich [33, 34], Roughgarden and Tardos [44]).[6] However, very few studies have investigated the implications of having the *property rights* over routes assigned to profit-maximizing providers. In Basar and Srikant [4], they analyze monopoly pricing under specific assumptions on the utility and latency functions. He and Walrand [19] study competition and cooperation among Internet service providers under specific demand models. Issues of efficient allocation of flows or traffic across routes do not arise in these papers. Our previous work (Acemoglu and Ozdaglar [1]) studies the monopoly problem and contains the efficiency of the monopoly result, but none of the other results here. More recent independent work by Hayrapetyan et al. [18] builds on Acemoglu and Ozdaglar [1] and also studies competition among service providers. Using a different mathematical approach, they provide nontight bounds on the efficiency loss for the case of elastic traffic. Finally, in current work, (Acemoglu and Ozdaglar [2]), we extend some of the results of this paper to a network with parallel-serial structure.

In the rest of this paper, we use the terminology of a (communication) network, though all of the analysis applies to resource allocation in transportation networks, electricity markets, and other economic applications. Section 2 describes the basic environment. Section 3 briefly characterizes the monopoly equilibrium (ME) and establishes its efficiency. Section 4 defines and characterizes the OE with competing profit-maximizing providers. Section 5 contains the main results and characterizes the efficiency properties of the OE and provide bounds on efficiency. Section 6 provides a tight efficiency bound when there may be positive latency at zero flow. Section 7 contains concluding comments.

Regarding notation, all vectors are viewed as column vectors, and inequalities are to be interpreted componentwise. We denote by $\mathbb{R}^I_+$ the set of nonnegative $I$-dimensional vectors. Let $C_i$ be a closed subset of $[0, \infty)$ and let $f\colon C_i \mapsto \mathbb{R}$ be a convex function. We use $\partial f(x)$ to denote the set of subgradients of $f$ at $x$, and $f^-(x)$ and $f^+(x)$ to denote the left and right derivatives of $f$ at $x$. For a function $f\colon \mathbb{R}^n \mapsto (-\infty, \infty]$, we say that $f$ is closed if the level set $\{x \mid f(x) \leq c\}$ is closed for every scalar $c$. Note that a function is closed if and only if it is lower semicontinuous over $\mathbb{R}^n$ (see Bertsekas et al. [9], Proposition 1.2.2).

**2. Model.** We consider a network with $I$ parallel links. Let $\mathcal{I} = \{1, \ldots, I\}$ denote the set of links. Let $x_i$ denote the total flow on link $i$, and $x = [x_1, \ldots, x_I]$ denote the vector of link flows. Each link in the network has a flow-dependent latency function $l_i(x_i)$, which measures the travel time (or delay) as a function of the total flow on link $i$. We denote the price per unit flow (bandwidth) of link $i$ by $p_i$. Let $p = [p_1, \ldots, p_I]$ 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 [56]) 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 (see the definition below).[7] We also assume

[4] 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 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 or profits). See Mascolell et al. [32].

[5] Using economics terminology, we could also say that the demand for route 1 becomes more *inelastic*. Because this term has a different meaning in the communication networks literature (see Shenker [47]), we do not use it here.

[6] Some of these papers also use prices (or tolls) to induce flow patterns that optimize overall system objective, and a number of studies have 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 Bergendorff et al. [8], Hearn and Ramana [20], Larsson and Patriksson [27, 28], and Hearn and Yildirim [21].

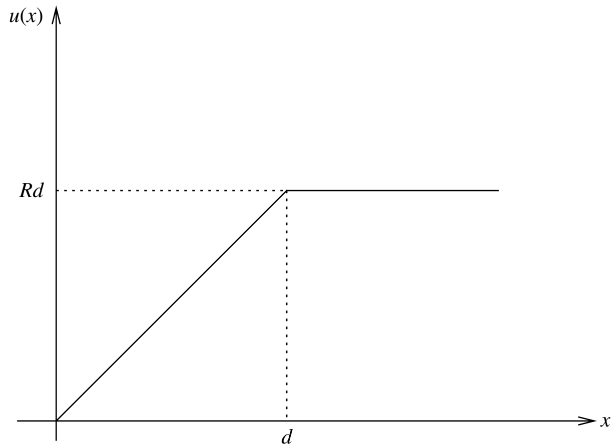


FIGURE 2. Aggregate utility function.

that the users have a *reservation utility* $R$ and decide not to send their flow if the effective cost exceeds the reservation utility. This implies that user preferences can be represented by the piecewise linear aggregate utility function $u(\cdot)$ depicted in Figure 2.[8]

To account for additional side constraints in the traffic equilibrium problem, including capacity constraints on the links, we use the following definition of a WE (see Larsson and Patriksson [26, 29]). Lemma 2.1 shows that this definition is equivalent to the more standard definition of a WE used in the literature under some assumptions.

DEFINITION 2.1. For a given price vector $p \geq 0$,[9] a vector $x^{WE} \in \mathbb{R}_+^I$ is a *Wardrop equilibrium* (WE) if

$$x^{WE} \in \underset{\substack{x \geq 0 \\ \sum_{i \in \mathcal{I}} x_i \leq d}}{\arg\max} \left\{ \sum_{i \in \mathcal{I}} (R - l_i(x_i^{WE}) - p_i) x_i \right\}. \tag{1}$$

We denote the set of WE at a given $p$ by $W(p)$.

ASSUMPTION 2.1. *For each $i \in \mathcal{I}$, the latency function $l_i$: $[0, \infty) \mapsto [0, \infty]$ is closed, convex, nondecreasing, and satisfies $l_i(0) = 0$.*

The assumption of zero latency at zero flow, i.e., $l_i(0) = 0$, implies that all latency is because of flow of traffic, and there are no fixed latency costs.[10] It is adopted to simplify the discussion, especially the characterization of equilibrium prices in Proposition 4.4 below. A trivial relaxation of this assumption to $l_i(0) = L$ for all $i \in \mathcal{I}$ for some $L > 0$ will have no effect on any of the results in this paper. Allowing for differential levels of $l_i(0)$ complicates the analysis, but has little effect on the major results. This case is discussed in §6, where we provide a slightly lower tight bound for the inefficiency of oligopoly equilibria without this assumption.

Another feature of Assumption 2.1 is that it allows latency functions to be extended real-valued, thus allowing for capacity constraints. Let $C_i = \{x \in [0, \infty) \mid l_i(x) < \infty\}$ denote the effective domain of $l_i$. By Assumption 2.1, $C_i$ is a closed interval of the form $[0, b]$ or $[0, \infty)$. Let $b_{C_i} = \sup_{x \in C_i} x$. Without loss of generality, we can add the constraint $x_i \in C_i$ in Equation (1). Using the optimality conditions for Problem (1), we see that a vector $x^{WE} \in \mathbb{R}_+^I$ is a WE if and only if $\sum_{i \in \mathcal{I}} x_i^{WE} \leq d$, and there exists some $\lambda \geq 0$ such that $\lambda(\sum_{i \in \mathcal{I}} x_i^{WE} - d) = 0$ and for all $i$,

$$\begin{aligned} R - l_i(x_i^{WE}) - p_i &\leq \lambda \quad \text{if } x_i^{WE} = 0, \\ &= \lambda \quad \text{if } 0 < x_i^{WE} < b_{C_i}, \\ &\geq \lambda \quad \text{if } x_i^{WE} = b_{C_i}. \end{aligned} \tag{2}$$

[7] Wardrop's principle is used extensively in modelling traffic behavior in transportation networks, e.g., Beckmann et al. [5], Dafermos and Sparrow [12], Patriksson [38], and Smith [49], and in communication networks, e.g., Roughgarden and Tardos [44] and Correa et al. [10].

[8] This simplifying assumption implies that all users are *homogeneous* in the sense that they have the same reservation utility, $R$. The analysis below will show that the value of this reservation utility $R$ has no effect on any of the results as long as it is strictly positive. We discuss potential issues in extending this work to users with elastic and heterogeneous requirements in the concluding section.

[9] Since the reservation utility of users is equal to $R$, we can also restrict attention to $p_i \leq R$ for all $i$. Throughout the paper, we use $p \geq 0$ and $p \in [0, R]^I$ interchangeably.

[10] This assumption is a good approximation to communication networks where queueing delays are more substantial than propagation delays.

When the latency functions are real-valued (i.e., $C_i = [0, \infty)$), we obtain the following characterization of a WE, which is often used as the definition of a WE in the literature. This lemma states that in the WE, *the effective costs*, defined as $l_i(x_i^{WE}) + p_i$, are equalized on all links with positive flows.

LEMMA 2.1. *Let Assumption 2.1 hold, and assume further that $C_i = [0, \infty)$ for all $i \in \mathcal{I}$. Then a nonnegative vector $x^* \in W(p)$ if and only if*

$$l_i(x_i^*) + p_i = \min_j \{l_j(x_j^*) + p_j\}, \quad \forall\, i \text{ with } x_i^* > 0, \tag{3}$$
$$l_i(x_i^*) + p_i \le R, \quad \forall\, i \text{ with } x_i^* > 0,$$
$$\sum_{i \in \mathcal{I}} x_i^* \le d,$$

*with $\sum_{i \in \mathcal{I}} x_i^* = d$ if $\min_j \{l_j(x_j) + p_j\} < R$.*

Example 2.1 below shows that Condition (3) in this lemma may not hold when the latency functions are not real-valued. The existence, uniqueness, and continuity properties of a WE are well studied (see Beckmann et al. [5], Dafermos and Sparrow [12], Smith [49]). We provide here the standard proof for existence, based on establishing the equivalence of WE and the optimal solutions of a convex optimization problem, which we will refer to later in our analysis.

PROPOSITION 2.1 (EXISTENCE AND CONTINUITY). *Let Assumption 2.1 hold. For any price vector $p \ge 0$, the set of WE, $W(p)$, is nonempty. Moreover, the correspondence $W: \mathbb{R}_+^I \rightrightarrows \mathbb{R}_+^I$ is upper semicontinuous.*

PROOF. Given any $p \ge 0$, consider the following optimization problem:

$$\begin{aligned} \underset{x \ge 0}{\text{maximize}} \quad & \sum_{i \in \mathcal{I}} \left( (R - p_i) x_i - \int_0^{x_i} l_i(z)\, dz \right) \\ \text{subject to} \quad & \sum_{i \in \mathcal{I}} x_i \le d. \\ & x_i \in C_i, \quad \forall\, i. \end{aligned} \tag{4}$$

In view of Assumption (2.1) (i.e., $l_i$ is nondecreasing for all $i$), it can be shown that the objective function of Problem (4) is convex over the constraint set, which is nonempty (since $0 \in C_i$) and convex. Moreover, the first-order optimality conditions of Problem (4), which are also sufficient conditions for optimality, are identical to the WE optimality conditions (cf. Equation (2)). Hence a flow vector $x^{WE} \in W(p)$ if and only if it is an optimal solution of Problem (4). Since the objective function of Problem (4) is continuous and the constraint set is compact, this problem has an optimal solution, showing that $W(p)$ is nonempty. The fact that $W$ is an upper semicontinuous correspondence at every $p$ follows by using the theorem of the maximum (see Berge [7], Chapter 6) for Problem (4). □

WE flows also satisfy intuitive monotonicity properties given in the following proposition. The proof follows from the optimality conditions (cf. Equation (2)) and is omitted (see Acemoglu and Ozdaglar [1]).

PROPOSITION 2.2 (MONOTONICITY). *Let Assumption 2.1 hold. For a given $p \ge 0$, let $p_{-j} = [p_i]_{i \ne j}$.*
(a) *For some $\bar{p} \le p$, let $\bar{x} \in W(\bar{p})$ and $x \in W(p)$. Then $\sum_{i \in \mathcal{I}} \bar{x}_i \ge \sum_{i \in \mathcal{I}} x_i$.*
(b) *For some $\bar{p}_j < p_j$, let $\bar{x} \in W(\bar{p}_j, p_{-j})$ and $x \in W(p_j, p_{-j})$. Then $\bar{x}_j \ge x_j$ and $\bar{x}_i \le x_i$ for all $i \ne j$.*
(c) *For some $\tilde{\mathcal{I}} \subset \mathcal{I}$, suppose that $\bar{p}_j < p_j$ for all $j \in \tilde{\mathcal{I}}$ and $\bar{p}_j = p_j$ for all $j \notin \tilde{\mathcal{I}}$, and let $\bar{x} \in W(\bar{p})$ and $x \in W(p)$. Then $\sum_{j \in \tilde{\mathcal{I}}} \bar{x}_j \ge \sum_{j \in \tilde{\mathcal{I}}} x_j$.*

For a given price vector $p$, the WE need not be unique in general. The following example illustrates some properties of the WE.

EXAMPLE 2.1. Consider a two-link network. Let the total flow be $d = 1$ and the reservation utility be $R = 1$. Assume that the latency functions are given by

$$l_1(x) = l_2(x) = \begin{cases} 0 & \text{if } 0 \le x \le \dfrac{2}{3} \\ \infty & \text{otherwise.} \end{cases}$$

At the price vector $(p_1, p_2) = (1, 1)$, the set of WE, $W(p)$, is given by the set of all vectors $(x_1, x_2)$ with $0 \le x_i \le 2/3$ and $\sum_i x_i \le 1$. At any price vector $(p_1, p_2)$ with $p_1 > p_2 = 1$, $W(p)$ is given by all $(0, x_2)$ with $0 \le x_2 \le 2/3$.

This example also illustrates that Lemma 1 need not hold when latency functions are not real-valued. Consider, for instance, the price vector $(p_1, p_2) = (1-\epsilon, 1-a\epsilon)$ for some scalar $a > 1$. In this case, the unique WE is $(x_1, x_2) = (1/3, 2/3)$, and clearly effective costs on the two routes are not equalized despite the fact that they both have positive flows. This arises because the path with the lower effective cost is capacity constrained, so no more traffic can use that path.

Under further restrictions on the $l_i$, the following standard result follows (proof omitted).

Proposition 2.3 (Uniqueness). *Let Assumption 2.1 hold. Assume further that $l_i$ is strictly increasing over $C_i$. For any price vector $p \geq 0$, the set of WE, $W(p)$, is a singleton. Moreover, the function $W: \mathbb{R}^I_+ \mapsto \mathbb{R}^I_+$ is continuous.*

Because we do not assume that the latency functions are strictly increasing, we need the following lemma in our analysis to deal with nonunique WE flows.

Lemma 2.2. *Let Assumption 2.1 hold. For a given $p \geq 0$, define the set*

$$\bar{\mathcal{I}} = \{i \in \mathcal{I} \mid \exists\, x, \hat{x} \in W(p) \text{ with } x_i \neq \hat{x}_i\}. \tag{5}$$

*Then*

$$l_i(x_i) = 0, \quad \forall\, i \in \bar{\mathcal{I}}, \ \ \forall\, x \in W(p),$$
$$p_i = p_j, \quad \forall\, i, j \in \bar{\mathcal{I}}.$$

Proof. Consider some $i \in \bar{\mathcal{I}}$ and $x \in W(p)$. Since $i \in \bar{\mathcal{I}}$, there exists some $\hat{x} \in W(p)$ such that $x_i \neq \hat{x}_i$. Assume without loss of generality that $x_i > \hat{x}_i$. There are two cases to consider:

(a) If $x_k \geq \hat{x}_k$ for all $k \neq i$, then $\sum_{j\in\mathcal{I}} x_j > \sum_{j\in\mathcal{I}} \hat{x}_j$, which implies that the WE optimality conditions (cf. Equation (2)) for $\hat{x}$ hold with $\hat{\lambda} = 0$. By Equation (2) and $x_i > \hat{x}_i$, we have

$$l_i(x_i) + p_i \leq R,$$
$$l_i(\hat{x}_i) + p_i \geq R,$$

which together imply that $l_i(x_i) = l_i(\hat{x}_i)$. By Assumption 2.1 (i.e., $l_i$ is convex and $l_i(0) = 0$), it follows that $l_i(x_i) = 0$.

(b) If $x_k < \hat{x}_k$ for some $k$, by the WE optimality conditions, we obtain

$$l_i(x_i) + p_i \leq l_k(x_k) + p_k,$$
$$l_i(\hat{x}_i) + p_i \geq l_k(\hat{x}_k) + p_k.$$

Combining the above with $x_i > \hat{x}_i$ and $x_k < \hat{x}_k$, we see that $l_i(x_i) = l_i(\hat{x}_i)$ and $l_k(x_k) = l_k(\hat{x}_k)$. By Assumption 2.1, this shows that $l_i(x_i) = 0$ (and also that $p_i = p_k$).

Next, consider some $i, j \in \bar{\mathcal{I}}$. We will show that $p_i = p_j$. Since $i \in \bar{\mathcal{I}}$, there exist $x, \hat{x} \in W(p)$ such that $x_i > \hat{x}_i$. There are three cases to consider:

- $x_j < \hat{x}_j$. Then a similar argument to part (b) above shows that $p_i = p_j$.
- $x_j > \hat{x}_j$. If $x_k \geq \hat{x}_k$ for all $k \neq i, j$, then $\sum_m \hat{x}_m < d$, implying that the WE optimality conditions hold with $\hat{\lambda} = 0$. Therefore we have

$$l_i(x_i) + p_i \leq R,$$
$$l_j(\hat{x}_j) + p_j \geq R,$$

which together with $l_i(x_i) = l_j(\hat{x}_j) = 0$ imply that $p_i = p_j$.

- $x_j = \hat{x}_j$. Since $j \in \bar{\mathcal{I}}$, by definition there must exist some other $\bar{x} \in W(p)$ such that $x_j \neq \bar{x}_j$. Repeating the above two steps with $\bar{x}_j$ instead of $\hat{x}_j$ yields the desired result. □

Intuitively, this lemma states that if there exist multiple WEs, $x, \hat{x} \in W(p)$ such that $x_i \neq \hat{x}_i$, then the latency function $l_i$ must be locally flat around $x_i$ (and $\hat{x}_i$). Given the assumption that $l_i(0) = 0$ and the convexity of latency functions, this immediately implies $l_i(x_i) = 0$.

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.

Definition 2.2. A flow vector $x^S$ is a *social optimum* if it is an optimal solution of the social problem

$$\begin{aligned} &\underset{x\geq 0}{\text{maximize}} \ \sum_{i\in\mathcal{I}}(R-l_i(x_i))x_i \\ &\text{subject to} \ \sum_{i\in\mathcal{I}} x_i \leq d. \end{aligned} \tag{6}$$

In view of Assumption 2.1, the social problem has a continuous objective function and a compact constraint set, guaranteeing the existence of a social optimum, $x^S$. Moreover, using the optimality conditions for a convex program (see Bertsekas et al. [9], §4.7), we see that a vector $x^S \in \mathbb{R}^I_+$ is a social optimum if and only if $\sum_{i\in\mathcal{I}} x_i^S \leq d$ and there exists a subgradient $g_{l_i} \in \partial l_i(x_i^S)$ for each $i$, and a $\lambda^S \geq 0$ such that $\lambda^S(\sum_{i\in\mathcal{I}} x_i^S - d) = 0$ and for each $i$,

$$\begin{aligned} R - l_i(x_i^S) - x_i^S g_{l_i} &\leq \lambda^S \quad \text{if } x_i^S = 0, \\ &= \lambda^S \quad \text{if } 0 < x_i^S < b_{C_i}, \\ &\geq \lambda^S \quad \text{if } x_i^S = b_{C_i}. \end{aligned} \tag{7}$$

For future reference, for a given vector $x \in \mathbb{R}^I_+$, we define the value of the objective function in the social problem,

$$\mathbb{S}(x) = \sum_{i\in\mathcal{I}}(R - l_i(x_i))x_i, \tag{8}$$

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

**3. Monopoly equilibrium and efficiency.** In this section, we assume that a monopolist service provider owns the $I$ links and charges a price of $p_i$ per unit bandwidth on link $i$. We considered a related problem in Acemoglu and Ozdaglar [1] for atomic users with inelastic traffic (i.e., the utility function of each of a finite set of users is a step function), and with increasing, real-valued and differentiable latency functions. Here we show that similar results hold for the more general latency functions and the demand model considered in §2.

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.

Definition 3.1. A vector $(p^{ME}, x^{ME}) \geq 0$ is a *monopoly equilibrium* (ME) if $x^{ME} \in W(p^{ME})$ and

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

Our definition of the ME is stronger than the standard subgame-perfect Nash equilibrium concept for dynamic games. With a slight abuse of terminology, let us associate a subgame-perfect Nash equilibrium with the on-the-equilibrium-path actions of the two-stage game.

Definition 3.2. A vector $(p^*, x^*) \geq 0$ is a subgame-perfect equilibrium (SPE) of the pricing congestion game if $x^* \in W(p^*)$ and for all $p \geq 0$, there exists $x \in W(p)$ such that

$$\Pi(p^*, x^*) \geq \Pi(p, x).$$

The following proposition shows that under Assumption 2.1, the two solution concepts coincide. Since the proof is not relevant for the rest of the argument, we provide it in Appendix A.

Proposition 3.1. *Let Assumption 2.1 hold. A vector $(p^{ME}, x^{ME})$ is an ME if and only if it is an SPE of the pricing congestion game.*

Since an ME $(p^*, x^*)$ is an optimal solution of the optimization problem

$$\begin{aligned} &\underset{p\geq 0,\, x\geq 0}{\text{maximize}} \ \sum_{i\in\mathcal{I}} p_i x_i \\ &\text{subject to} \ \ x \in W(p), \end{aligned} \tag{9}$$

it is easier to work with than an SPE. Therefore we use ME as the solution concept in this paper.

The preceding problem has an optimal solution, which establishes the existence of an ME. Moreover, we have:

Proposition 3.2. *Let Assumption 2.1 hold. A vector $x$ is the flow vector at an ME if and only if it is a social optimum. Moreover, if $(p, x)$ is an ME, then for all $i$ with $x_i > 0$, we have $p_i = R - l_i(x_i)$.*

This proposition therefore establishes that the flow allocation at an ME and the social optimum are the same. Its proof is similar to an analogous result in Acemoglu and Ozdaglar [1] and is omitted.

In addition to the social surplus defined above, it is also useful to define the *consumer surplus*, as the difference between users' willingness to pay and effective cost, i.e., $\sum_{i=1}^{I}(R - l_i(x_i) - p_i)x_i$ (see Mascolell et al. [32]). By Proposition 3.2, it is clear that even though the ME achieves the social optimum, all of the surplus is captured by the monopolist, and users are just indifferent between sending their information or not (i.e., receive no consumer surplus).

Our major motivation for the study of oligopolistic settings is that they provide a better approximation to reality, where there is typically competition among service providers. A secondary motivation is to see whether an OE will achieve an efficient allocation like the ME, while also transferring some or all of the surplus to the consumers.

**4. Oligopoly equilibrium.** We suppose 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 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*.

Definition 4.1. A vector $(p^{OE}, x^{OE}) \geq 0$ is a (pure strategy) *oligopoly equilibrium* (OE) if $x^{OE} \in W(p_s^{OE}, p_{-s}^{OE})$ and for all $s \in \mathcal{S}$,

$$\Pi_s(p_s^{OE}, p_{-s}^{OE}, x^{OE}) \geq \Pi_s(p_s, p_{-s}^{OE}, x), \quad \forall\, p_s \geq 0, \ \ \forall\, x \in W(p_s, p_{-s}^{OE}). \tag{10}$$

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

As for the monopoly case, there is a close relation between a pure strategy OE and a pure strategy SPE. Again, associating the SPE with the on-the-equilibrium-path actions, we have:

Definition 4.2. A vector $(p^*, x^*) \geq 0$ is a SPE of the price competition game if $x^* \in W(p^*)$ and there exists a function $x: \mathbb{R}_+^I \mapsto \mathbb{R}_+^I$ such that $x(p) \in W(p)$ for all $p \geq 0$ and for all $s \in \mathcal{S}$,

$$\Pi_s(p_s^*, p_{-s}^*, x^*) \geq \Pi_s(p_s, p_{-s}^*, x(p_s, p_{-s}^*)) \quad \forall\, p_s \geq 0. \tag{11}$$

The following proposition generalizes Proposition 3.1 and enables us to work with the OE definition, which is more convenient for the subsequent analysis. The proof parallels that of Proposition 3.1 and is omitted.

Proposition 4.1. *Let Assumption 2.1 hold. A vector* $(p^{OE}, x^{OE})$ *is an OE if and only if it is an SPE of the price competition game.*

The price competition game is neither concave nor supermodular. Therefore, classical arguments that are used to show the existence of a pure strategy equilibrium do not hold (see Fudenberg and Tirole [16], Topkis [51]). In the next proposition, we show that for linear latency functions, there exists a pure strategy OE. The proof is provided in Appendix B.

Proposition 4.2. *Let Assumption 2.1 hold, and assume further that the latency functions are linear. Then the price competition game has a pure strategy OE.*

The existence result cannot be generalized to piecewise linear latency functions or to latency functions that are linear over their effective domain, as illustrated in the following example.

Example 4.1. Consider a two-link network. Let the total flow be $d = 1$. Assume that the latency functions are given by

$$l_1(x) = 0, \qquad l_2(x) = \begin{cases} 0 & \text{if } 0 \leq x \leq \delta \\ \dfrac{x - \delta}{\epsilon} & \text{if } x \geq \delta, \end{cases}$$

for some $\epsilon > 0$ and $\delta > 1/2$, with the convention that when $\epsilon = 0$, $l_2(x) = \infty$ for $x > \delta$. We first show that there exists no pure strategy OE for small $\epsilon$ (i.e., there exists no pure strategy SPE). The following list considers all candidate oligopoly price equilibria $(p_1, p_2)$ and profitable unilateral deviations for $\epsilon$ sufficiently small, thus establishing the nonexistence of an OE:

(i) $p_1 = p_2 = 0$: A small increase in the price of provider 1 will generate positive profits, thus provider 1 has an incentive to deviate.

(ii) $p_1 = p_2 > 0$: Let $x$ be the flow allocation at the OE. If $x_1 = 1$, then provider 2 has an incentive to decrease its price. If $x_1 < 1$, then provider 1 has an incentive to decrease its price.

(iii) $0 \le p_1 < p_2$: Player 1 has an incentive to increase its price because its flow allocation remains the same.

(iv) $0 \le p_2 < p_1$: For $\epsilon$ sufficiently small, the profit function of player 2, given $p_1$, is strictly increasing as a function of $p_2$, showing that provider 2 has an incentive to increase its price.

We next show that a mixed-strategy OE always exists. We define a mixed-strategy OE as a mixed-strategy SPE of the price competition game (see Dasgupta and Maskin [13]). Let $\mathscr{B}^n$ be the space of all (Borel) probability measures on $[0, R]^n$. Let $I_s$ denote the cardinality of $\mathscr{I}_s$, i.e., the number of links controlled by service provider $s$. Let $\mu_s \in \mathscr{B}^{I_s}$ be a probability measure, and denote the vector of these probability measures by $\mu$ and the vector of these probability measures excluding $s$ by $\mu_{-s}$.

Definition 4.3. $(\mu^*, x^*(p))$ is a mixed-strategy OE if the function $x^*(p) \in W(p)$ for every $p \in [0, R]^I$ and

$$\int_{[0,R]^I} \Pi_s(p_s, p_{-s}, x^*(p_s, p_{-s}))\, d(\mu_s^*(p_s) \times \mu_{-s}^*(p_{-s})) \ge \int_{[0,R]^I} \Pi_s(p_s, p_{-s}, x^*(p_s, p_{-s}))\, d(\mu_s(p_s) \times \mu_{-s}^*(p_{-s}))$$

for all $s$ and $\mu_s \in \mathscr{B}^{I_s}$.

Therefore, a mixed-strategy OE simply requires that there be no profitable deviation to a different probability measure for each oligopolist.

Example 4.1 (continued). We now show that the following strategy profile is the unique mixed-strategy OE for the above game when $\epsilon \to 0$ (a mixed-strategy OE also exists when $\epsilon > 0$, but its structure is more complicated and less informative):

$$\mu_1(p) = \begin{cases} 0 & 0 \le p \le R(1-\delta), \\ 1 - \dfrac{(1-\delta)R}{p} & R(1-\delta) \le p < R, \\ 1 & \text{otherwise.} \end{cases}$$

$$\mu_2(p) = \begin{cases} 0 & 0 \le p \le R(1-\delta), \\ \dfrac{1}{\delta} - \dfrac{(1-\delta)R}{\delta p} & R(1-\delta) \le p \le R, \\ 1 & \text{otherwise.} \end{cases}$$

Notice that $\mu_1$ has an atom equal to $1-\delta$ at $R$. To verify that this profile is a mixed-strategy OE, let $\mu'$ be the density of $\mu$, with the convention that $\mu' = \infty$ when there is an atom at that point. Let $M_i = \{p \mid \mu_i'(p) > 0\}$. To establish that $(\mu_1, \mu_2)$ is a mixed-strategy equilibrium, it suffices to show that the expected payoff to player $i$ is constant for all $p_i \in M_i$ when the other player chooses $p_{-i}$ according to $\mu_{-i}$ (see Osborne and Rubinstein [37]). These expected payoffs are

$$\overline{\Pi}_i(p_i \mid \mu_{-i}) \equiv \int_0^R \Pi_i(p_i, p_{-i}, x(p_i, p_{-i}))\, d\mu_{-i}(p_{-i}). \tag{12}$$

The WE demand $x(p_1, p_2)$ takes the simple form of $x_1(p_1, p_2) = 1$ if $p_1 < p_2$ and $x_1(p_1, p_2) = 1-\delta$ if $p_1 > p_2$. The exact value of $x_1(p_1, p_2)$ when $p_1 = p_2$ is immaterial since this event happens with zero probability. It is evident that the expression in (12) is constant for all $p_i \in M_i$, for $i = 1, 2$ given $\mu_1$ and $\mu_2$ above. This establishes that $(\mu_1, \mu_2)$ is a mixed-strategy OE. It can also be verified that there are no other mixed-strategy equilibria.

The next proposition, which is proved in Appendix C, establishes that a mixed-strategy equilibrium always exists.

Proposition 4.3. *Let Assumption 2.1 hold. Then the price competition game has a mixed-strategy OE,* $(\mu^{OE}, x^{OE}(p))$.

We next provide an explicit characterization of pure strategy OE. Though of also independent interest, these results are most useful for us to quantify the efficiency loss of oligopoly in the next section.

The following lemma shows that an equivalent to Lemma 1 (which required real-valued latency functions) also holds with more general latency functions at the pure strategy OE.

Lemma 4.1. *Let Assumption 2.1 hold. If* $(p^{OE}, x^{OE})$ *is a pure strategy OE, then*

$$l_i(x_i^{OE}) + p_i^{OE} = \min_j \{l_j(x_j^{OE}) + p_j^{OE}\}, \quad \forall\, i \text{ with } x_i^{OE} > 0, \tag{13}$$

$$l_i(x_i^{OE}) + p_i^{OE} \le R, \quad \forall\, i \text{ with } x_i^{OE} > 0, \tag{14}$$

$$\sum_{i\in\mathcal{I}} x_i^{OE} \le d, \tag{15}$$

*with* $\sum_{i\in\mathcal{I}} x_i^{OE} = d$ *if* $\min_j\{l_j(x_j^{OE}) + p_j\} < R$.

Proof. Let $(p^{OE}, x^{OE})$ be an OE. Since $x^{OE} \in W(p^{OE})$, Conditions (14) and (15) follow by the definition of a WE. Consider Condition (13). Assume that there exist some $i, j \in \mathcal{I}$ with $x_i^{OE} > 0$, $x_j^{OE} > 0$ such that

$$l_i(x_i^{OE}) + p_i^{OE} < l_j(x_j^{OE}) + p_j^{OE}.$$

Using the optimality conditions for a WE (cf. Equation (2)), this implies that $x_i^{OE} = b_{C_i}$. Consider changing $p_i^{OE}$ to $p_i^{OE} + \epsilon$ for some $\epsilon > 0$. By checking the optimality conditions, we see that we can choose $\epsilon$ sufficiently small such that $x^{OE} \in W(p_i^{OE} + \epsilon, p_{-i}^{OE})$. Hence the service provider who owns link $i$ can deviate to $p_i^{OE} + \epsilon$ and increase its profits, contradicting the fact that $(p^{OE}, x^{OE})$ is an OE. Finally, assume to arrive at a contradiction that $\min_j\{l_j(x_j^{OE}) + p_j\} < R$ and $\sum_{i\in\mathcal{I}} x_i^{OE} < d$. Using the optimality conditions for a WE (Equation (2) with $\lambda = 0$ since $\sum_{i\in\mathcal{I}} x_i^{OE} < d$), this implies that we must have $x_i^{OE} = b_{C_i}$ for some $i$. With a similar argument to the above, a deviation to $p_i^{OE} + \epsilon$ keeps $x^{OE}$ as a WE, and is more profitable, completing the proof. □

We need the following additional assumption for our price characterization.

Assumption 4.1. *Given a pure strategy OE* $(p^{OE}, x^{OE})$, *if for some* $i \in \mathcal{I}$ *with* $x_i^{OE} > 0$, *we have* $l_i(x_i^{OE}) = 0$, *then* $\mathcal{I}_s = \{i\}$.

Note that this assumption is automatically satisfied if all latency functions are strictly increasing or if all service providers own only one link.

Lemma 4.2. *Let* $(p^{OE}, x^{OE})$ *be a pure strategy OE. Let Assumptions 2.1 and 4.1 hold. Let* $\Pi_s$ *denote the profit of service provider* $s$ *at* $(p^{OE}, x^{OE})$.

(a) *If* $\Pi_{s'} > 0$ *for some* $s' \in \mathcal{S}$, *then* $\Pi_s > 0$ *for all* $s \in \mathcal{S}$.

(b) *If* $\Pi_s > 0$ *for some* $s \in \mathcal{S}$, *then* $p_j^{OE} x_j^{OE} > 0$ *for all* $j \in \mathcal{I}_s$.

Proof.

(a) For some $j \in \mathcal{I}_{s'}$, define $K = p_j^{OE} + l_j(x_j^{OE})$, which is positive since $\Pi_{s'} > 0$. Assume $\Pi_s = 0$ for some $s$. For $k \in \mathcal{I}_s$, consider the price $\bar{p}_k = K - \epsilon > 0$ for some small $\epsilon > 0$. By the assumption that $l_k(0) = 0$, it can be seen that at the price vector $(\bar{p}_k, p_{-k}^{OE})$, the corresponding WE link flow will satisfy $\bar{x}_k > 0$. Hence, service provider $s$ has an incentive to deviate to $\bar{p}_k$ at which he will make positive profit, contradicting the fact that $(p^{OE}, x^{OE})$ is a pure strategy OE.

(b) Since $\Pi_s > 0$, we have $p_m^{OE} x_m^{OE} > 0$ for some $m \in \mathcal{I}_s$. By Assumption 4.1, we can assume without loss of generality that $l_m(x_m^{OE}) > 0$ (otherwise, we are done). Let $j \in \mathcal{I}_s$ and assume to arrive at a contradiction that $p_j^{OE} x_j^{OE} = 0$. The profit of service provider $s$ at the pure strategy OE can be written as

$$\Pi_s = \overline{\Pi}_s + p_m^{OE} x_m^{OE},$$

where $\overline{\Pi}_s$ denotes the profits from links other than $m$ and $j$. Let $p_m^{OE} = K - l_m(x_m^{OE})$ for some $K$. Consider changing the prices $p_m^{OE}$ and $p_j^{OE}$ such that the new profit is

$$\widetilde{\Pi}_s = \overline{\Pi}_s + (K - l_m(x_m^{OE} - \epsilon))(x_m^{OE} - \epsilon) + \epsilon(K - l_j(\epsilon)).$$

Note that $\epsilon$ units of flow are moved from link $m$ to link $j$ such that the flows of other links remain the same at the new WE. Hence the change in the profit is

$$\tilde{\Pi}_s - \Pi_s = (l_m(x_m^{OE}) - l_m(x_m^{OE} - \epsilon))x_m^{OE} + \epsilon(l_m(x_m^{OE} - \epsilon) - l_j(\epsilon)).$$

Since $l_m(x_m^{OE}) > 0$ and $l_j(0) = 0$, $\epsilon$ can be chosen sufficiently small such that the above is strictly positive, contradicting the fact that $(p^{OE}, x^{OE})$ is an OE. □

The following example shows that Assumption 4.1 cannot be dispensed with for part (b) of this lemma.

Example 4.2. Consider a three-link network with two providers, where provider 1 owns links 1 and 3 and provider 2 owns link 2. Let the total flow be $d = 1$ and the reservation utility be $R = 1$. Assume that the latency functions are given by

$$l_1(x_1) = 0, \qquad l_2(x_2) = x_2, \qquad l_3(x_3) = ax_3,$$

for some $a > 0$. Any price vector $(p_1, p_2, p_3) = (2/3, 1/3, b)$ with $b \geq 2/3$ and $(x_1, x_2, x_3) = (2/3, 1/3, 0)$ is a pure strategy OE, so $p_3 x_3 = 0$ contrary to part (b) of the lemma. To see why this is an equilibrium, note that provider 2 is clearly playing a best response. Moreover, in this allocation $\Pi_1 = 4/9$. We can represent any deviation of provider 1 by

$$(p_1, p_3) = (2/3 - \delta, 2/3 - a\epsilon - \delta),$$

for two scalars $\epsilon$ and $\delta$ , which will induce a WE of $(x_1, x_2, x_3) = (2/3 + \delta - \epsilon, 1/3 - \delta, \epsilon)$. The corresponding profit of provider 1 at this deviation is $\Pi_1 = 4/9 - \delta^2 < 4/9$, establishing that provider 1 is also playing a best response and we have a pure strategy OE.

We next establish that, under an additional mild assumption, a pure strategy OE will never be at a point of nondifferentiability of the latency functions.

Assumption 4.2. *There exists some $s \in \mathcal{S}$ such that $l_i$ is real-valued and continuously differentiable for all $i \in \mathcal{I}_s$.*

Lemma 4.3. *Let $(p^{OE}, x^{OE})$ be an OE with $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} < R$ and $p_i^{OE} x_i^{OE} > 0$ for some $i$. Let Assumptions 2.1, 4.1, and 4.2 hold. Then*

$$l_i^+(x_i^{OE}) = l_i^-(x_i^{OE}), \quad \forall\, i \in \mathcal{I},$$

*where $l_i^+(x_i^{OE})$ and $l_i^-(x_i^{OE})$ are the right and left derivatives of the function $l_i$ at $x_i^{OE}$, respectively.*

Because the proof of this lemma is long, it is given in Appendix D. Note that Assumption 4.2 cannot be dispensed within this lemma. This is illustrated in the next example.

Example 4.3. Consider a two-link network. Let the total flow be $d = 1$ and the reservation utility be $R = 2$. Assume that the latency functions are given by

$$l_1(x) = l_2(x) = \begin{cases} 0 & \text{if } 0 \leq x \leq \dfrac{1}{2} \\ 2\left(x - \dfrac{1}{2}\right) & \text{otherwise.} \end{cases}$$

It can be verified that the vector $(p_1^{OE}, p_2^{OE}) = (1, 1)$, with $(x_1^{OE}, x_2^{OE}) = (1/2, 1/2)$ is a pure strategy OE, and is at a point of nondifferentiability for both latency functions.

We next provide an explicit characterization of the OE prices, which is essential in our efficiency analysis in §5. The proof is given in Appendix D.

Proposition 4.4. *Let $(p^{OE}, x^{OE})$ be an OE such that $p_i^{OE} x_i^{OE} > 0$ for some $i \in \mathcal{I}$. Let Assumptions 2.1, 4.1, and 4.2 hold.*

(a) *Assume that $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} < R$. Then, for all $s \in \mathcal{S}$ and $i \in \mathcal{I}_s$, we have*

$$p_i^{OE} = \begin{cases} x_i^{OE} l_i'(x_i^{OE}), & \text{if } l_j'(x_j^{OE}) = 0 \text{ for some } j \notin \mathcal{I}_s, \\ x_i^{OE} l_i'(x_i^{OE}) + \dfrac{\sum_{j \in \mathcal{I}_s} x_j^{OE}}{\sum_{j \notin \mathcal{I}_s} 1/l_j'(x_j^{OE})}, & \text{otherwise}. \end{cases} \tag{16}$$

(b) *Assume that* $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} = R$. *Then, for all* $s \in \mathcal{S}$ *and* $i \in \mathcal{I}_s$, *we have*

$$p_i^{OE} \geq x_i^{OE} l_i^-(x_i^{OE}). \tag{17}$$

*Moreover, if there exists some* $i \in \mathcal{I}$ *such that* $\mathcal{I}_s = \{i\}$ *for some* $s \in \mathcal{S}$, *then*

$$p_i^{OE} \leq x_i^{OE} l_i^+(x_i^{OE}) + \frac{x_i^{OE}}{\sum_{j\neq i} 1/l_j^-(x_j^{OE})}. \tag{18}$$

If the latency functions $l_i$ are all real-valued and continuously differentiable, then analysis of Karush-Kuhn-Tucker (KKT) condition for oligopoly problem (Problem (A24) in Appendix E) immediately yields the following result.

Corollary 4.1. *Let* $(p^{OE}, x^{OE})$ *be an OE such that* $p_i^{OE} x_i^{OE} > 0$ *for some* $i \in \mathcal{I}$. *Let Assumptions 2.1 and 4.1 hold. Assume also that* $l_i$ *is real-valued and continuously differentiable for all* $i$. *Then, for all* $s \in \mathcal{S}$ *and* $i \in \mathcal{I}_s$, *we have*

$$p_i^{OE} = \begin{cases} x_i^{OE} l_i'(x_i^{OE}), & \text{if } l_j'(x_j^{OE}) = 0 \text{ for some } j \notin \mathcal{I}_s, \\ \min\left\{ R - l_i(x_i^{OE}),\, x_i^{OE} l_i'(x_i^{OE}) + \dfrac{\sum_{j\in\mathcal{I}_s} x_j^{OE}}{\sum_{j\notin\mathcal{I}_s} 1/l_j'(x_j^{OE})} \right\}, & \text{otherwise}. \end{cases} \tag{19}$$

This corollary also implies that in the two-link case with real-valued and continuously differentiable latency functions and with minimum effective cost less than $R$, the OE prices are

$$p_i^{OE} = x_i^{OE}(l_1'(x_1^{OE}) + l_2'(x_2^{OE})) \tag{20}$$

as claimed in the introduction.

**5. Efficiency of oligopoly equilibria.** This section contains our main results, providing tight bounds on the inefficiency of OE. We take as our measure of efficiency the ratio of the social surplus of the equilibrium flow allocation to the social surplus of the social optimum, $\mathbb{S}(x^*)/\mathbb{S}(x^S)$, where $x^*$ refers to the ME or the OE (cf. Equation (8)). Section 3 established that the flow allocation at a ME is a social optimum. Hence, in congestion games with monopoly pricing, there is no efficiency loss. The following example shows that this is not necessarily the case with oligopoly pricing.

Example 5.1. Consider a two-link network. Let the total flow be $d = 1$ and the reservation utility be $R = 1$. The latency functions are given by

$$l_1(x) = 0, \qquad l_2(x) = \frac{3}{2}x.$$

The unique social optimum for this example is $x^S = (1, 0)$. The unique ME $(p^{ME}, x^{ME})$ is $x^{ME} = (1, 0)$ and $p^{ME} = (1, 1)$. As expected, the flow allocations at the social optimum and the ME are the same. Next, consider a duopoly where each of these links is owned by a different provider. Using Corollary 4.1 and Lemma 4.1, it follows that the flow allocation at the OE, $x^{OE}$, satisfies

$$l_1(x_1^{OE}) + x_1^{OE}[l_1'(x_1^{OE}) + l_2'(x_2^{OE})] = l_2(x_2^{OE}) + x_2^{OE}[l_1'(x_1^{OE}) + l_2'(x_2^{OE})].$$

Solving this together with $x_1^{OE} + x_2^{OE} = 1$ shows that the flow allocation at the unique OE is $x^{OE} = (2/3, 1/3)$. The social surplus at the social optimum, the ME, and the OE are given by 1, 1, and 5/6, respectively.

Before providing a more thorough analysis of the efficiency properties of the OE, the next proposition proves that, as claimed in the introduction and suggested by Example 5.1, a change in the market structure from monopoly to duopoly in a two-link network typically reduces efficiency.

Proposition 5.1. *Consider a two-link network where each link is owned by a different provider. Let Assumption 2.1 hold. Let* $(p^{OE}, x^{OE})$ *be a pure strategy OE such that* $p_i^{OE} x_i^{OE} > 0$ *for some* $i \in \mathcal{I}$ *and* $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} < R$. *If* $l_1'(x_1^{OE})/x_1^{OE} \neq l_2'(x_2^{OE})/x_2^{OE}$, *then* $\mathbb{S}(x^{OE})/\mathbb{S}(x^S) < 1$.

Proof. Combining the OE prices with the WE conditions, we have

$$l_1(x_1^{OE}) + x_1^{OE}(l_1'(x_1^{OE}) + l_2'(x_2^{OE})) = l_2(x_2^{OE}) + x_2^{OE}(l_1'(x_1^{OE}) + l_2'(x_2^{OE})),$$

where we use the fact that $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} < R$. Moreover, we can use optimality Conditions (7) to prove that a vector $(x_1^S, x_2^S) > 0$ is a social optimum if and only if

$$l_1(x_1^S) + x_1^S l_1'(x_1^S) = l_2(x_2^S) + x_2^S l_2'(x_2^S).$$

Since $l_1'(x_1^{OE})/x_1^{OE} \neq l_2'(x_2^{OE})/x_2^{OE}$, the result follows. □

We next quantify the efficiency of OE by providing a tight bound on the efficiency loss in congestion games with oligopoly pricing. As we have shown in §4, such games do not always have a pure strategy OE. In the following, we first provide bounds on congestion games that have pure strategy equilibria. We next study efficiency properties of mixed-strategy equilibria.

**5.1. Pure strategy equilibria.** We consider price competition games that have pure strategy equilibria (this set includes, but is substantially larger than, games with linear latency functions, see §4). We consider latency functions that satisfy Assumptions 2.1, 4.1, and 4.2. Let $\mathscr{L}_I$ denote the set of latency functions for which the associated price competition game has a pure strategy OE and the individual $l_i$'s satisfy Assumptions 2.1, 4.1, and 4.2.[11] We refer to an element of the set $\mathscr{L}_I$ by $\{l_i\}_{i\in\mathscr{I}}$. Given a parallel link network with $I$ links and latency functions $\{l_i\}_{i\in\mathscr{I}} \in \mathscr{L}_I$, let $\overrightarrow{OE}(\{l_i\})$ denote the set of flow allocations at an OE. We define the efficiency metric at some $x^{OE} \in \overrightarrow{OE}(\{l_i\})$ as

$$r_I(\{l_i\}, x^{OE}) = \frac{R\sum_{i\in\mathscr{I}} x_i^{OE} - \sum_{i\in\mathscr{I}} l_i(x_i^{OE})x_i^{OE}}{R\sum_{i\in\mathscr{I}} x_i^S - \sum_{i\in\mathscr{I}} l_i(x_i^S)x_i^S}, \tag{21}$$

where $x^S$ is a social optimum given the latency functions $\{l_i\}_{i\in\mathscr{I}}$ and $R$ is the reservation utility. In other words, our 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 [25], we are interested in the worst performance in an OE, so we look for a lower bound on

$$\inf_{\{l_i\}\in\mathscr{L}_I} \inf_{x^{OE}\in\overrightarrow{OE}(\{l_i\})} r_I(\{l_i\}, x^{OE}).$$

We first prove two lemmas, which reduce the set of latency functions that need to be considered in bounding the efficiency metric. The next lemma allows us to use the oligopoly price characterization given in Proposition 4.4.

Lemma 5.1. *Let $(p^{OE}, x^{OE})$ be a pure strategy OE such that $p_i^{OE} x_i^{OE} = 0$ for all $i \in \mathscr{I}$. Then $x^{OE}$ is a social optimum.*

Proof. We first show that $l_i(x_i^{OE}) = 0$ for all $i \in \mathscr{I}$. Assume that $l_j(x_j^{OE}) > 0$ for some $j \in \mathscr{I}$. This implies that $x_j^{OE} > 0$, and therefore $p_j^{OE} = 0$. Since $l_j(x_j^{OE}) > 0$, it follows by Lemma 2.2 that for all $x \in W(p)$, we have $x_j = x_j^{OE}$. Consider increasing $p_j^{OE}$ to some small $\epsilon > 0$. By the upper semicontinuity of $W(p)$, it follows that there exists some $\epsilon > 0$ sufficiently small such that for all $x \in W(\epsilon, p_{-j}^{OE})$, we have $|x_j - x_j^{OE}| < \delta$ for some $\delta > 0$. Moreover, by Proposition 2.2, we have, for all $x \in W(\epsilon, p_{-j}^{OE})$, $x_i \geq x_i^{OE}$ for all $i \neq j$. Hence the profit of the provider that owns link $j$ is strictly higher at price vector $(\epsilon, p_{-j}^{OE})$ than at $p^{OE}$, contradicting the fact that $(p^{OE}, x^{OE})$ is an OE.

Clearly, $x_j^{OE} > 0$ for some $j$, and hence $\min_{i\in\mathscr{I}}\{p_i^{OE} + l_i(x_i^{OE})\} = p_j^{OE} + l_j(x_j^{OE}) = 0$, which implies by Lemma 4.1 that $\sum_{i\in\mathscr{I}} x_i^{OE} = d$. Using $l_i(x_i^{OE}) = 0$ and $0 \in \partial l_i(x_i^{OE})$ for all $i$, we have

$$R - l_i(x_i^{OE}) - x_i^{OE} g_{l_i} = R, \quad \forall\, i \in \mathscr{I}$$

for some $g_{l_i} \in \partial l_i(x_i^{OE})$. Hence $x^{OE}$ satisfies the sufficient optimality conditions for a social optimum (cf. Equation (7) with $\lambda^S = R$), and the result follows. □

The next lemma allows us to assume without loss of generality that $R\sum_{i\in\mathscr{I}} x_i^S - \sum_{i\in\mathscr{I}} l_i(x_i^S)x_i^S > 0$ and $\sum_{i\in\mathscr{I}} x_i^{OE} = d$ in the subsequent analysis.

[11] More explicitly, Assumption 4.1 implies that if any OE $(p^{OE}, x^{OE})$ associated with $\{l_i\}_{i\in\mathscr{I}}$ has $x_i^{OE} > 0$ and $l_i(x_i^{OE}) = 0$, then $\mathscr{I}_s = \{i\}$.

LEMMA 5.2. *Let $\{l_i\}_{i\in\mathcal{I}} \in \mathcal{L}_I$. Assume that*
*either* (i) $\sum_{i\in\mathcal{I}} l_i(x_i^S)x_i^S = R\sum_{i\in\mathcal{I}} x_i^S$ *for some social optimum $x_s$,*
*or* (ii) $\sum_{i\in\mathcal{I}} x_i^{OE} < d$ *for some* $x^{OE} \in \overrightarrow{OE}(\{l_i\})$.
*Then every* $x^{OE} \in \overrightarrow{OE}(\{l_i\})$ *is a social optimum, implying that* $r_I(\{l_i\}, x^{OE}) = 1$.

PROOF. Assume that $\sum_{i\in\mathcal{I}} l_i(x_i^S)x_i^S = R\sum_{i\in\mathcal{I}} x_i^S$. Since $x^S$ is a social optimum and every $x^{OE} \in \overrightarrow{OE}(\{l_i\})$ is a feasible solution to the social problem (Problem (6)), we have

$$0 = \sum_{i\in\mathcal{I}}(R - l_i(x_i^S))x_i^S \geq \sum_{i\in\mathcal{I}}(R - l_i(x_i^{OE}))x_i^{OE}, \quad \forall\, x^{OE} \in \overrightarrow{OE}(\{l_i\}).$$

By the definition of a WE, we have $x_i^{OE} \geq 0$ and $R - l_i(x_i^{OE}) \geq p_i^{OE} \geq 0$ (where $p_i^{OE}$ is the price of link $i$ at the OE) for all $i$. This combined with the preceding relation shows that $x^{OE}$ is a social optimum.

Assume next that $\sum_{i\in\mathcal{I}} x_i^{OE} < d$ for some $x^{OE} \in \overrightarrow{OE}(\{l_i\})$. Let $p^{OE}$ be the associated OE price. Assume that $p_j^{OE}x_j^{OE} > 0$ for some $j \in \mathcal{I}$ (otherwise we are done by Lemma 5.1). Since $\sum_{i\in\mathcal{I}} x_i^{OE} < d$, we have by Lemma 4.1 that $\min_{j\in\mathcal{I}}\{p_j + l_j(x_j^{OE}\} = R$. Moreover, by Lemma 4.2, it follows that $p_i x_i^{OE} > 0$ for all $i \in \mathcal{I}$. Hence, for all $s \in \mathcal{S}$, $((p_i^{OE})_{i\in\mathcal{I}_s}, x^{OE})$ is an optimal solution of the problem

$$\begin{aligned} \underset{((p_i)_{i\in\mathcal{I}_s},\, x)}{\text{maximize}} \quad & \sum_{i\in\mathcal{I}_s} p_i x_i \\ \text{subject to} \quad & p_i + l_i(x_i) = R, \quad \forall\, i \in \mathcal{I}_s, \\ & p_i^{OE} + l_i(x_i) = R, \quad \forall\, i \notin \mathcal{I}_s, \\ & \sum_{i\in\mathcal{I}} x_i^{OE} \leq d. \end{aligned}$$

Substituting for $(p_i)_{i\in\mathcal{I}_s}$ in the above, we obtain

$$\begin{aligned} \underset{x\geq 0}{\text{maximize}} \quad & \sum_{i\in\mathcal{I}_s}(R - l_i(x_i))x_i \\ \text{subject to} \quad & x_i \in T_i, \quad \forall\, i \notin \mathcal{I}_s, \\ & \sum_{i\in\mathcal{I}} x_i^{OE} \leq d, \end{aligned}$$

where $T_i = \{x_i \mid p_i^{OE} + l_i(x_i) = R\}$ is either a singleton or a closed interval. Since this is a convex problem, using the optimality conditions, we obtain

$$R - l_i(x_i^{OE}) - x_i^{OE} g_{l_i} = 0, \quad \forall\, i \in \mathcal{I}_s, \ \forall\, s \in \mathcal{S},$$

where $g_{l_i} \in \partial l_i(x_i^{OE})$. By Equation (7), it follows that $x^{OE}$ is a social optimum. ☐

This lemma implies that in finding a lower bound on the efficiency metric, we can restrict ourselves, without loss of generality, to latency functions $\{l_i\} \in \mathcal{L}_I$ such that $\sum_{i\in\mathcal{I}} l_i(x_i^S)x_i^S < R\sum_{i\in\mathcal{I}} x_i^S$ for some social optimum $x^S$, and $\sum_{i\in\mathcal{I}} x_i^{OE} = d$ for all $x^{OE} \in \overrightarrow{OE}(\{l_i\})$. By the following lemma, we can also assume that $\sum_{i\in\mathcal{I}} x_i^S = d$.

LEMMA 5.3. *For a set of latency functions $\{l_i\}_{i\in\mathcal{I}}$, let Assumption 2.1 hold. Let $(p^{OE}, x^{OE})$ be an OE and $x^S$ be a social optimum. Then*

$$\sum_{i\in\mathcal{I}} x_i^{OE} \leq \sum_{i\in\mathcal{I}} x_i^S.$$

PROOF. Assume to arrive at a contradiction that $\sum_{i\in\mathcal{I}} x_i^{OE} > \sum_{i\in\mathcal{I}} x_i^S$. This implies that $x_j^{OE} > x_j^S$ for some $j$. We also have $l_j(x_j^{OE}) > l_j(x_j^S)$. (Otherwise, we would have $l_j(x_j^S) = l_j'(x_j^S) = 0$, which yields a contradiction by the optimality Conditions (7) and the fact that $\sum_{i\in\mathcal{I}} x_i^S < d$.) Using the optimality Conditions (2) and (7), we obtain

$$R - l_j(x_j^{OE}) - p_j^{OE} \geq R - l_j(x_j^S) - x_j^S g_{l_j},$$

for some $g_{l_j} \in \partial l_j(x_j^S)$. Combining the preceding with $l_j(x_j^{OE}) > l_j(x_j^S)$ and $p_j^{OE} \geq x_j^{OE} l_j^-(x_j^{OE})$ (cf. Proposition 4.4), we see that

$$x_j^{OE} l_j^-(x_j^{OE}) < x_j^S g_{l_j},$$

contradicting $x_j^{OE} > x_j^S$ and completing the proof. ☐

**5.1.1. Two links.** We first consider a parallel link network with two links owned by two service providers. The next theorem provides a tight lower bound of $5/6$ on $r_2(\{l_i\}, x^{OE})$ (cf. Equation (21)).

Starting with the two-link network is useful for two reasons: (1) the two-link network avoids the additional layer of optimization over the allocation of links to service providers in characterizing the bound on inefficiency and (2) we will prove the result for the general case by reducing it to the proof of the two-link case.

Although the details of the proof of the theorem are involved, the structure is straightforward. The problem of finding a lower bound on $r_2(\{l_i\}, x^{OE})$ is an infinite-dimensional problem, since the minimization is over latency functions. The proof first provides a lower bound to the infinite-dimensional problem by the optimal value of a finite-dimensional optimization problem using the relations between the flows at social optimum and equilibrium, and convexity of the latency functions. It then shows that the solution will involve one of the links having zero latency. Finally, using this fact and the price characterization from Proposition 4.4, it reduces the problem of characterizing the bound on inefficiency to a simple minimization problem, with optimal value 5/6. An intuition for this value is provided below.

In the following, we assume without loss of generality that $d = 1$. Also, recall that latency functions in $\mathcal{L}_2$ satisfy Assumptions 2.1, 4.1, and 4.2.

Theorem 5.1. *Consider a two-link network where each link is owned by a different provider. Then*

$$r_2(\{l_i\}, x^{OE}) \geq \frac{5}{6}, \quad \forall \{l_i\}_{i=1,2} \in \mathcal{L}_2, \quad x^{OE} \in \overrightarrow{OE}(\{l_i\}), \tag{22}$$

*and the bound is tight. In particular, there exists $\{l_i\}_{i=1,2} \in \mathcal{L}_2$ and $x^{OE} \in \overrightarrow{OE}(\{l_i\})$ that attains the lower bound in Equation (22).*

Proof. The proof follows a number of steps.

*Step* 1. We are interested in finding a lower bound for the problem

$$\inf_{\{l_i\}\in\mathcal{L}_2} \inf_{x^{OE}\in\overrightarrow{OE}(\{l_i\})} r_2(\{l_i\}, x^{OE}). \tag{23}$$

Given $\{l_i\} \in \mathcal{L}_2$, let $x^{OE} \in \overrightarrow{OE}(\{l_i\})$ and let $x^S$ be a social optimum. By Lemmas 5.2 and 5.3, we can assume that $\sum_{i=1}^2 x_i^{OE} = \sum_{i=1}^2 x_i^S = 1$. This implies that there exists some $i$ such that $x_i^{OE} < x_i^S$. Since the problem is symmetric, we can restrict ourselves to $\{l_i\} \in \mathcal{L}_2$ such that $x_1^{OE} < x_1^S$, i.e., we restrict ourselves to $\{l_i\} \in \mathcal{L}_2$ such that $x_1^{OE} \leq x_1^S - \epsilon$ for some $\epsilon > 0$. We claim

$$\inf_{\{l_i\}\in\mathcal{L}_2} \inf_{x^{OE}\in\overrightarrow{OE}(\{l_i\})} r_2(\{l_i\}, x^{OE}) \geq \inf_{\epsilon>0} r_{2,t}^{OE}(\epsilon), \tag{24}$$

where we define problem ($\mathrm{E}^\epsilon$) as

$$(\mathrm{E}^\epsilon) \quad r_{2,t}^{OE}(\epsilon) = \underset{\substack{l_i^S,\,(l_i^S)'\geq 0\\ l_i,\,l_i'\geq 0\\ y_i^S,\,y_i^{OE}\geq 0}}{\text{minimize}} \quad \frac{R - l_1 y_1^{OE} - l_2 y_2^{OE}}{R - l_1^S y_1^S - l_2^S y_2^S}$$

$$\text{subject to} \quad l_i^S \leq y_i^S (l_i^S)', \quad i = 1, 2, \tag{25}$$

$$l_i \leq y_i^{OE} l_i', \quad i = 1, 2, \tag{26}$$

$$l_2^S + y_2^S (l_2^S)' = l_1^S + y_1^S (l_1^S)', \tag{27}$$

$$l_1^S + y_1^S (l_1^S)' \leq R, \tag{28}$$

$$\sum_{i=1}^2 y_i^S \leq 1, \tag{29}$$

$$l_1 + l_1'(y_2^{OE} - y_2^S) \leq l_1^S, \tag{30}$$

$$y_2^{OE} \geq y_2^S + \epsilon, \tag{31}$$

$$\sum_{i=1}^2 y_i^{OE} = 1, \tag{32}$$

$$+\{\text{OE constraints}\}_t, \quad t = 1, 2.$$

Problem ($\mathrm{E}^\epsilon$) can be viewed as a finite-dimensional problem that captures the equilibrium and the social optimum characteristics of the infinite-dimensional problem given in Equation (23). This implies that instead of optimizing over the entire function $l_i$, we optimize over the possible values of $l_i(\cdot)$ and $\partial l_i(\cdot)$ at the equilibrium and the social optimum, which we denote by $l_i$, $l_i'$, $l_i^S$, $(l_i^S)'$ (i.e., $(l_i^S)'$ is a variable that represents all possible values of $g_{l_i} \in \partial l_i(y_i^S)$). The constraints of the problem guarantee that these values satisfy the necessary optimality conditions for a social optimum and an OE. In particular, Conditions (25) and (26) capture the convexity assumption on $l_i(\cdot)$ by relating the values $l_i$, $l_i'$ and $l_i^S$, $(l_i^S)'$ (note that the assumption $l_i(0)=0$ is essential here). Conditions (27) and (28) follow from the optimality conditions for the social optimum. Condition (30) follows by the convexity of the function $l_1(\cdot)$, which implies the relation

$$l_1(x_1^S) \geq l_1(x_1^{OE}) + g_{l_1}(x_1^S - x_1^{OE}),$$

where $g_{l_1} \in \partial l_1(x_1^{OE})$. Using the relation $\sum_{i=1}^2 x_i^{OE} = \sum_{i=1}^2 x_i^S = 1$, we write the preceding constraint as

$$l_1(x_1^S) \geq l_1(x_1^{OE}) + g_{l_1}(x_2^{OE} - x_2^S),$$

which turns out to be more convenient in the analysis of the optimality conditions (see Step 3). Similarly, Condition (31) follows by the facts that we are considering $\{l_i\}$ such that $x_1^{OE} \leq x_1^S - \epsilon$ for some $\epsilon > 0$ and $\sum_{i=1}^2 x_i^{OE} = \sum_{i=1}^2 x_i^S = 1$. Note that we use the relaxed constraint $\sum_{i=1}^2 x_i^S \leq 1$ in the optimization problem (which provides a lower bound to the original problem) since this makes the analysis of the optimality conditions easier.

Finally, the last set of constraints are the necessary conditions for a pure strategy OE. These are written separately for $t = 1, 2$, for the two cases characterized in Proposition 4.4, giving us two bounds, which we will show to be equal.

More explicitly, the OE constraints are given by:

*For* $t = 1$: (corresponding to a lower bound for pure strategy OE, $(p^{OE}, y^{OE})$, with $\min_j\{p_j^{OE} + l_j(y_j^{OE})\} < R$),

$$\begin{gathered} l_1 + y_1^{OE}[l_1' + l_2'] = l_2 + y_2^{OE}[l_1' + l_2'], \\ l_1 + y_1^{OE}[l_1' + l_2'] \leq R, \end{gathered} \tag{33}$$

where $l_1' = l_1'(y_1^{OE})$ and $l_2' = l_2'(y_2^{OE})$ (cf. Equation (16)).

*For* $t = 2$: (corresponding to a lower bound for pure strategy OE, $(p^{OE}, y^{OE})$, with $\min_j\{p_j^{OE} + l_j(y_j^{OE})\} = R$),

$$\begin{gathered} R - l_2 \geq y_2^{OE} l_2', \\ R - l_1 \leq y_1^{OE}[l_1' + l_2'], \end{gathered} \tag{34}$$

where $l_1' = l_1^+(y_1^{OE})$ and $l_2' = l_2^-(y_2^{OE})$ (cf. Equations (17), (18)). We will show in Step 4 that $r_{2,1}^{OE}(\epsilon) = r_{2,2}^{OE}(\epsilon)$.

Note that given any feasible solution of Problem (23), there exists some $\epsilon > 0$ such that we have a feasible solution for problem ($\mathrm{E}^\epsilon$) with the same objective function value. Therefore the optimum value of problem $\inf_{\epsilon>0} r_{2,t}^{OE}(\epsilon)$ is indeed a lower bound on the optimum value of Problem (23).

*Step* 2. Let $(l_i^S, y_i^S)_{i=1,2}$ satisfy Equations (25)–(29). We show that

$$l_1^S y_1^S + l_2^S y_2^S < R. \tag{35}$$

Using Equations (27), (28), and (29), we obtain

$$l_1^S y_1^S + l_2^S y_2^S + (y_1^S)^2 (l_1^S)' + (y_2^S)^2 (l_2^S)' \leq R.$$

If $(y_1^S)^2(l_1^S)' + (y_2^S)^2(l_2^S)' > 0$, then the result follows. If $(y_1^S)^2(l_1^S)' + (y_2^S)^2(l_2^S)' = 0$, then we have using Equation (25) that $l_i^S$=0 for all $i$, again showing the result.

Next, let $(l_i, y_i^{OE})_{i=1,2}$ satisfy Equation (32) and one of the OE constraints (i.e., Equations (33) or (34)). Using a similar argument, we can show that

$$l_1 y_1^{OE} + l_2 y_2^{OE} < R. \tag{36}$$

*Step* 3. Let $(\bar{l}_i^S, (\bar{l}_i^S)', \bar{l}_i, \bar{l}_i', \bar{y}_i^S, \bar{y}_i^{OE})$ denote an optimal solution of problem (E). We show that $\bar{l}_i^S = 0$ for $i = 1, 2$.

We assign the Lagrange multipliers $\mu_i^S \geq 0$, $\lambda^S$, $\gamma^S \geq 0$ to Equations (25), (27), (28), respectively, and $\theta^S \geq 0$ to Equation (29). Using the first-order optimality conditions, we obtain

$$\begin{aligned} \bar{y}_2^S \frac{(R - \bar{l}_1 \bar{y}_1^{OE} - \bar{l}_2 \bar{y}_2^{OE})}{(R - \bar{l}_1^S \bar{y}_1^S - \bar{l}_2^S \bar{y}_2^S)^2} + \mu_2^S + \lambda^S &= 0 \quad \text{if } \bar{l}_2^S > 0 \\ &\geq 0 \quad \text{if } \bar{l}_2^S = 0, \end{aligned} \tag{37}$$

$$\begin{aligned} -\mu_2^S \bar{y}_2^S + \lambda^S \bar{y}_2^S &= 0 \quad \text{if } (\bar{l}_2^S)' > 0 \\ &\geq 0 \quad \text{if } (\bar{l}_2^S)' = 0, \end{aligned} \tag{38}$$

$$\begin{aligned} -\mu_1^S \bar{y}_1^S - \lambda^S \bar{y}_1^S + \gamma^S \bar{y}_1^S &= 0 \quad \text{if } (\bar{l}_1^S)' > 0 \\ &\geq 0 \quad \text{if } (\bar{l}_1^S)' = 0, \end{aligned} \tag{39}$$

$$\begin{aligned} \bar{l}_1^S \frac{(R - \bar{l}_1 \bar{y}_1^{OE} - \bar{l}_2 \bar{y}_2^{OE})}{(R - \bar{l}_1^S \bar{y}_1^S - \bar{l}_2^S \bar{y}_2^S)^2} - \mu_1^S (\bar{l}_1^S)' - \lambda^S (\bar{l}_1^S)' + \gamma^S (\bar{l}_1^S)' + \theta^S &= 0 \quad \text{if } \bar{y}_1^S > 0 \\ &\geq 0 \quad \text{if } \bar{y}_1^S = 0. \end{aligned} \tag{40}$$

We first show that $\bar{l}_2^S = 0$. If $\bar{y}_2^S = 0$ or $(\bar{l}_2^S)' = 0$, we are done by Equation (25). Assume that $\bar{y}_2^S > 0$ and $(\bar{l}_2^S)' > 0$. By Equation (38), this implies that $\lambda^S = \mu_2^S \geq 0$. We claim that in this case, Equation (37) cannot be equal to 0. Assume to arrive at a contradiction that it is. Using Step 2 and the fact that $\bar{y}_2^S > 0$, we have $\mu_2^S + \lambda^S < 0$, which is a contradiction and shows that Equation (37) is strictly positive. This establishes that $\bar{l}_2^S = 0$.

We next show that $\bar{l}_1^S = 0$. If $\bar{y}_1^S = 0$ or $(\bar{l}_1^S)' = 0$, we are done by Equation (25). Assume that $\bar{y}_1^S > 0$ and $(\bar{l}_1^S)' > 0$. By Equation (39), this implies that $-\mu_1^S - \lambda^S + \gamma^S = 0$. Substituting this in Equation (40) and using $\theta^S \geq 0$ together with Equation (36), we obtain $\bar{l}_1^S = 0$.

*Step* 4. Since $\bar{l}_1^S = 0$, in view of Equation (30), we have $\bar{l}_1 = 0$. Moreover, since $\bar{y}_2^{OE} \geq \bar{y}_2^S + \epsilon$, by Equation (30), we have $\bar{l}_1' = 0$. Using in addition $\bar{l}_2^S = 0$, we see that for all $\epsilon > 0$ and $t = 2$,

$$\begin{aligned} r_{2,2}^{OE}(\epsilon) = \underset{\substack{l_2, l_2' \geq 0 \\ y_1^{OE}, y_2^{OE} \geq 0}}{\text{minimize}} \quad & 1 - \frac{l_2 y_2^{OE}}{R} \\ \text{subject to} \quad & l_2 \leq y_2^{OE} l_2', \\ & l_2 + y_2^{OE} l_2' \leq R, \\ & y_1^{OE} l_2' \geq R. \\ & \sum_{i=1}^{2} y_i^{OE} = 1, \end{aligned} \tag{41}$$

which follows by Equation (34). It is straightforward to show that the optimal solution of this problem is $(\bar{l}_2, \bar{l}_2', \bar{y}_1^{OE}, \bar{y}_2^{OE}) = (R/2, 3R/2, 2/3, 1/3)$ with optimal value 5/6. (One can write a similar optimization problem for $t = 1$ using the constraints in (33) and show that the optimal value is still 5/6.) Therefore it follows that $r_{2,t}^{OE}(\epsilon) = 5/6$ for $t = 1, 2$ and all $\epsilon > 0$. By Equation (24), this implies that

$$\inf_{\{l_i\} \in \mathcal{L}_2} \inf_{x^{OE} \in \overrightarrow{OE}(\{l_i\})} r_2(\{l_i\}, x^{OE}) \geq \frac{5}{6}.$$

We next show that this bound is tight. Consider the latency functions $l_1(x) = 0$ and $l_2(x) = (3/2)x$. As shown in Example 5.1, the corresponding OE flow vector is $x^{OE} = (2/3, 1/3)$, and the social optimum is $x^S = (1, 0)$. Hence the efficiency metric for these latency functions is $r_2(\{l_i\}, x^{OE}) = 5/6$, thus showing that the bound is tight. □

It is instructive to briefly consider the intuition underlying the 5/6 bound. The efficiency loss is maximized when as much of the traffic as possible goes through route 2 and when the latency on route 2 is as high as possible, i.e., when $x_2 l_2(x_2)$ is maximized. But these two requirements are in conflict in the sense that when the latency on route 2 is high, there will be less traffic on that route, because in a WE, we must have $p_1 + l_1(x_1) = p_2 + l_2(x_2)$. Moreover, with zero latency on route 1, equilibrium prices will satisfy $p_1 = x_1 l_2'(x_2)$ and $p_2 = x_2 l_2'(x_2)$. So the problem is to maximize $x_2 l_2(x_2)$, while satisfying $x_1 l_2'(x_2) = x_2 l_2'(x_2) + l_2(x_2)$. This constraint immediately implies that $x_1 > x_2$, and since $l_2(x_2) \le R$, the efficiency loss can never exceed $1/2$. But the bound is, in fact, much tighter than this. Since $x_1 > x_2$, convexity of $l_2$, i.e., a greater $l_2'$ given $l_2$, is harmful for the objective, since it tends to increase $x_1$ (as the inspection of the condition $x_1 l_2'(x_2) = x_2 l_2'(x_2) + l_2(x_2)$ shows). This reasoning suggests that the worst case will happen when $l_2$ is linear, which is exactly the case in our Example 5.1, leading to the efficiency loss of $1/6$ and the bound of $5/6$.

**5.1.2. Multiple links.** We next consider the general case where we have a parallel link network with $I$ links and $S$ service providers, and provider $s$ owns a set of links $\mathcal{I}_s \subset \mathcal{I}$. It can be seen by augmenting a two-link network with links that have latency functions

$$l(x) = \begin{cases} 0 & \text{if } x = 0, \\ \infty & \text{otherwise,} \end{cases}$$

that the lower bound in the general network case can be no higher than $5/6$. However, this is a degenerate example in the sense that at the OE, the flows of the links with latency functions given above are equal to 0. We next give an example of an $I$ link network, which has positive flows on all links at the OE and an efficiency metric of $5/6$.

Example 5.2. 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

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

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

$$x^{OE} = \left[\frac{2}{3}, \frac{1}{3(I-1)}, \dots, \frac{1}{3(I-1)}\right].$$

Hence the efficiency metric for this example is

$$r_I(\{l_i\}, x^{OE}) = \frac{5}{6}.$$

The next theorem generalizes Theorem 5.1 to a parallel link network with $I \ge 2$ links. The new feature here is not only the existence of more than two links, but also the fact that to find the worst-case bound, we have to optimize over the allocation of links across service providers. The strategy of the proof is again to reduce the infinite-dimensional problem to a finite-dimensional optimization problem, and then show that it reduces to the case in Theorem 5.1.

Theorem 5.2. *Consider a general parallel link network with $I$ links and $S$ service providers, where provider $s$ owns a set of links $\mathcal{I}_s \subset \mathcal{I}$. Then*

$$r_I(\{l_i\}, x^{OE}) \ge \frac{5}{6}, \qquad \forall \{l_i\}_{i\in\mathcal{I}} \in \mathcal{L}_I, \;\; x^{OE} \in \overrightarrow{OE}(\{l_i\}), \tag{42}$$

*and the bound is tight. In particular, there exists $\{l_i\}_{i\in\mathcal{I}} \in \mathcal{L}_I$ and $x^{OE} \in \overrightarrow{OE}(\{l_i\})$ that attains the lower bound in Equation (42).*

Proof. The proof again follows a number of steps.

*Step* 1. Consider the problem

$$\inf_{\{l_i\}\in\mathcal{L}_I} \inf_{x^{OE}\in\overrightarrow{OE}(\{l_i\})} r_I(\{l_i\}, x^{OE}). \tag{43}$$

Given $\{l_i\} \in \mathcal{L}_I$, let $x^{OE} \in \overrightarrow{OE}(\{l_i\})$ and let $x^S$ be a social optimum. By Lemmas 5.2 and 5.3, we can assume without loss of generality that $\sum_{i\in\mathcal{I}} x_i^{OE} = \sum_{i\in\mathcal{I}} x_i^S = 1$. Hence there exists some $i$ such that $x_i^{OE} < x_i^S$. Without loss of any generality, we restrict ourselves to the set of latency functions $\{l_i\}_{i\in\mathcal{I}} \in \mathcal{L}_I$ such that $x_1^{OE} < x_1^S$. Similar to the proof of Proposition 5.1, it can be seen that Problem (43) can be lower bounded by

$$\inf_{\{l_i\}\in\mathcal{L}_I} \inf_{x^{OE}\in\overrightarrow{OE}(\{l_i\})} r_I(\{l_i\}, x^{OE}) \ge \inf_{\epsilon>0} r_{I,t}^{OE}(\epsilon),$$

where $r_{I,t}^{OE}(\epsilon)$ is the optimum value of the following finite-dimensional problem, which we denote by $(\mathrm{E}_I^\epsilon)$:

$$(\mathrm{E}_I^\epsilon) \quad r_{I,t}^{OE}(\epsilon) = \underset{\substack{l_i^S,\,(l_i^S)'\geq 0\\ l_i,\,l_i'\geq 0\\ y_i^S,\,y_i^{OE}\geq 0\\ \mathcal{I}_s\subset\mathcal{I}}}{\text{minimize}} \quad \frac{R-\sum_{i\in\mathcal{I}} l_i y_i^{OE}}{R-\sum_{i\in\mathcal{I}} l_i^s y_i^S}$$

$$\text{subject to} \quad l_i^S \leq y_i^S (l_i^S)', \quad i=1,\ldots,I, \tag{44}$$

$$l_i^S + y_i^S (l_i^S)' = l_1^S + y_1^S (l_1^S)', \quad i=2,\ldots,I,$$

$$l_1^S + y_1^S (l_1^S)' \leq R,$$

$$\sum_{i\in\mathcal{I}} y_i^S \leq 1$$

$$l_1 + l_1'\left(\sum_{i\neq 1} y_i^{OE} - \sum_{i\neq 1} y_i^S\right) \leq l_1^S, \tag{45}$$

$$\sum_{i\neq 1} y_i^{OE} \geq \sum_{i\neq 1} y_i^S + \epsilon, \tag{46}$$

$$\sum_{i\in\mathcal{I}} y_i^{OE} = 1,$$

$$\mathcal{I}_s = \{1\} \quad \text{for some } s \quad \text{if } l_1^S = 0, \tag{47}$$

$$+\{\text{OE constraints}\}_t, \quad t=1,2.$$

The new feature relative to the two-link case is the presence of $\mathcal{I}_s$'s as choice variables to allow a choice over possible distribution of links across service providers (with the constraint $\bigcup_s \mathcal{I}_s = \mathcal{I}$ left implicit). The OE constraints, which are again written separately for $t=1,2$ for the two cases in Proposition 4.4, depend on $\mathcal{I}_i$'s. In addition, we have added Constraint (47) to impose Assumption 2 (recall that $x_1^{OE} > 0$ by Lemma 4.2).

*Step* 2. Let $(\bar{l}_i^S, (\bar{l}_i^S)', \bar{l}_i, \bar{l}_i', \bar{y}_i^S, \bar{y}_i^{OE})$ be an optimal solution of the preceding problem. Note that the constraints that involve $(l_i^S, (l_i^S)', y_i^S)$ for $i=2,\ldots,I$ are decoupled and have the same structure as in problem $(\mathrm{E}^\epsilon)$. Therefore, by the same argument used to show $\bar{l}_2^S = 0$ in Step 3 of the proof of Proposition 5.1, one can show that $\bar{l}_i^S = 0$ for each $i=2,\ldots,I$. Similarly, one can extend the same argument given in the proof of Proposition 5.1 to show that $\bar{l}_1^S = 0$.

*Step* 3. Since $\bar{l}_1^S = 0$, it follows that $\bar{l}_1 = 0$ and $\bar{l}_1' = 0$ (cf. Equations (45) and (46)), and also $\mathcal{I}_s = \{1\}$. Therefore, using the price characterization from Proposition 4.4, the structure of the problem simplifies to

$$r_{I,t}^{OE} \geq \underset{\substack{l_i,\,l_i'\geq 0,\, i=2,\ldots,I\\ y_i^{OE}\geq 0,\, i=1,\ldots,I}}{\text{minimize}} \quad 1 - \frac{\sum_{i=2}^I l_i y_i^{OE}}{R} \tag{48}$$

$$\text{subject to} \quad l_i \leq y_i^{OE} l_i', \quad i=2,\ldots,I, \tag{49}$$

$$l_i + y_i^{OE} l_i' \leq R, \quad i=2,\ldots,I, \tag{50}$$

$$\frac{y_1^{OE}}{\sum_{j=2}^I 1/l_j'} \geq R, \tag{51}$$

$$\sum_{i\in\mathcal{I}} y_i^{OE} = 1,$$

where we have also used the fact that $\bar{l}_i^S = 0$ for $i=2,\ldots,I$.

The first set of constraints are because of the convexity assumptions on the $l_i$. Similar to the two-link case, the second set of constraints are because of the OE constraints (given $\bar{l}_1' = 0$, see the OE price characterization in Proposition 4.4).

*Step* 4. Let $((\bar{l}_i, \bar{l}_i')_{i=2,\ldots,I}, (\bar{y}_i^{OE})_{i=1,\ldots,I})$ denote an optimal solution of the preceding problem. Assign the Lagrange multipliers $\mu_i \geq 0$, $\lambda_i \geq 0$, $i=2,\ldots,I$, and $\gamma \geq 0$ and $\theta$ consecutively to the constraints of the problem. We will show that $\bar{l}_i = \bar{y}_i^{OE}\bar{l}_i'$ for all $i = 2, \ldots, I$. Assume the contrary, i.e., $\bar{l}_i < \bar{y}_i^{OE}\bar{l}_i'$ for some $i = 2, \ldots, I$. This implies that $\mu_i = 0$. Using the optimality conditions, we have

$$-\frac{\gamma}{\sum_{j=2}^{I} 1/\bar{l}_j'} + \theta = 0 \quad \text{if } \bar{y}_1^{OE} > 0 \qquad \geq 0 \quad \text{if } \bar{y}_1^{OE} = 0, \tag{52}$$

$$-\frac{\bar{y}_i^{OE}}{R} + \lambda_i = 0 \quad \text{if } \bar{l}_i > 0 \qquad \geq 0 \quad \text{if } \bar{l}_i = 0, \tag{53}$$

$$-\frac{\bar{l}_i}{R} + \lambda_i \bar{l}_i' + \theta = 0 \quad \text{if } \bar{y}_i^{OE} > 0 \qquad \geq 0 \quad \text{if } \bar{y}_i^{OE} = 0. \tag{54}$$

By feasibility (cf. Equation (51)), we have $\bar{y}_1^{OE} > 0$. Moreover, by our assumption $[\bar{l}_i < \bar{y}_i^{OE}\bar{l}_i']$, we have $\bar{y}_i^{OE} > 0$ and $\bar{l}_i' > 0$. Equation (52) implies that $\theta \geq 0$. We also have from Equation (53) that $\lambda_i \geq \bar{y}_i^{OE}/R$, which when substituted in Equation (54), yields a contradiction in view of $\theta \geq 0$.

Hence, for all $i = 2, \ldots, I$, we have $\bar{l}_i = \bar{y}_i^{OE}\bar{l}_i'$. It is also straightforward to see that Constraint (50) is binding at the optimal solution (otherwise it would be possible to decrease the objective function value), which implies that $\bar{l}_i = R/2$ and $\bar{y}_i^{OE}\bar{l}_i' = R/2$ for all $i = 2, \ldots, I$. By using the transformation of variables

$$y^{OE} = \sum_{i=2}^{I} y_i^{OE} \qquad \text{and} \qquad \frac{1}{l'} = \sum_{i=2}^{I} \frac{1}{l_i'},$$

it can be seen that the optimal value of Problem (48) is the same as the optimal value of the following problem:

$$\begin{aligned} \underset{\substack{l,\, l' \geq 0 \\ y_1^{OE},\, y^{OE} \geq 0}}{\text{minimize}} \quad & 1 - \frac{l y^{OE}}{R} \\ \text{subject to} \quad & l \leq y^{OE} l', \\ & l + y^{OE} l' \leq R, \\ & y_1^{OE} l' \geq R, \\ & y_1^{OE} + y^{OE} = 1. \end{aligned}$$

which is identical to Problem (41) in the two-link case, showing that for all $t = 1, 2$ and all $\epsilon > 0$,

$$r_{I,t}^{OE}(\epsilon) \geq \frac{5}{6}.$$

Hence we have

$$\inf_{\{l_i\}_{i\in\mathcal{I}} \in \mathcal{L}_I} \; \inf_{x^{OE} \in \overrightarrow{OE}(\{l_i\})} r_I(\{l_i\}, x^{OE}) \geq \frac{5}{6}.$$

Finally, Example 5.2 shows that the preceding bound is tight. □

A notable feature of Example 5.2 and this theorem is that the (tight) lower bound on inefficiency is independent of the number of links $I$. Thus arbitrarily large networks can feature as much inefficiency as small networks.[12]

[12] This result superficially contrasts with theorems in the economics literature that large oligopolistic markets approach competitive behavior (e.g., Roberts and Postlewaite [41], Hart [17], Novshek [35], Vives [54, 55]). These theorems do not consider arbitrary large markets, but replicas of a given market structure. In our model as well, if we take a given network and replicate it $n$ times (i.e., increase $d$ to $nd$ and the number of service providers by $n$), then as $n \to \infty$, the efficiency metric tends to 1. In fact, in Example 5.1, replicating the network once, i.e., $n = 2$, achieves full efficiency, because of Bertrand competition between two oligopolists with zero latencies.

**5.2. Mixed-strategy equilibria.** As we illustrated in §4, pure strategy OE may fail to exist (cf. Example 4.1). Nevertheless, as shown in Proposition 4.3, such games always have a mixed-strategy equilibrium. In this section, we discuss the efficiency properties of mixed-strategy OE.

Although there has been much less interest in the efficiency properties of mixed-strategy equilibria, two different types of efficiency metrics present themselves as natural candidates. The first considers the worst realization of the strategies, while the second focuses on average inefficiency across different realizations of mixed strategies. We refer to the first metric as worst realization metric, and denote it by $\tilde{r}_I^W(\{l_i\})$, and to the second as the average metric, and denote it by $\tilde{r}_I^A(\{l_i\})$.

Given a set of latency functions $\{l_i\}_{i\in\mathcal{I}}$, let $OM(\{l_i\})$ denote the set of mixed-strategy equilibria. For some $\mu \in OM(\{l_i\})$, let $M_i(\mu)$ denote the support of $\mu_i$ as defined before in Example 4.1 (in particular, recall that $M_i = \{p \mid \mu_i'(p) > 0\}$). Further, let

$$\overrightarrow{OE}_m(\{l_i\}, \mu) = \{x \mid x \in W(p), \text{ for some } p \text{ s.t. } p_i \in M_i(\mu) \text{ for all } i\}.$$

We define the worst realization efficiency metric as

$$\tilde{r}_I^W(\{l_i\}) = \inf_{\{l_i\}\in\mathcal{L}_I} \inf_{\mu\in OM(\{l_i\})} \inf_{x^{OE}\in\overrightarrow{OE}_m(\{l_i\},\mu)} r_I(\{l_i\}, x^{OE}),$$

where $r_I$ is given by Equation (21).

Similarly, the average efficiency metric is defined as

$$\tilde{r}_I^A(\{l_i\}) = \inf_{\{l_i\}\in\mathcal{L}_I} \inf_{\mu\in OM(\{l_i\})} \int \cdots \int r_I(\{l_i\}, x^{OE}(p))\, d\mu_1 \cdots d\mu_S.$$

In the next example, we show that the worst realization efficiency metric for games with no pure strategy equilibrium can be arbitrarily low.

Example 4.1 (continued). Consider the prices $p_1 = R$ and $p_2 = R(1-\delta)$ that satisfy $p_i \in M_i$ for the unique mixed-strategy equilibrium given in Example 4.1 as $\epsilon \to 0$. The WE at these prices is given by

$$x^{OE} = (1-\delta, \delta),$$

and the worst realization efficiency metric is

$$\tilde{r}_I^W(\{l_i\}) = 1-\delta^2,$$

which as $\delta \to 1$ goes to 0.

On the other hand, as $\epsilon \to 0$, the average efficiency metric, $\tilde{r}_I^A(\{l_i\})$ is given by

$$\tilde{r}_I^A(\{l_i\}) = \int_{(1-\delta)R}^{R} \int_{(1-\delta)R}^{R} r(p_1, p_2)\, d\mu_1 \times d\mu_2,$$

where $r(p_1, p_2)$ is the inefficiency at the price vector $(p_1, p_2)$ at the unique mixed-strategy OE characterized above. Therefore

$$r(p_1, p_2) = \begin{cases} 1 & \text{if } p_1 < p_2 \\ 1 - \dfrac{\delta(p_1 - p_2)}{R} & \text{if } p_1 > p_2, \end{cases}$$

and thus

$$\tilde{r}_I^A(\{l_i\}) \to 1 - \int_{(1-\delta)R}^{R} \int_{(1-\delta)R}^{R} \frac{\delta(p_1 - p_2)}{R}\, d\mu_1 \times d\mu_2.$$

Thus, to calculate $\tilde{r}_I^A(\{l_i\})$, we need to compute the last integral. Denoting this by $A$, we have

$$\begin{aligned} A &= \frac{\delta}{R} \int_{(1-\delta)R}^{R} \int_{p_2}^{R} (p_1 - p_2)\, d\mu_1 \times d\mu_2 \\ &= \frac{\delta}{R}\left[\int_{(1-\delta)R}^{R} \int_{(1-\delta)R}^{p_1} p_1\, d\mu_2 \times d\mu_1 - \int_{(1-\delta)R}^{R} \int_{p_2}^{R} p_2\, d\mu_1 \times d\mu_2\right] \\ &= \frac{\delta}{R}\left[\int_{(1-\delta)R}^{R} p_1 \left(\frac{1}{\delta} - \frac{(1-\delta)R}{\delta p_1}\right) d\mu_1(p_1) - \int_{(1-\delta)R}^{R} (1-\delta)R\, d\mu_2(p_2)\right]. \end{aligned}$$

Now, recall that $\mu_1$ has an atom equal to $1-\delta$ at $R$, so

$$\begin{aligned} A &= \frac{\delta}{R}\left[R(1-\delta)+\int_{(1-\delta)R}^{R} p_1\left(\frac{1}{\delta}-\frac{(1-\delta)R}{\delta p_1}\right)\left(\frac{(1-\delta)R}{p_1^2}\right)dp_1 - R(1-\delta)\right] \\ &= (1-\delta)^2-(1-\delta)+(1-\delta)[\ln R-\ln((1-\delta)R)] \\ &= -(1-\delta)\delta-(1-\delta)\ln(1-\delta). \end{aligned}$$

It can be calculated that $A$ reaches a maximum of approximately 0.16 for $\delta\approx 0.8$. Therefore, in this example, $\tilde{r}_I^A(\{l_i\})$ reaches $0.84\approx 5/6$ (in fact, slightly greater than 5/6). We conjecture, but are unable to prove, that 5/6 is also a lower bound for the average efficiency metric, $\tilde{r}_I^A(\{l_i\})$, in mixed-strategy OE. This is left as an open research question.[13]

**6. Bound for positive latency at zero flow.** In this section, we relax the assumption $l_i(0)=0$, and allow positive latency at zero flow. To simplify the exposition in this section, we focus on continuously differentiable latency functions, but as our previous analysis indicates, the main result, Theorem 6.1, holds for general convex latency functions.

Assumption 6.1. *For each $i\in\mathcal{I}$, the latency function $l_i$: $[0,\infty)\mapsto[0,\infty)$ is continuously differentiable, convex, and nondecreasing.*

We first provide an equilibrium price characterization, which generalizes Corollary 4.1.

Proposition 6.1. *Let $(p^{OE},x^{OE})$ be an OE such that $p_i^{OE}x_i^{OE}>0$ for some $i\in\mathcal{I}$. Define the index set*

$$\mathcal{N}=\{j\in\mathcal{I}\mid p_i^{OE}+l_i(x_i^{OE})<p_j^{OE}+l_j(0)\}. \tag{55}$$

*Let Assumptions 4.1 and 6.1 hold. Then, for all $s\in\mathcal{S}$ and $i\in\mathcal{I}_s$ and $i\notin\mathcal{N}$, we have*

$$p_i^{OE}=\begin{cases} x_i^{OE}l_i'(x_i^{OE}), & \text{if } l_j'(x_j^{OE})=0 \text{ for some } j\notin\mathcal{I}_s,\ j\notin\mathcal{N} \\ \min\left\{R-l_i(x_i^{OE}),\ x_i^{OE}l_i'(x_i^{OE})+\dfrac{\sum_{j\in\mathcal{I}_s}x_j^{OE}}{\sum_{j\notin\mathcal{I}_s,\,j\notin\mathcal{N}}1/l_j'(x_j^{OE})}\right\}, & \text{otherwise}. \end{cases} \tag{56}$$

The proof of this theorem follows immediately from the proof of Corollary 4.1. In particular, $\mathcal{N}$ is the set of all latencies where $x_i^{OE}=0$, so that any $i\in\mathcal{N}$ can be discarded when considering the individual optimization problem of each service provider. In what follows, let $\mathcal{L}_I^*$ denote the set of latency functions for which the associated price competition game has a pure strategy OE and the individual $l_i$'s satisfy Assumptions 4.1 and 6.1.

Theorem 6.1. *Consider a general parallel link network with $I$ links and $S$ service providers, where provider $s$ owns a set of links $\mathcal{I}_s\subset\mathcal{I}$. Then*

$$r_I(\{l_i\},\ x^{OE})\ge 2\sqrt{2}-2,\quad \forall\,\{l_i\}_{i\in\mathcal{I}}\in\mathcal{L}_I^*,\quad x^{OE}\in\overrightarrow{OE}(\{l_i\}), \tag{57}$$

*and the bound is tight, i.e., there exists $\{l_i\}_{i\in\mathcal{I}}\in\mathcal{L}_I^*$ and $x^{OE}\in\overrightarrow{OE}(\{l_i\})$ that attains the lower bound in Equation (57).*

Proof. The proof follows those of Theorems 5.1 and 5.2 closely. Once again, the Problem (23) is lower bounded by a modified version of the finite-dimensional problem ($\mathrm{E}_I^\epsilon$) (see the proof of Theorem 2), in which we introduce additional variables $l_i^0\ge 0$, which represent the value of the latency function, $l_i(\cdot)$ at 0. Using the convexity of the latency functions, we replace Constraint (44) by

$$l_i^S\le y_i^S(l_i^S)'+l_i^0.$$

Following the same line of argument, it can be seen that Problem (23) can further be bounded below by a problem identical to (48) except that Constraint (49) is replaced by

$$l_i\le y_i^{OE}l_i'+l_i^0.$$

[13] As pointed out by one of our anonymous referees, the intuition provided following Theorem 5.1 suggests that even in the case of a mixed-strategy equilibrium, the average efficiency metric should not fall below 1/2. Nevertheless, proving this conjecture has not been possible because the equilibrium conditions for a mixed-strategy OE are considerably more involved than those for a pure strategy only.

Using a similar transformation, this problem can be seen to be equivalent to

$$\begin{aligned} \underset{\substack{l,\,l',\,l^0\geq 0\\ y_1^{OE},\,y^{OE}\geq 0}}{\text{minimize}} \quad & 1-\frac{ly^{OE}}{R} \\ \text{subject to} \quad & l\leq y^{OE}l'+l^0, \\ & l+y^{OE}l'\leq R, \\ & y_1^{OE}l'\geq R, \\ & y_1^{OE}+y^{OE}=1. \end{aligned}$$

The optimal solution of this problem is $(\bar{l},\bar{l}',\bar{l}^0,\bar{y}_1^{OE},\bar{y}^{OE})=(2-\sqrt{2},\sqrt{2},3-2\sqrt{2},\sqrt{2}/2,1-\sqrt{2}/2)$ and the corresponding optimal value is $2\sqrt{2}-2$.

We next show that this bound is tight. Consider an $I$ link parallel 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

$$l_1(x)=0 \qquad \text{and} \qquad l_i(x)=(I-1)\sqrt{2}x+(3-2\sqrt{2}), \quad \forall\, i=2,\ldots,I.$$

The corresponding OE flow vector is

$$x^{OE}=\left[\frac{\sqrt{2}}{2},\frac{1}{I-1}\left(1-\frac{\sqrt{2}}{2}\right),\ldots,\frac{1}{I-1}\left(1-\frac{\sqrt{2}}{2}\right)\right],$$

and the social optimum is $x^S=(1,0)$. Hence the efficiency metric for these latency functions is $r_I(\{l_i\},x^{OE})=2\sqrt{2}-2$, thus showing that

$$\min_{\{l_i\}_{i\in\mathcal{I}}\in\mathcal{L}_I^*}\ \min_{x^{OE}\in\overrightarrow{OE}(\{l_i\})} r_I(\{l_i\},x^{OE})=2\sqrt{2}-2. \quad \square$$

It is interesting to note that $2\sqrt{2}-2\approx 0.828\leq 5/6$. Therefore, relaxing the assumption $l_i(0)=0$ has a small effect on the worst-case performance of a pure strategy OE. In terms of the intuition, we provided for Theorem 5.1, the fact that $l_2(0)$ can be positive allows us to increase $l_2$ slightly for a given $l_2'$, leading to a small deterioration in performance.

**7. Conclusions.** In this paper, we presented an analysis of competition in congested networks. We established a number of results. First, despite the potential inefficiencies of flow routing without prices, price setting by a monopolist always achieves the social optimum. Second, and in contrast to the monopoly result, OE where multiple service providers compete are typically inefficient. Third and most importantly, when latency at zero flow is zero, there is a tight bound of $5/6$ on inefficiency in pure strategy OE. When latency at zero flow can be positive, the bound is slightly lower at $2\sqrt{2}-2\approx 0.828$. These bounds apply even for arbitrarily large parallel link networks.

A number of concluding comments are useful:

- Our motivating example has been the flow of information in a communication network, but our results apply equally to traffic assignment problems and oligopoly in product markets with negative externalities, congestion, or snob effects (as originally suggested by Veblen [52]).
- Our analysis has been quite general, in particular, allowing for constant latencies and capacity constraints. Some of the analysis simplifies considerably when we specialize the network to increasing and real-valued (noncapacity constrained) latencies.
- On the other hand, our analysis has been simplified by our focus on parallel link networks. We have started extending this analysis in ongoing work (Acemoglu and Ozdaglar [2]) for topologies consisting of parallel-serial structure. This parallel-serial topology, however, rules out many interesting cases, including those that could potentially lead to Braess' paradox, and the analysis for more general topologies is an open area for future research.
- One simplifying feature of our analysis is the assumption that users are *homogeneous* in the sense that the same reservation utility, $R$, applies to all users. It is possible to conduct a similar analysis with elastic and heterogeneous users (or traffic), but this raises a number of new and exciting challenges. For example, monopoly or oligopoly providers might want to use nonlinear pricing (designed as a mechanism subject to incentive compatibility constraints of different types of users, e.g., Wilson [57]). This is an important research area for understanding equilibria in communication networks, where users often have heterogeneous quality of service requirements.
- While we have established that worst realization efficiency metric in mixed-strategy OE can be arbitrarily low, a bound for average efficiency metric is an open research question.

**Appendix A. Proof of Proposition 3.1.** If $(p^{ME}, x^{ME})$ is an ME, then it is an SPE by definition. Let $(p^{ME}, x^{ME})$ be an SPE. Assume to arrive at a contradiction that there exists some $p \geq 0$ and $\tilde{x} \in W(p)$ such that

$$\Pi(p^{ME}, x^{ME}) < \Pi(p, \tilde{x}). \tag{A1}$$

If $W(p)$ is a singleton, we immediately obtain a contradiction. Assume that $W(p)$ is not a singleton and $\sum_{i\in\mathcal{I}} x_i = \sum_{i\in\mathcal{I}} \bar{x}_i$ for all $x, \bar{x} \in W(p)$. By Lemma 2.2, it follows that $\Pi(p, \tilde{x}) = \Pi(p, x)$ for all $x \in W(p)$, which contradicts the fact that $(p^{ME}, x^{ME})$ is an SPE.

Assume finally that $W(p)$ is not a singleton and

$$\sum_{i\in\mathcal{I}} \hat{x}_i < \sum_{i\in\mathcal{I}} \bar{x}_i \quad \text{for some } \hat{x}, \bar{x} \in W(p). \tag{A2}$$

For this case, we have $p_i = R$ for all $i \in \bar{\mathcal{I}}$, where

$$\bar{\mathcal{I}} = \{i \in \mathcal{I} \mid \exists x,\ \bar{x} \in W(p) \text{ with } x_i \neq \bar{x}_i\},$$

(cf. Equation (5)). To see this, note that since $\sum_{i\in\mathcal{I}} \hat{x}_i < d$, the WE optimality conditions for $\hat{x}$ (cf. Equation (2)) hold with $\lambda = 0$. Assume that $\tilde{p} < R$. By Lemma 2.2, $l_i(x_i) = 0$ for all $i \in \bar{\mathcal{I}}$. If $b_{C_i} = \infty$ for some $i \in \bar{\mathcal{I}}$, we get a contradiction by Equation (2). Otherwise, Equation (2) implies that $\hat{x}_i = b_{C_i}$ for all $i \in \bar{I}$. Since $\hat{x}_i = \bar{x}_i$ for all $i \notin \bar{\mathcal{I}}$, this contradicts Equation (A2).

We show that given $\delta > 0$, there exists some $\epsilon > 0$ such that

$$\Pi(p^{\epsilon}, x^{\epsilon}) \geq \Pi(p, \tilde{x}) - \delta, \quad \forall\, x^{\epsilon} \in W(p^{\epsilon}), \tag{A3}$$

where

$$p_i^{\epsilon} = \begin{cases} p_i & i \notin \bar{\mathcal{I}}, \\ R - \epsilon & i \in \bar{\mathcal{I}}. \end{cases} \tag{A4}$$

The preceding relation together with Equation (A1) contradicts the fact that $(p^{ME}, x^{ME})$ is an SPE, thus establishing our claim.

We first show that

$$\sum_{i\in\bar{\mathcal{I}}} x_i^{\epsilon} \geq \sum_{i\in\bar{\mathcal{I}}} \tilde{x}_i. \tag{A5}$$

Assume to arrive at a contradiction that

$$\sum_{i\in\bar{\mathcal{I}}} x_i^{\epsilon} < \sum_{i\in\bar{\mathcal{I}}} \tilde{x}_i. \tag{A6}$$

This implies that there exists some $j \in \bar{\mathcal{I}}$ such that $x_j^{\epsilon} < \tilde{x}_j$ (which also implies that $x_j^{\epsilon} < b_{C_j}$). We use the WE optimality conditions (Equation (2)) for $\tilde{x}$ and $x^{\epsilon}$ to obtain the following:

- There exists some $\tilde{\lambda} \geq 0$ such that for some $i \in \bar{\mathcal{I}}$,

$$R - l_i(\tilde{x}_i) - p_i = 0 \geq \tilde{\lambda},$$

where we used the facts that $l_i(\tilde{x}_i) = 0$, $p_i = R$ (cf. Lemma 2.2) and $\tilde{x}_i > 0$ for some $i \in \bar{\mathcal{I}}$ (cf. Equation (A1)). Since $\tilde{\lambda} = 0$, we have, for all $i \notin \bar{\mathcal{I}}$,

$$\begin{aligned} R - l_i(\tilde{x}_i) - p_i &\leq 0 \quad \text{if } \tilde{x}_i \leq b_{C_i}, \\ &\geq 0 \quad \text{if } \tilde{x}_i = b_{C_i}. \end{aligned} \tag{A7}$$

- There exists some $\lambda^{\epsilon} \geq 0$ such that

$$\epsilon - l_j(x_j^{\epsilon}) \leq \lambda^{\epsilon} \tag{A8}$$

(since $x_j^{\epsilon} < \tilde{x}_j$ and $p_j = R - \epsilon$), and for all $i \notin \bar{\mathcal{I}}$,

$$\begin{aligned} R - l_i(x_i^{\epsilon}) - p_i &\leq \lambda^{\epsilon} \quad \text{if } x_i^{\epsilon} = 0, \\ &\geq \lambda^{\epsilon} \quad \text{if } x_i^{\epsilon} > 0. \end{aligned} \tag{A9}$$

If $\lambda^{\epsilon} = 0$, then by Equation (A8) and the fact that $l_j(\tilde{x}_j) = 0$ (Lemma 2.2), we obtain

$$l_j(x_j^{\epsilon}) \geq \epsilon > 0 = l_j(\tilde{x}_j),$$

which is a contradiction. If $\lambda^{\epsilon} > 0$, then $\sum_{i\in\mathcal{I}} x_i^{\epsilon} = d$. Assume first that $x_i^{\epsilon} \leq \tilde{x}_i$ for all $i \notin \bar{\mathcal{I}}$. Then

$$\sum_{i\in\bar{\mathcal{I}}} x_i^{\epsilon} = d - \sum_{i\notin\bar{\mathcal{I}}} x_i^{\epsilon} \geq d - \sum_{i\notin\bar{\mathcal{I}}} \tilde{x}_i \geq \sum_{i\in\bar{\mathcal{I}}} \tilde{x}_i,$$

which yields a contradiction by Equation (A6). Assume next that $x_k^\epsilon > \tilde{x}_k$ for some $k \notin \bar{\mathcal{I}}$. By Equations (A7) and (A9), we have

$$R - l_k(x_k^\epsilon) - p_k \geq \lambda^\epsilon, \qquad R - l_k(\tilde{x}_k) - p_k \leq 0,$$

which together implies that $l_k(\tilde{x}_k) > l_k(x_k^\epsilon)$, yielding a contradiction and proving Equation (A5).

Since $W(p)$ is an upper semicontinuous correspondence and the $i$th component of $W(p)$ is uniquely defined for all $i \notin \bar{\mathcal{I}}$, it follows that $x_i(\cdot)$ is continuous at $p$ for all $i \notin \bar{\mathcal{I}}$. Together with Equation (5), this implies that

$$\begin{aligned}
\Pi(p^\epsilon, x^\epsilon) &= \sum_{i\notin\bar{\mathcal{I}}} p_i\tilde{x}_i + \sum_{i\notin\bar{\mathcal{I}}} p_i(x_i^\epsilon - \tilde{x}_i) + \sum_{i\in\bar{\mathcal{I}}}(R-\epsilon)x_i^\epsilon \\
&\geq \sum_{i\notin\bar{\mathcal{I}}} p_i\tilde{x}_i + \sum_{i\in\bar{\mathcal{I}}}(R-\epsilon)\tilde{x}_i + \sum_{i\notin\bar{\mathcal{I}}} p_i(x_i^\epsilon - \tilde{x}_i) \\
&= \sum_{i\in\mathcal{I}} p_i\tilde{x}_i - \epsilon\sum_{i\in\bar{\mathcal{I}}}\tilde{x}_i + \sum_{i\notin\bar{\mathcal{I}}} p_i(x_i^\epsilon - \tilde{x}_i) \\
&\geq \sum_{i\in\mathcal{I}} p_i\tilde{x}_i - \delta,
\end{aligned}$$

where the last inequality holds for sufficiently small $\epsilon$, establishing (A3), and completing the proof.

**Appendix B. Proof of Proposition 4.2.** For all $i \in \mathcal{I}$, let $l_i(x) = a_i x$. Define the set

$$\mathcal{I}_0 = \{i \in \mathcal{I} \mid a_i = 0\}.$$

Let $I_0$ denote the cardinality of set $\mathcal{I}_0$. There are two cases to consider:

*Case* 1. $I_0 \geq 2$: Assume that there exist $i, j \in \mathcal{I}_0$ such that $i \in \mathcal{I}_s$ and $j \in \mathcal{I}_{s'}$ for some $s \neq s' \in \mathcal{S}$. Then, it can be seen that a vector $(p^{OE}, x^{OE})$ with $p_i^{OE} = 0$ for all $i \in \mathcal{I}_0$ and $x^{OE} \in W(p^{OE})$ is an OE. Assume next that for all $i \in \mathcal{I}_0$, we have $i \in \mathcal{I}_s$ for some $s \in \mathcal{S}$. Then, we can assume without loss of generality that provider $s$ owns a single link $i'$, which has $a_{i'} = 0$ and consider the case $I_0 = 1$.

*Case* 2. $I_0 \leq 1$: Let $B_s(p_{-s}^{OE})$ be the set of $p_s^{OE}$ such that

$$(p_s^{OE}, x^{OE}) \in \arg\max_{\substack{p_s\geq 0\\ x\in W(p_s,\, p_{-s}^{OE})}} \sum_{i\in\mathcal{I}_s} p_i x_i. \tag{A10}$$

Let $B(p^{OE}) = [B_s(p_{-s}^{OE})]_{s\in\mathcal{S}}$. By the theorem of the maximum (Berge [7]), it follows that $B(p^{OE})$ is an upper semicontinuous correspondence. We next show that it is convex-valued.

Lemma B.1. *For all $s \in \mathcal{S}$ and $p_{-s}^{OE} \geq 0$, the set $B_s(p_{-s}^{OE})$ is a convex set.*

Proof. For some $s \in \mathcal{S}$ and $p_{-s}^{OE} \geq 0$, let $p_s \in B_s(p_{-s}^{OE})$ and $\bar{p}_s \in B_s(p_{-s}^{OE})$ such that $(p_s, x)$ and $(\bar{p}_s, \bar{x})$ are optimal solutions of Problem (A10). Denote $x_s = [x_i]_{i\in\mathcal{I}_s}$ and $\bar{x}_s = [\bar{x}_i]_{i\in\mathcal{I}_s}$. If $p_s^T x_s = \bar{p}_s^T \bar{x}_s = 0$, then the vector $\gamma p_s + (1-\gamma)\bar{p}_s \in B_s(p_{-s}^{OE})$ for all $\gamma \in [0, 1]$, and we are done.

Assume that $p_s^T x_s = \bar{p}_s^T \bar{x}_s > 0$. We will show that $p_s = \bar{p}_s$. Using a similar argument as in the proof of Lemma 4.2(b), it follows that $x_i > 0$ for all $i \in \mathcal{I}_s$, and $\bar{x}_i > 0$ for all $i \in \mathcal{I}_s$. By checking the first-order optimality conditions of Problem (A10) and using the linearity of the latency functions, it can be seen that $p_i = p$ for all $i \in \mathcal{I}_s$, and $\bar{p}_i = \bar{p}$ for all $i \in \mathcal{I}_s$. Assume to arrive at a contradiction that $p > \bar{p}$. Since $p_s^T x_s = \bar{p}_s^T \bar{x}_s$, this implies that $\sum_{i\in\mathcal{I}_s} x_i < \sum_{i\in\mathcal{I}_s} \bar{x}_i$. There are two cases to consider:

- $p_i + a_i x_i < \bar{p}_i + a_i\bar{x}_i$ for $i \in \mathcal{I}_s$:

Since $p_i + a_i x_i < R$, it follows by the definition of a WE that $\sum_{i\in\mathcal{I}} x_i = d$. One can also immediately see that $\bar{x}_j \geq x_j$ for all $j \notin \mathcal{I}_s$. Together with $\sum_{i\in\mathcal{I}_s} x_i < \sum_{i\in\mathcal{I}_s} \bar{x}_i$, this implies that $\sum_{i\in\mathcal{I}} x_i < \sum_{i\in\mathcal{I}} \bar{x}_i$, which contradicts the fact that $\sum_{i\in\mathcal{I}} x_i = d$.

- $p_i + a_i x_i \geq \bar{p}_i + a_i\bar{x}_i$ for $i \in \mathcal{I}_s$:

If both effective costs are equal to $R$, i.e.,

$$p_i + a_i x_i = \bar{p}_i + a_i\bar{x}_i = R, \quad \forall\, i \in \mathcal{I}_s,$$

then the optimization problem of provider $i$ (cf. Problem (A10)) can be shown to have a strictly concave objective function over polyhedral constraints, thus implying that $p = \bar{p}$.

Assume next that $p_i + a_i x_i \le R$ and $\bar{p}_i + a_i \bar{x}_i < R$. Define the sets

$$\mathcal{N} = \{j \notin \mathcal{I}_s \mid p_i + l_i(x_i) < p_j^{OE}\}, \qquad \bar{\mathcal{N}} = \{j \notin \mathcal{I}_s \mid \bar{p}_i + l_i(\bar{x}_i) < p_j^{OE}\}.$$

The first-order conditions of Problem (A10) in this case yields

$$p_i \le a_i x_i + \frac{\sum_{j\in\mathcal{I}_s} x_j}{\sum_{j\notin\mathcal{I}_s,\, j\notin\mathcal{N}} 1/a_j}, \quad \forall\, i \in \mathcal{I}_s,$$
$$\bar{p}_i = a_i \bar{x}_i + \frac{\sum_{j\in\mathcal{I}_s} \bar{x}_j}{\sum_{j\notin\mathcal{I}_s,\, j\notin\bar{\mathcal{N}}} 1/a_j}, \quad \forall\, i \in \mathcal{I}_s.$$

Moreover, in view of the relation between the effective costs, it can be seen that $\mathcal{N} \subset \bar{\mathcal{N}}$. Since $\sum_{i\in\mathcal{I}_s} x_i < \sum_{i\in\mathcal{I}_s} \bar{x}_i$, we have $x_i < \bar{x}_i$ for some $i \in \mathcal{I}_s$, which by the preceding implies that $p_i < \bar{p}_i$, yielding a contradiction. ☐

Proof of Proposition 4.2. Since $B(p^{OE})$ is an upper semicontinuous and convex-valued correspondence, we can use Kakutani's fixed-point theorem to assert the existence of a $p^{OE}$ such that $B(p^{OE}) = p^{OE}$ (see Berge [7]). To complete the proof, it remains to show that there exists $x^{OE} \in W(p^{OE})$ such that Equation (10) holds.

If $\mathcal{I}_0 = \varnothing$, we have by Proposition 2.3 that $W(p^{OE})$ is a singleton, and therefore Equation (10) holds and $(p^{OE}, W(p^{OE}))$ is an OE.

Assume finally that exactly one of the $a_i$'s (without loss of generality $a_1$) is equal to 0. We show that for all $\bar{x}, \tilde{x} \in W(p^{OE})$, we have $\bar{x}_i = \tilde{x}_i$ for all $i \ne 1$. Let $EC(x, p^{OE}) = \min_j\{l_j(x_j) + p_j^{OE}\}$. If at least one of

$$EC(\tilde{x}, p^{OE}) < R \qquad \text{or} \qquad EC(\bar{x}, p^{OE}) < R$$

holds, then one can show that $\sum_{i\in\mathcal{I}} \tilde{x}_i = \sum_{i\in\mathcal{I}} \bar{x}_i = d$. Substituting $x_1 = d - \sum_{i\in\mathcal{I},\, i\ne 1} x_i$ in Problem (4), we see that the objective function of Problem (4) is strictly convex in $x_{-1} = [x_i]_{i\ne 1}$, thus showing that $\tilde{x} = \bar{x}$. If both $EC(\tilde{x}, p^{OE}) = R$ and $EC(\bar{x}, p^{OE}) = R$, then $\bar{x}_i = \tilde{x}_i = l_i^{-1}(R - p_i^{OE})$ for all $i \ne 1$, establishing our claim.

For some $x \in W(p^{OE})$, consider the vector $x^{OE} = (d - \sum_{i\ne 1} x_i, x_{-1})$. Since $x_{-1}$ is uniquely defined and $x_1$ is chosen such that the provider that owns link 1 has no incentive to deviate, it follows that $(p^{OE}, x^{OE})$ is an OE.

**Appendix C. Proof of Proposition 4.3.** We will prove Proposition 4.3 using Theorem 5* of Dasgupta and Maskin [13]. We start by stating a slightly simplified version of this theorem. Consider an $S$ player game. Let the strategy space of player $s$, denoted by $P_s$, be a closed interval of $\mathbb{R}^{n_s}$ for some $n_s \in \mathbb{N}$, and its payoff function by $\pi_s(p_s, p_{-s})$. We also denote $p = (p_s, p_{-s})$, $P = \prod_{s=1}^{S} P_s$, and $P_{-s} = \prod_{k=1,\, k\ne s,}^{S} P_s$. To state Theorem 5* in Dasgupta and Maskin [13], we need the following three definitions.

Definition C1. Let $\pi(p) = \sum_{s\in\mathcal{S}} \pi_s(p_s, p_{-s})$. $\pi(p)$ is upper semicontinuous in $p$ if for all $\bar{p}$,

$$\limsup_{p\to\bar{p}} \pi(p) \le \pi(\bar{p}).$$

Definition C2. The profit function $\pi_s(p_s, p_{-s})$ is weakly lower semicontinuous in $p_s$ if for all $\bar{p}_s \in P_s$, there exists $\lambda \in [0, 1]$ such that for all $p_{-s} \in P_{-s}$,

$$\lambda \liminf_{p_s \downarrow \bar{p}_s} \pi_s(p_s, p_{-s}) + (1-\lambda) \liminf_{p_s \uparrow \bar{p}_s} \pi_s(p_s, p_{-s}) \ge \pi_s(\bar{p}_s, p_{-s}).$$

Definition C3. For each player $s$, let $D_s \in \mathbb{N}$. For each $D$ with $0 \le D \le D_s$ and each $k \ne s$ with $1 \le k \le S$, let $f_{sk}^D$ be a one-to-one continuous function. Let $\bar{P}(s)$ be a subset of $P$ such that

$$\bar{P}(s) = \{(p_1, \ldots, p_S) \in P \mid \exists\, k \ne s,\ \exists\, D,\ 0 \le D \le D_s \text{ s.t. } p_k = f_{sk}^D(p_s)\}.$$

In other words, $\bar{P}(s)$ is a lower dimensional subset of $P$ (which is also of Lebesgue measure zero). Theorem 5* in Dasgupta and Maskin [13] states:

Theorem C1 (Dasgupta-Maskin). *Assume that $\pi_s(p_s, p_{-s})$ is continuous in $p$ except on a subset $P^{**}$ of $\bar{P}(s)$, weakly lower semicontinuous in $p_s$ for all $s$ and bounded, and that $\pi(p)$ is upper semicontinuous in $p$. Then the game $[(P_s, \pi_s);\ s = 1, 2, \ldots, S]$ has a mixed-strategy equilibrium.*

We show that our game satisfies the hypotheses of Theorem C1. We will select a function $x^*(p_s, p_{-s})$ from the set of WE, $W(p_s, p_{-s})$, such that

$$\pi_s(p_s, p_{-s}) = \Pi_s(p_s, p_{-s}, x^*(p_s, p_{-s})), \quad \forall s \in \mathcal{S},$$

that will satisfy these hypotheses. First, since $P_s = [0, R]^{I_s}$ and $\sum_i x_i(p) \le d$ for all $p$, and all $x \in W(p)$, $\pi_s(p_s, p_{-s})$ is clearly bounded.

Since $W(p_s, p_{-s})$ is an upper semicontinuous correspondence, we select $x^*(\cdot)$ such that

$$\liminf_{p_s \uparrow \bar{p}_s} \sum_{j \in \mathcal{I}_s} x_j^*(p_s, p_{-s}) = \sum_{j \in \mathcal{I}_s} x_j^*(\bar{p}_s, p_{-s}), \quad \forall \bar{p}_s \ge 0, \ \forall p_{-s} \ge 0. \tag{A11}$$

Given $p \ge 0$, since $p_j = p_k$ for all $j, k \in \bar{\mathcal{I}}$, where $\bar{\mathcal{I}}$ is defined in Equation (5) in Lemma 2.2, it follows that

$$\liminf_{p_s \uparrow \bar{p}_s} \pi_s(p_s, p_{-s}) = \pi_s(\bar{p}_s, p_{-s}), \quad \forall \bar{p}_s \ge 0, \ \forall p_{-s} \ge 0,$$

hence ensuring that $\pi_s(p_s, p_{-s}) = \Pi_s(p_s, p_{-s}, x^*(p_s, p_{-s}))$ is weakly lower semicontinuous. We claim that we have

$$\sum_j x_j^*(p) \ge \sum_j x_j(p), \quad \forall p \ge 0, \ \forall x \in W(p). \tag{A12}$$

Assume the contrary. This implies that there exist some $\bar{p} \ge 0$, $s \in \mathcal{S}$, and $x \in W(\bar{p})$ such that

$$\sum_{j \in \mathcal{I}_s} x_j(\bar{p}) > \sum_{j \in \mathcal{I}_s} x_j^*(\bar{p}). \tag{A13}$$

By Equation (A11), we have that $\sum_{j \in \mathcal{I}_s} x_j^*(p_s^n, \bar{p}_{-s}) \to \sum_{j \in \mathcal{I}_s} x_j^*(\bar{p}_s, \bar{p}_{-s})$ for some sequence $\{p_s^n\} \uparrow \bar{p}_s$. Combined with Equation (A13), this implies that

$$\sum_{j \in \mathcal{I}_s} x_j(\bar{p}) > \sum_{j \in \mathcal{I}_s} x_j^*(\hat{p}_s, \bar{p}_{-s})$$

for some $\hat{p}_s < p_s$, contradicting the monotonicity of WE by Proposition 2.2.

Next, we show that $\pi_s(p_s, p_{-s})$ is continuous in $p$ except on a set $P^{**}$. We define the set

$$P^{**} = \{p \mid W(p) \text{ is not a singleton}\}.$$

By the upper semicontinuity of $W(p)$, we see that $\pi_s(p_s, p_{-s})$ is continuous at all $p \notin P^{**}$. Moreover, by Lemma 2.2, it follows that $P^{**} \subset \bar{P}$, where

$$\bar{P} = \{p \mid p_j = p_k, \text{ for some } j \ne k\} \cup \{p \mid p_j = R, \text{ for some } j\},$$

which is a lower dimensional set. This establishes the desired condition for Theorem C1.

Finally, we show that

$$\pi(p) = \sum_{s \in \mathcal{S}} \pi_s(p_s, p_{-s}) = \sum_{i \in \mathcal{I}} p_i x_i^*(p)$$

is continuous at all $p$. Given some $p \ge 0$, define $\bar{\mathcal{I}}$ as in Equation (5) of Lemma 2.2. If $\bar{\mathcal{I}} = \varnothing$, then we automatically have that $\pi$ is continuous at $p$. Assume that $\bar{\mathcal{I}} \ne \varnothing$. Since $x_i^{OE}(\cdot)$ is continuous at $p$ for all $i \notin \bar{\mathcal{I}}$ and $p_j = p_k$ for all $j, k \in \bar{\mathcal{I}}$, it is sufficient to show that $\sum_{i \in \bar{\mathcal{I}}} x_i^*(p)$ is continuous at $p$, i.e., for a sequence $\{p^n\}$ with $p^n \in [0, \mathbb{R}]^I$ and $p^n \to p$, we show that

$$\lim_{n \to \infty} \sum_{i \in \bar{\mathcal{I}}} x_i^*(p^n) = \sum_{i \in \bar{\mathcal{I}}} x_i^*(p).$$

Define

$$\tilde{d}(p^n) = \sum_{i \notin \bar{\mathcal{I}}} x_i^*(p^n).$$

Since $x_i(\cdot)$ is continuous at $p$ for all $i \notin \bar{\mathcal{I}}$, we have $\tilde{d}(p^n) \to d(p) = \sum_{i \notin \bar{\mathcal{I}}} x_i(p)$. Consider two cases:

- $\sum_{i \in \bar{\mathcal{I}}} b_{C_i} > d - \tilde{d}(p)$. Since $x^*(p)$ is the maximum $l_1$-norm element of $W(p)$ (cf. Equation (A12)) and $l_i(x_i^*) = 0$ for all $i \in \bar{\mathcal{I}}$, this implies that $\sum_{i \in \mathcal{I}} x_i^*(p) = d$, and for all $n$ sufficiently large $\sum_{i \in \mathcal{I}} x_i^*(p^n) = d$, establishing the claim.

- $\sum_{i\in\bar{\mathcal{I}}} b_{C_i} \le d - \tilde{d}(p)$. By the same reasoning as in the previous part, this implies that $\sum_{i\in\mathcal{I}} x_i^*(p) = \sum_{i\in\bar{\mathcal{I}}} b_{C_i}$. Moreover, for all $\epsilon > 0$, there exists some $n$ sufficiently large such that

$$\left|\sum_{i\in\mathcal{I}} x_i^*(p^n) - \sum_{i\in\bar{\mathcal{I}}} b_{C_i}\right| \le \epsilon,$$

establishing the claim.
The preceding enable us to apply the theorem, completing the proof.

**Appendix D. Proof of Lemma 4.3.** We first prove the following lemma.

Lemma D.1. *Let $(p^{OE}, x^{OE})$ be an OE such that $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} < R$. Let Assumptions 2.1 and 4.1 hold. If $p_j^{OE} x_j^{OE} > 0$ for some $j \in \mathcal{I}$, then $W(p^{OE})$ is a singleton.*

Proof. Since $p_j^{OE} x_j^{OE} > 0$ for some $j \in \mathcal{I}$, it follows by Lemma 4.2 that $p_i^{OE} x_i^{OE} > 0$ for all $i \in \mathcal{I}$. We first show that for all $x \in W(p^{OE})$, we have $x_i \le x_i^{OE}$ for all $i$. If $l_i(x_i^{OE}) > 0$, then by Lemma 2.2, $x_i = x_i^{OE}$ for all $x \in W(p^{OE})$. If $l_i(x_i^{OE}) = 0$, then $\mathcal{I}_s = \{i\}$ for some $s$ by the fact that $x_i^{OE} > 0$ and Assumption 4.1, which implies that $x_i \le x_i^{OE}$ by the definition of an OE (cf. Definition 4.1).
Since $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} < R$, we have $\sum_{i\in\mathcal{I}} x_i^{OE} = d$. Moreover, the fact that $x_i \le x_i^{OE}$ for all $x \in W(p^{OE})$ implies that $\min_j\{p_j^{OE} + l_j(x_j)\} < R$ as well, and therefore $\sum_{i\in\mathcal{I}} x_i = d$, showing that $x_i = x_i^{OE}$, for all $x \in W(p^{OE})$ for all $i \in \mathcal{I}$. □

Proof of Lemma 4.3. We first prove this result for a network with two links. Assume to arrive at a contradiction that

$$l_2^+(x_2^{OE}) > l_2^-(x_2^{OE}). \tag{A14}$$

Let $\{\epsilon^k\}$ be a scalar sequence with $\epsilon^k \downarrow 0$. Consider the sequence $\{x_1(\epsilon^k)\}$, where $x_1(\epsilon^k)$ is the load of link 1 at a WE given price vector $(p_1^{OE} + \epsilon^k, p_2^{OE})$. By Proposition 2.1 and Lemma D.1, the WE correspondence $W(p)$ is upper semicontinuous and $W(p^{OE})$ is a singleton. Therefore it follows that $x_1(\epsilon^k) \to x_1^{OE}$. Define

$$\frac{\partial^+ x_1(p_1^{OE}, p_2^{OE})}{\partial p_1} = \lim_{k\to\infty} \frac{x_1(\epsilon^k) - x_1^{OE}}{\epsilon^k}. \tag{A15}$$

Similarly, let $x_1(-\epsilon^k)$ be the load of link 1 at a WE given price vector $(p_1^{OE} - \epsilon^k, p_2^{OE})$. Since $W(p^{OE})$ is a singleton, we also have $x_1(-\epsilon^k) \to x_1^{OE}$. Define

$$\frac{\partial^- x_1(p_1^{OE}, p_2^{OE})}{\partial p_1} = \lim_{k\to\infty} \frac{x_1^{OE} - x_1(-\epsilon^k)}{\epsilon^k}. \tag{A16}$$

Since $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} < R$, it can be seen using Lemma 4.1 that

$$\frac{\partial^+ x_1(p_1^{OE}, p_2^{OE})}{\partial p_1} \ge \frac{-1}{l_1^-(x_1^{OE}) + l_2^+(x_2^{OE})}, \quad \text{and} \quad \frac{\partial^- x_1(p_1^{OE}, p_2^{OE})}{\partial p_1} \le \frac{-1}{l_1^+(x_1^{OE}) + l_2^-(x_2^{OE})}.$$

Since $l_1^+(x_1^{OE}) = l_1^-(x_1^{OE})$ by Assumption 4.2, this combined with Equation (A14) yields

$$\frac{\partial^+ x_1(p_1^{OE}, p_2^{OE})}{\partial p_1} > \frac{\partial^- x_1(p_1^{OE}, p_2^{OE})}{\partial p_1}. \tag{A17}$$

Consider the profit of service provider 1, $\Pi_1(p_1^{OE}, p_2^{OE}) = p_1^{OE} x_1^{OE}$. Define

$$\frac{\partial^+ \Pi_1(p_1^{OE}, p_2^{OE})}{\partial p_1} = \lim_{k\to\infty} \frac{\Pi_1(p_1^{OE} + \epsilon^k, p_2^{OE}) - \Pi_1(p_1^{OE}, p_2^{OE})}{\epsilon^k},$$

$$\frac{\partial^- \Pi_1(p_1^{OE}, p_2^{OE})}{\partial p_1} = \lim_{k\to\infty} \frac{\Pi_1(p_1^{OE}, p_2^{OE}) - \Pi_1(p_1^{OE} - \epsilon^k, p_2^{OE})}{\epsilon^k}.$$

Since $p_1^{OE}$ is a maximum of $\Pi_1(\cdot, p_2^{OE})$, we have

$$\frac{\partial^+ \Pi_1(p_1^{OE}, p_2^{OE})}{\partial p_1} = x_1^{OE} + p_1^{OE} \frac{\partial^+ x_1(p_1^{OE}, p_2^{OE})}{\partial p_1} \le 0, \tag{A18}$$

and

$$\frac{\partial^- \Pi_1(p_1^{OE}, p_2^{OE})}{\partial p_1} = x_1^{OE} + p_1^{OE} \frac{\partial^- x_1(p_1^{OE}, p_2^{OE})}{\partial p_1} \ge 0, \tag{A19}$$

which, when combined, yields

$$\frac{\partial^+ x_1(p_1^{OE}, p_2^{OE})}{\partial p_1} \le \frac{\partial^- x_1(p_1^{OE}, p_2^{OE})}{\partial p_1}, \tag{A20}$$

which is a contradiction by Equation (A17), thus showing that we have $l_2^+(x_2^{OE}) = l_2^-(x_2^{OE})$.

We next consider a network with multiple links. As in Equations (A15) and (A16), we define for all $i \in \mathcal{I}$,

$$\frac{\partial^+ x_i(p^{OE})}{\partial p_1} = \lim_{k\to\infty} \frac{x_i(\epsilon^k) - x_i^{OE}}{\epsilon^k}, \qquad \frac{\partial^- x_i(p^{OE})}{\partial p_1} = \lim_{k\to\infty} \frac{x_i^{OE} - x_i(-\epsilon^k)}{\epsilon^k}.$$

Using the same line of argument as above, we obtain

$$\begin{aligned} \frac{\partial^+ x_1(p^{OE})}{\partial p_1} &\ge \frac{-1}{l_1^-(x_1^{OE}) + 1/\sum_{i\neq 1}(1/l_i^+(x_i^{OE}))}, \\ \frac{\partial^- x_1(p^{OE})}{\partial p_1} &\le \frac{-1}{l_1^+(x_1^{OE}) + 1/\sum_{i\neq 1}(1/l_i^-(x_i^{OE}))}. \end{aligned} \tag{A21}$$

Let $1 \in \mathcal{I}_s$, and without loss of any generality, assume that all $l_i$'s for $i \in \mathcal{I}_s$ are smooth (recall Assumption 4.2). For all $i \in \mathcal{I}_s$, $i \neq 1$, we obtain

$$\begin{aligned} \frac{\partial^+ x_i(p^{OE})}{\partial p_1} &\ge \frac{1}{l_i^+(x_i^{OE})(1 + l_1^-(x_1^{OE})(\sum_{j\neq 1} 1/l_j^+(x_j^{OE})))}, \\ \frac{\partial^- x_i(p^{OE})}{\partial p_1} &\le \frac{1}{l_i^-(x_i^{OE})(1 + l_1^+(x_1^{OE})(\sum_{j\neq 1} 1/l_j^-(x_j^{OE})))}. \end{aligned}$$

To arrive at a contradiction, assume that $l_j^+(x_j^{OE}) > l_j^-(x_j^{OE})$ for some $j \notin \mathcal{I}_s$. Then the preceding two sets of equations imply that

$$\frac{\partial^+ x_i(p^{OE})}{\partial p_1} > \frac{\partial^- x_i(p^{OE})}{\partial p_1} \tag{A22}$$

for all $i \in \mathcal{I}_s$.

Next, Equations (A18) and (A19) for multiple link case are given by

$$\begin{aligned} \frac{\partial^+ \Pi_1(p^{OE})}{\partial p_1} &= x_1^{OE} + p_1^{OE} \frac{\partial^+ x_1(p^{OE})}{\partial p_1} + \sum_{i\in\mathcal{I}_s,\, i\neq 1} p_i^{OE} \frac{\partial^+ x_i(p^{OE})}{\partial p_1} \le 0, \\ \frac{\partial^- \Pi_1(p^{OE})}{\partial p_1} &= x_1^{OE} + p_1^{OE} \frac{\partial^- x_1(p^{OE})}{\partial p_1} + \sum_{i\in\mathcal{I}_s,\, i\neq 1} p_i^{OE} \frac{\partial^- x_i(p^{OE})}{\partial p_1} \ge 0, \end{aligned} \tag{A23}$$

which are inconsistent with Equation (A22), leading to a contradiction. This proves the claim for the multiple link case.

**Appendix E. Proof of Proposition 4.4.** We first assume that $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} < R$. Consider service provider $s$ and assume without loss of generality that $1 \in \mathcal{I}_s$. Since $p_j^{OE} x_j^{OE} > 0$ for some $j \in \mathcal{I}_{s'}$ and $s' \in \mathcal{S}$, it follows by Lemma 4.2 that $p_i^{OE} x_i^{OE} > 0$ for all $i \in \mathcal{I}$. Together with Lemma 4.1, this implies that $((p_i^{OE})_{i\in\mathcal{I}_s}, x^{OE})$ is an optimal solution of the problem

$$\underset{((p_i)_{i\in\mathcal{I}_s},\, x)\ge 0}{\text{maximize}} \quad \sum_{i\in\mathcal{I}_s} p_i x_i \tag{A24}$$

$$\begin{aligned} \text{subject to} \quad & l_1(x_1) + p_1 = l_i(x_i) + p_i, \quad i \in \mathcal{I}_s - \{1\}, \\ & l_1(x_1) + p_1 = l_i(x_i) + p_i^{OE}, \quad i \notin \mathcal{I}_s, \\ & l_1(x_1) + p_1 \le R, \\ & \sum_{i\in\mathcal{I}} x_i \le d. \end{aligned} \tag{A25}$$

By Lemma 4.3, we have that $l_i$ is continuously differentiable in a neighborhood of $x_i^{OE}$ for all $i$ (since the gradient mapping of a convex function is continuous over the set the function is differentiable, see Rockafellar [42]). Therefore, by examining the KKT conditions of this problem, we obtain

$$p_i^{OE} = x_i^{OE} l_i'(x_i^{OE}) - \theta, \quad \forall\, i \in \mathcal{I}_s, \tag{A26}$$

where

$$\theta = \begin{cases} 0, & \text{if } l'_j(x_j^{OE}) = 0 \text{ for some } j \notin \mathcal{I}_s, \\ -\dfrac{\sum_{j\in\mathcal{I}_s} x_j^{OE}}{\sum_{j\notin\mathcal{I}_s} 1/l'_j(x_j^{OE})}, & \text{otherwise,} \end{cases} \tag{A27}$$

showing the result in Equation (16).

We next assume that $\min_j\{p_j^{OE} + l_j(x_j^{OE})\} = R$. Using the assumption that $p_j^{OE} x_j^{OE} > 0$ for some $j \in \mathcal{I}$ and Lemma 4.1, this implies that

$$p_i^{OE} = R - l_i(x_i^{OE}), \quad \forall\, i,$$

and thus for all $s \in \mathcal{S}$, $x^{OE}$ is an optimal solution of

$$\begin{aligned} \text{maximize} \quad & \sum_{i\in\mathcal{I}_s} (R - l_i(x_i))x_i \\ \text{subject to} \quad & x_i \in T_i, \quad \forall\, i \notin \mathcal{I}_s \\ & \sum_{i\in\mathcal{I}} x_i \le d, \end{aligned} \tag{A28}$$

where $T_i = \{x_i \mid p_i^{OE} + l_i(x_i) = R\}$ is either a singleton or a closed interval. Since this is a convex problem, using the optimality conditions, we obtain

$$R - l_i(x_i^{OE}) - x_i^{OE} g_{l_i} = \theta_s, \quad \forall\, i \in \mathcal{I}_s,$$

where $\theta_s \ge 0$ is the Lagrange multiplier associated with Constraint (A28) and $g_{l_i} \in \partial l_i(x_i^{OE})$. Since $l_i^-(x_i^{OE}) \le g_{l_i}$, the preceding implies

$$p_i^{OE} = R - l_i(x_i^{OE}) \ge x_i^{OE} l_i^-(x_i^{OE}), \quad \forall\, i \in \mathcal{I},$$

proving (17).

To prove (18), consider some $i \in \mathcal{I}$ with $\mathcal{I}_s = \{i\}$ for some $s$ and the sequence of price vectors $\{p^k\}$ with $p^k = (p_i^{OE} - \epsilon^k, p_{-i}^{OE})$. Let $\{x^k\}$ be a sequence such that $x^k \in W(p^k)$ for all $k$. By the upper semicontinuity of $W(p)$, it follows that $x^k \to \bar{x}$ with $\bar{x} \in W(p^{OE})$ and $\bar{x} \le x^{OE}$ (see the proof of Lemma D.1). Moreover, by Lemma 2.2, we have $x_i^k \ge x_i^{OE}$ for all $k$, which implies that $\bar{x}_i \ge x_i^{OE}$, showing that $x_i^k \to x_i^{OE}$. We can now use Equations (A19) and (A23) (by substituting $i$ instead of 1 and using $\mathcal{I}_s = \{i\}$) to conclude that

$$p_i^{OE} \le x_i^{OE} l_i^+(x_i^{OE}) + \frac{x_i^{OE}}{\sum_{j\neq 1} 1/l_j^-(x_j^{OE})}.$$

**Acknowledgments.** The authors thank Xin Huang, Ramesh Johari, Eric Maskin, Eilon Solan, Nicolas Stier Moses, Jean Tirole, John Tsitsiklis, Ivan Werning, Muhamet Yildiz, two anonymous referees, and participants at various seminars and conferences for useful comments.

## References

[1] Acemoglu, D., A. Ozdaglar. 2004. Flow control, routing, and performance from service provider viewpoint. Technical Report WP-1696, LIDS, Massachusetts Institute of Technology, Cambridge, MA.
[2] Acemoglu, D., A. Ozdaglar. 2006. Price competition in communication networks. *Proc. IEEE INFOCOM*, Barcelona, Catalunya, Spain.
[3] Anshelevich, E., A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, T. Roughgarden. 2004. The price of stability for network design with selfish agents. *Proc. IEEE Sympos. Foundations Comput. Sci.,* 295–304.
[4] Basar, T., R. Srikant. 2002. Revenue-maximizing pricing and capacity expansion in a many-users regime. *Proc. IEEE INFOCOM*, New York.
[5] Beckmann, M., C. B. Mcguire, C. B. Winsten. 1956. *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.
[6] Benassy, J. P. 1989. Market size and substitutability in imperfect competition: A Bertrand-Edgeworth-Chamberlin model. *Rev. Econom. Stud.* **56** 217–234.
[7] Berge, C. 1963. *Topological Spaces*. Reprint, Dover Publications, New York.
[8] Bergendorff, P., D. W. Hearn, M. V. Ramana. 1997. Congestion toll pricing of traffic networks. P. M. Pardalos, D. W. Hearn, W. W. Hager, eds. *Network Optimization*. Springer-Verlag, Berlin, Germany, 51–71.
[9] Bertsekas, D. P., A. Nedic, A. E. Ozdaglar. 2003. *Convex Analysis and Optimization*. Athena Scientific, Cambridge, MA.
[10] Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2002. Selfish routing in capacitated networks. *Math. Oper. Res.* **29**(4) 961–976.
[11] Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2005. On the inefficiency of equilibria in congestion games. *Proc. 11th Conf. Integer Programming and Combin. Optim.* **3509** 167–181.

[12] Dafermos, S., F. T. Sparrow. 1969. The traffic assignment problem for a general network. *J. Res. National Bureau Standards-B. Math. Sci.* **73**(2) 91–118.
[13] Dasgupta, P., E. Maskin. 1986. The existence of equilibrium in discontinuous economic games: Theory. *Rev. Econom. Stud.* **53** 1–26.
[14] Edgeworth, F. 1925. The pure theory of monopoly. *Papers Relating to Political Economy*, Vol. 1. Macmillan (for the Royal Economic Society, London, UK), 111–143.
[15] Friedman, E. 2004. A generic analysis of selfish routing. *Proc. IEEE CDC*, Bahamas.
[16] Fudenberg, D., J. Tirole. 1991. *Game Theory*. MIT Press, Cambridge, MA.
[17] Hart, O. D. 1979. Monopolistic competition in a large economy with differentiated commodities. *Rev. Econom. Stud.* **46** 1–30.
[18] Hayrapetyan, A., E. Tardos, T. Wexler. 2005. A network pricing game for selfish traffic. *Proc. Annual ACM SIGACT-SIGOPS Sympos. Principles of Distributed Comput.*, Las Vegas, NV.
[19] He, L., J. Walrand. 2003. Pricing Internet services with multiple providers. *Proc. Allerton Conf.*, Allerton-Park, Urbana-Champaign, IL.
[20] Hearn, D. W., M. V. Ramana. 1998. Solving congestion toll pricing models. P. Marcotte, S. Nguyen, eds. *Proc. Equilibrium and Advanced Transportation Modelling Colloquium*. Kluwer Academic Publishers, 109–124.
[21] Hearn, D. W., M. Yildirim. 2002. A toll pricing framework for traffic assignment problems with elastic demand. *Current Trends in Transportation and Network Analysis*.
[22] Johari, R., J. Tsitsiklis. 2004. Network resource allocation and a congestion game. *Math. Oper. Res.* **29**(3) 407–435.
[23] Kelly, F. P., A. K. Maulloo, D. K. Tan. 1998. Rate control for communication networks: Shadow prices, proportional fairness, and stability. *J. Oper. Res. Soc.* **49** 237–252.
[24] Korilis, Y. A., A. A. Lazar, A. Orda. 1997. Achieving network optima using Stackelberg routing strategies. *IEEE/ACM Trans. Networking* **5** 161–173.
[25] Koutsoupias, E., C. Papadimitriou. 1999. Worst-case equilibria. *Proc. 16th Annual Sympos. Theoret. Aspects of Comput. Sci.*, 404–413.
[26] Larsson, T., M. Patriksson. 1994. Equilibrium characterizations of solutions to side constrained asymmetric traffic assignment models. *Le Matematiche* **49** 249–280.
[27] Larsson, T., M. Patriksson. 1998a. Price-directive traffic management: Applications of side constrained traffic equilibrium models. M. G. H. Bell, ed. *Transportation Networks: Recent Technological Advances*, 83–98.
[28] Larsson, T., M. Patriksson. 1998b. Side constrained traffic equilibrium models: Traffic management through link tolls. P. Marcotte, S. Nguyen, eds. *Proc. Equilibrium and Advanced Transportation Model. Colloquium*, 125–151.
[29] Larsson, T., M. Patriksson. 1999. Side constrained traffic equilibrium models-analysis, computation and applications. *Transportation Res.* **33** 233–264.
[30] Low, S., D. E. Lapsley. 1999. Optimization flow control, *i*: Basic algorithm and convergence. *IEEE/ACM Trans. Networking* **7**(6) 861–874.
[31] Maheswaran, R., T. Basar. 2003. Nash equilibirum and decentralized negotiation in auctioning divisible resources. *Group Decision Negotiation* **12** 361–395.
[32] Mascolell, A., M. D. Whinston, J. R. Green. 1995. *Microeconomic Theory*. Oxford University Press, Oxford, UK.
[33] Milchtaich, I. 1996a. Congestion games with player-specific payoff functions. *Games Econom. Behav.* **13** 111–124.
[34] Milchtaich, I. 1996b. Congestion models of competition. *Amer. Naturalist* **147**(5) 760–783.
[35] Novshek, W. 1985. Perfectly competitive markets as the limits of Cournot markets. *J. Econom. Theory* **19** 223–266.
[36] Orda, A., R. Rom, N. Shimkin. 1993. Competitive routing in multi-user communication networks. *IEEE/ACM Trans. Networking* **1** 510–521.
[37] Osborne, M. J., A. Rubinstein. 1994. *Game Theory*. MIT Press, Cambridge, MA.
[38] Patriksson, M. 1994. *The Traffic Assignment Problem: Models and Methods*. VSP, Utrecht, The Netherlands.
[39] Perakis, G. 2004. The price of anarchy when costs are non-separable and asymmetric. *Proc. 10th Conf. Integer Programming and Combin. Optim.*, Vol. 3064. MIT, Cambridge, MA, 46–58.
[40] Pigou, A. C. 1920. *The Economics of Welfare*. Macmillan, London, UK.
[41] Roberts, D., A. Postlewaite. 1976. The incentives for price taking in large exchange economies. *Econometrica* **44** 115–127.
[42] Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.
[43] Rosenthal, R. W. 1973. A class of games possessing pure-strategy Nash equilibria. *Internat. J. Game Theory* **2** 65–67.
[44] Roughgarden, T., É. Tardos. 2002. How bad is selfish routing? *J. ACM* **49**(2) 236–259.
[45] Samuelson, P. A. 1952. Spatial price equilibrium and linear programming. *Amer. Econom. Rev.* **42** 283–303.
[46] Sanghavi, S., B. Hajek. 2004. Optimal allocation of a divisible good to strategic buyers. *Proc. IEEE CDC*, Paradise Island, Bahamas.
[47] Shenker, S. 1995. Fundamental design issues for the future Internet. *IEEE J. Selected Areas in Comm.* **13** 1176–1188.
[48] Shubik, M. 1959. *Strategy and Market Structure*. Wiley, New York.
[49] Smith, M. J. 1979. The existence, uniqueness and stability of traffic equilibria. *Transportation Res.* **13B** 295–304.
[50] Tirole, J. 1990. *The Theory of Industrial Organization*. MIT Press, Cambridge, MA.
[51] Topkis, D. 1979. Equilibrium points in nonzero-sum *n*-person submodular games. *SIAM J. Control Optim.* **17** 773–787.
[52] Veblen, T. 1959. *The Theory of the Leisure Class: An Economic Study of Institutions*. Vanguard Press, New York.
[53] Vetta, A. 2002. Nash equilibria in competitive societies with applications to facility location, traffic routing, and auctions. *Proc. IEEE Sympos. Foundations of Comput. Sci.*, McGill University, Montreal, Canada, 416–428.
[54] Vives, X. 1985. On the efficiency of Bertrand and Cournot equilibria with product differentiation. *J. Econom. Theory* **36** 166–175.
[55] Vives, X. 2001. *Oligopoly Pricing*. MIT Press, Cambridge, MA.
[56] Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proc. Inst. Civil Engineers*, London, UK, 325–378.
[57] Wilson, R. B. 1993. *Nonlinear Pricing*. Oxford University Press, New York.
[58] Yang, S., B. Hajek. 2005. Revenue and stability of a mechanism for efficient allocation of a divisible good. Preprint,