Partially Optimal Routing

Daron Acemoglu, Ramesh Johari, *Member, IEEE,* Asuman Ozdaglar, *Member, IEEE*

***Abstract*— Most large-scale communication networks, such as the Internet, consist of interconnected administrative domains. While source (or selfish) routing, where transmission follows the least cost path for each source, is reasonable across domains, service providers typically engage in traffic engineering to improve operating performance within their own network. Motivated by this observation, we develop and analyze a model of *partially optimal routing*, where optimal routing within subnetworks is overlaid with selfish routing across domains. We demonstrate that optimal routing within a subnetwork does not necessarily improve the performance of the overall network. In particular, when Braess' paradox occurs in the network, partially optimal routing may lead to worse overall network performance. We provide bounds on the worst-case loss of efficiency that can occur due to partially optimal routing. For example, when all congestion costs can be represented by affine latency functions and all administrative domains have a single entry and exit point, the worst-case loss of efficiency is no worse than 25% relative to the optimal solution. In the presence of administrative domains incorporating multiple entry and/or exit points, however, the performance of partially optimal routing can be arbitrarily inefficient even with linear latencies. We also provide conditions for traffic engineering to be individually optimal for service providers.**

***Index Terms*— Traffic engineering, selfish routing, Wardrop equilibrium, Braess' paradox.**

I. Introduction

SINCE the passage of the Telecommunications Act in 1996, the Internet has undergone a dramatic transformation and experienced increasing decentralization. Today, thousands of network providers cooperate and compete to provide end-to-end network service to billions of users worldwide. While end-users care about the performance across the entire network, individual network providers optimize their own objectives. The Internet's architecture provides no guarantees that provider incentives will be aligned with end-user objectives.

The emergence of *overlay routing* over the past five years has further highlighted the potentially conflicting objectives of the service provider and the end-users. In overlay routing, end-user software (such as peer-to-peer file-sharing software) makes route selection decisions on the basis of the best end-to-end performance available at any given time, while administrative domains control the routing of traffic within their own (sub)networks. Network operators use traffic engineering to optimize performance, and also to react to the global routing decisions of overlay networks (e.g., [2]).

These considerations make it clear that the study of routing patterns and performance in large-scale communication networks requires an analysis of *partially optimal routing*, where end-to-end route selection is selfish and responds to aggregate route latency, but network providers redirect traffic within their own networks to achieve minimum intradomain total latency.

We develop and analyze a model of partially optimal routing, combining selfish across-domain routing and *traffic engineering* by service providers within their administrative domains. While recent research (e.g., [3]–[7]) has studied the interactions of overlay routing and traffic engineering, it has neither provided a formal model of partially optimal routing nor theoretically addressed the central question of whether partially optimal routing improves overall network performance.

We consider routing flows between multiple source-destination pairs through a network. Each link is endowed with a *latency function* describing the congestion level (e.g., delay or probability of packet loss) as a function of the total flow passing through the link (e.g., [8]). Each source-destination pair in the network has a fixed amount of flow, and flows follow the minimum delay route among the available paths as captured by the familiar notion of *Wardrop equilibrium* (e.g., [8]). Our innovation is to allow subsets of the links in the network ("subnetworks") to be independently owned and operated by different providers, and consider the possibility that these providers engage in traffic engineering and route traffic to minimize the total (or average) latency within their subnetworks. Source-destination pairs sending traffic across subnetworks perceive only the *effective latency* resulting from the traffic engineering of the service providers. The resulting equilibrium, which we call a *partially optimal routing* (POR) equilibrium, is a Wardrop equilibrium according to the effective latencies seen by the source-destination pairs. This model provides a stylized description of the practice of traffic engineering in the Internet.

Because of the congestion externalities created by selfish routing, the Wardrop equilibrium without traffic engineering within subnetworks is typically inefficient and leads to a level of total delay in excess of the system optimum (see, for example, [9]–[11]). It may therefore be conjectured that the addition of traffic engineering within parts of the network will improve performance. Our first set of results show that this is not necessarily the case. In particular, when the Braess'

Manuscript received May 27, 2006; revised January 15, 2007. This work was partially supported by an Okawa Foundation research grant, the National Science Foundation grant 0428868, and the National Science Foundation Career award DMI-0545910. A preliminary version of a subset of this work was presented at the Conference on Information Sciences and Systems, Princeton, N.J., March 22-24, 2006 [1].

D. Acemoglu is with the Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, 02139 (e-mail: daron@mit.edu).

R. Johari, is with the Department of Management Science and Engineering, Stanford University, Stanford, CA, 94305 (e-mail: ramesh.johari@stanford.edu).

A. Ozdaglar is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 (e-mail: asuman@mit.edu).

Digital Object Identifier 10.1109/JSAC.2007.0708xx.

paradox occurs within the global network, partially optimal routing may be less efficient than the Wardrop equilibrium.[1]

Motivated by this finding, we study the extent of inefficiency of partially optimal routing relative to the system optimum. For the case in which all independently-operated subnetworks have unique entry and exit points, and latency functions belong to certain subclasses, we provide tight bounds for the inefficiency of partially optimal routing that exactly match the corresponding bounds for the performance of Wardrop equilibria. For example, with affine latency functions, the worst-case performance of partially optimal routing is no worse than 25% relative to the system optimum, matching the same bound provided for Wardrop equilibrium by Roughgarden and Tardos [10]. Similarly, we provide bounds for the cases in which latency functions are nonnegative polynomials of bounded degree.

Interestingly, however, in the case of subnetworks with multiple entry and exit points, the performance of partially optimal routing can be arbitrarily inefficient even with *linear* latency functions. This contrasts with the tight bound of 25% efficiency loss for Wardrop equilibria with linear or affine latency functions (see [10], [11]). In this general case, we can only provide bounds for some special classes of subnetworks with multiple entry and exit points.

We conclude by investigating subnetwork performance measured in terms of total delay (latency) under partially optimal routing. We show that, in the absence of prices per unit of transmission, a service provider may prefer not to engage in traffic engineering in order to reduce total flow and delay in their subnetwork. In addition, we provide conditions for service providers to prefer traffic engineering to selfish routing within their domain.

The remainder of the paper is organized as follows. Section II introduces the three basic routing paradigms: socially optimal routing, where total (or average) latency is minimized across the entire network; selfish routing, where end-to-end route selection is made based on minimum route latency; and partially optimal routing, where end-to-end route selection is still dependent on aggregate route latency, but providers engage in traffic engineering within their subnetworks to achieve minimum intradomain total latency.

Section III analyzes the performance of partially optimal routing. We show that there may exist situations where optimization within a subnetwork leads to lower global network performance. We prove that this can only be the case when the Braess' paradox occurs within the global network. Section IV then analyzes the worst-case efficiency loss that can occur at the partially optimal routing solution and establishes bounds on efficiency loss when all latency functions are affine, and when all latency functions are nonnegative polynomials of bounded degree. In the special case where all latency functions are affine, we find that the ratio of partially optimal routing cost to the social optimum is no worse than $4/3$.

In Section V, we consider the case where subnetworks may have multiple entry and exit points, and show how partially optimal routing leads to further inefficiencies in this case. Section VI considers the choice of routing policy by a single service provider and provides conditions under which traffic engineering is (individually) optimal for a provider in parallel link topologies. We conclude in Section VII.

[1]Throughout, by *Wardrop equilibrium* we refer to the equilibrium of the same network structure without any traffic engineering—without any optimal routing within subnetworks.

II. Preliminaries: Different Routing Paradigms

We consider a directed network $G = (V, A)$, with node set V, link (or edge) set A, and w source-destination node pairs $\{s_1, t_1\}, \ldots, \{s_w, t_w\}$. Let $W = \{1, \ldots, w\}$. Let P_i denote the set of paths available from s_i to t_i using the edges in A; we view each path $p \in P_i$ as a subset of A, $p \subset A$. Define $P = \cup_{i \in W} P_i$. Each link $j \in A$ has a strictly increasing, nonnegative *latency function* $l_j(x_j)$ as a function of the flow on link j.[2] We assume that X_i units of flow are to be routed from s_i to t_i, for all $i \in W$, and we define $\mathbf{X} = [X_1, \ldots, X_w]$. We call the tuple $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$ a *routing instance*. We denote the set of routing instances $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$ by the set $\mathcal{R}$. In the following, we will also be interested in routing instances in which the latency functions of all links are restricted to belong to a certain class of functions. We denote the set of routing instances R in which all latency functions are convex (affine and concave, respectively) by $\mathcal{R}^{conv}$ ($\mathcal{R}^{aff}$ and $\mathcal{R}^{conc}$, respectively).

A. Socially Optimal Routing

Given a routing instance $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$, we define a *social optimum*, denoted by $\mathbf{x}^{SO}(R)$, as an optimal solution of the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{j \in A} x_j l_j(x_j) \\ \text{subject to} \quad & \sum_{p \in P : j \in p} y_p = x_j, \quad j \in A; \\ & \sum_{p \in P_i} y_p = X_i, \quad i \in W; \\ & y_p \geq 0, \quad p \in P. \end{aligned} \tag{1}$$

This optimization problem minimizes the total (or equivalently the average) delay experienced over all paths. Under our assumption that each latency function is continuous, it follows that at least one social optimum always exists. The total latency cost at a social optimum is given by:

$$C(\mathbf{x}^{SO}(R)) = \sum_{j \in A} x_j^{SO}(R) l_j(x_j^{SO}(R)).$$

B. Selfish Routing

When traffic routes "selfishly"—that is, when sources choose minimum delay end-to-end paths—all paths with nonzero flow must have the same total delay. A flow configuration with this property is called a *Wardrop equilibrium*. Under the assumptions on the latency functions (i.e., each l_j is continuous and strictly increasing), it is well-known that the Wardrop equilibrium flow vector for a given routing instance

[2]Throughout the paper, we will refer to l_j as the latency function, though l_j can be used to model congestion metrics other than latency (e.g., loss).

R, denoted $\mathbf{x}^{WE}(R)$, is the unique optimal solution to the following optimization problem (see e.g., [10], [12]):

$$\begin{aligned} \text{minimize} \quad & \sum_{j\in A}\int_0^{x_j} l_j(z)\,dz \qquad (2)\\ \text{subject to} \quad & \sum_{p\in P:j\in p} y_p = x_j, \quad j\in A;\\ & \sum_{p\in P_i} y_p = X_i, \quad i\in W;\\ & y_p \geq 0, \quad p\in P. \end{aligned}$$

The total latency cost at the Wardrop equilibrium is given by

$$C(\mathbf{x}^{WE}(R)) = \sum_{j\in A} x_j^{WE}(R) l_j(x_j^{WE}(R)). \qquad (3)$$

Equivalently a feasible solution $\mathbf{x}^{WE}$ for a routing instance R is a Wardrop equilibrium if and only if it satisfies

$$\sum_{j\in A} l_j(x_j^{WE})(x_j^{WE} - x_j) \leq 0, \qquad (4)$$

for all feasible solutions $\mathbf{x}$ for the same routing instance; see, for example, [13], [14].

C. Partially Optimal Routing

Let us now assume that a single network provider controls a subnetwork with unique entry and exit points; within this domain, the provider optimizes performance of traffic flow. Formally, we assume there is a collection of directed subgraphs (subnetworks) inside of G. Within a subnetwork $G_0 = (V_0, A_0)$, a service provider *optimally* routes all incoming traffic. Let $s_0 \in V_0$ denote the unique entry point to G_0, and let $t_0 \in V_0$ denote the unique exit point from G_0. Let P_0 denote the set of available paths from s_0 to t_0 using the edges in A_0. We make the assumption that every path in P passing through any node in V_0 must contain a path in P_0 from s_0 to t_0; this is consistent with our assumption that G_0 is an independent autonomous system, with a unique entry and exit point. We call $R_0 = (V_0, A_0, P_0, s_0, t_0)$ a *subnetwork* of G, and with a slight abuse of notation, we say that $R_0 \subset R$.

Given an incoming amount of flow X_0, the network provider chooses a routing of flow to solve the following optimization problem to minimize total (or average) latency:

$$\begin{aligned} \text{minimize} \quad & \sum_{j\in A_0} x_j l_j(x_j) \qquad (5)\\ \text{subject to} \quad & \sum_{p\in P_0:j\in p} y_p = x_j, \quad j\in A_0;\\ & \sum_{p\in P_0} y_p = X_0;\\ & y_p \geq 0, \quad p\in P_0. \end{aligned}$$

In this optimization problem, the subnetwork owner sees an incoming traffic amount X_0, and chooses the optimal routing of this flow through the subnetwork. This is a formal abstraction of the process of *traffic engineering* via link weight optimization, carried out by many network providers to optimize intradomain performance; see, e.g., [15].

Let $L(X_0)$ denote the optimal value of the preceding optimization problem. We define $l_0(X_0) = L(X_0)/X_0$ as the *effective latency* of partially optimal routing in the subnetwork R_0, with flow $X_0 > 0$. If traffic in the entire network G routes selfishly, while traffic is optimally routed within G_0, then replacing G_0 by a single link with latency l_0 will leave the Wardrop equilibrium flow unchanged elsewhere in G.

We have the following simple lemma that provides basic properties of l_0 and L. The proof is straightforward and is omitted; details can be found in [16].

Lemma 1 Assume that every latency function, l_j, is a strictly increasing, nonnegative, and continuous function. Then:

(a) The effective latency $l_0(X_0)$ is a strictly increasing function of $X_0 > 0$.
(b) Assume further that each l_j is a convex function. The total cost $L(X_0)$ is a convex function of X_0.

In light of the preceding lemma, we can extend the definition of l_0 so that $l_0(0) = \lim_{x_0\downarrow 0} l_0(x_0)$; the preceding limit is well defined since l_0 is strictly increasing.

To define the overall network performance under partially optimal routing, first suppose that there is a single independently-operated subnetwork. Given a routing instance $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$, and a subnetwork $R_0 = (V_0, A_0, P_0, s_0, t_0)$ defined as above, we define a new routing instance $R' = (V', A', P', s, t, \mathbf{X}, \mathbf{l}')$ as follows:

$$\begin{aligned} V' &= (V \setminus V_0) \bigcup \{s_0, t_0\};\\ A' &= (A \setminus A_0) \bigcup \{(s_0, t_0)\}; \end{aligned}$$

P' corresponds to all paths in P, where any subpath in P_0 is replaced by the link (s_0, t_0); and $\mathbf{l}'$ consists of latency functions l_j for all edges in $A \setminus A_0$, and latency l_0 for the edge (s_0, t_0). Thus R' is the routing instance R with the subgraph G_0 replaced by a single link with latency l_0; we call R' the *equivalent POR instance* for R with respect to R_0. The overall network flow in R with partially optimal routing in R_0, $\mathbf{x}^{POR}(R, R_0)$, is defined to be the Wardrop equilibrium flow in the routing instance R':

$$\mathbf{x}^{POR}(R, R_0) = \mathbf{x}^{WE}(R').$$

In other words, it is equilibrium with traffic routed selfishly given the effective latency l_0 of the subnetwork R_0. Note also that this formulation leaves undefined the exact flow in the subnetwork R_0; this is to be expected, since problem (5) may not have a unique solution.

The total latency cost of the equivalent POR instance for R with respect to R_0 is given by

$$C(\mathbf{x}^{POR}(R, R_0)) = \sum_{j\in A'} x_j^{POR}(R, R_0) l_j(x_j^{POR}(R, R_0)).$$

The definition immediately generalizes when there are multiple independently-operated subnetworks. Let $R_0^j = (V_0^j, A_0^j, P_0^j, s_0^j, t_0^j)$ for $j = 1, 2, ..., J$ denote the subnetworks, each represented by a directed subgraph G_0^j. Define

$$\begin{aligned} V' &= (V \setminus \bigcup_{j=1}^{J} V_0^j) \bigcup_{j=1}^{J} \{s_0^j, t_0^j\};\\ A' &= (A \setminus \bigcup_{j=1}^{J} A_0^j) \bigcup_{j=1}^{J} \{(s_0^j, t_0^j)\}. \end{aligned}$$

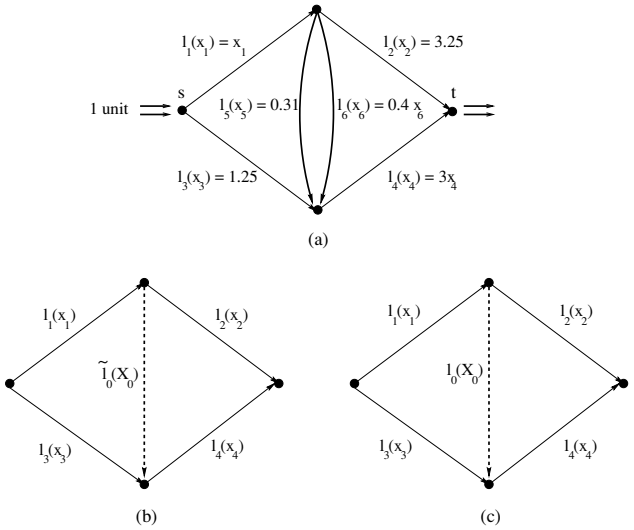

Fig. 1. A network for which POR leads to a worse performance relative to selfish routing. Figures (b) and (c) illustrate representing the subnetwork with a single link with Wardrop effective latency $\tilde{l}_0(X_0)$ and optimal effective latency $l_0(X_0)$, respectively.

Let R' be the routing instance R with each subgraph G_0^j replaced by a single link with effective latency l_0^j. The partially optimal routing flow $\mathbf{x}^{POR}(R, \{R_0^j\}_{j=1}^J)$, is again the Wardrop equilibrium flow in the routing instance R'. In the remainder of the paper, we assume without loss of generality that there is a single subnetwork in the overall network.

III. Partially Optimal Routing and Global Performance

We first consider the effect of optimal routing within subnetworks on the performance of the overall network. One might conjecture that optimally routing traffic within subnetworks should improve the overall performance. The following example shows that this need not be the case.

Example 1 Consider the network $G = (V, A)$ with source and destination nodes $s, t \in V$ illustrated in Figure 1(a). Let $R = (V, A, P, s, t, 1, \mathbf{l})$ be the corresponding routing instance, i.e., one unit of flow is to be routed over this network. The subnetwork G_0 consists of the two parallel links in the middle, links 5 and 6, with latency functions

$$l_5(x_5) = 0.31, \qquad l_6(x_6) = 0.4\ x_6.$$

The latency functions for the remaining links in the network are given by

$$l_1(x_1) = x_1, \qquad l_2(x_2) = 3.25,$$

$$l_3(x_3) = 1.25, \qquad l_4(x_4) = 3x_4.$$

Assume first that the flow through the subnetwork G_0 is routed selfishly, i.e., according to the Wardrop equilibrium. Given a total flow X_0 through the subnetwork G_0, the effective Wardrop latency can be defined as

$$\tilde{l}_0(X_0) = \frac{1}{X_0}\ C(x^{WE}(R_0)), \tag{6}$$

[cf. Eq. (3)], where R_0 is the routing instance corresponding to the subnetwork G_0 with total flow X_0. The effective Wardrop latency for this example is given by

$$\tilde{l}_0(X_0) = \min\{0.31, 0.4X_0\}.$$

Substituting the subnetwork with a single link with latency function $\tilde{l}_0$ yields the network in Figure 1(b). It can be seen that selfish routing over the network of Figure 1(b) leads to the link flows $x_1^{WE} = 0.94$ and $X_0^{WE} = 0.92$, with a total cost of $C(\mathbf{x}^{WE}(R)) = 4.19$. It is clear that this flow configuration arises from a Wardrop equilibrium in the original network.

Assume next that the flow through the subnetwork G_0 is routed optimally, i.e., as the optimal solution of problem (5) for the routing instance corresponding to G_0. Given a total flow X_0 through the subnetwork G_0, the effective latency of optimal routing within the subnetwork G_0 can be defined as

$$l_0(X_0) = \frac{L(X_0)}{X_0},$$

where $L(X_0)$ is the optimal value of problem (5). The effective optimal routing latency for this example is given by

$$l_0(X_0) = \begin{cases} 0.4X_0, & \text{if } 0 \le X_0 \le 0.3875; \\ 0.31 - \frac{0.0961}{1.6X_0}, & \text{if } X_0 \ge 0.3875. \end{cases}$$

Substituting the subnetwork with a single link with latency function l_0 yields the network in Figure 1(c). Note that selfish routing over this network leads to the partially optimal routing (POR) equilibrium. It can be seen that at the POR equilibrium, the link flows are given by $x_1^{POR} = 1$ and $X_0^{POR} = 1$, with a total cost of $C(\mathbf{x}^{POR}(R)) = 4.25$, which is strictly greater than $C(\mathbf{x}^{WE}(R))$.

In the preceding example, when the subnetwork optimizes intradomain performance, we see a degradation in *global* network performance. This is reminiscent of *Braess' paradox*, a classic example of degradation in global network performance despite local improvements (see, for example, [12]). Braess' paradox occurs in a network if reducing the link latency functions increases the total latency in the network. We now investigate the relationship between Braess' paradox and the performance degradation observed in Example 1.

Definition 1 (Braess' paradox) Consider a routing instance $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$ and a subnetwork $R_0 = (V_0, A_0, P_0, s_0, t_0) \subset R$. We say that *Braess' paradox* occurs in R *centered* at R_0 if there exists another routing instance $R_m = (V, A, P, s, t, \mathbf{X}, \mathbf{m})$, with a vector of strictly increasing, nonnegative latency functions, $\mathbf{m} = (m_j, j \in A)$, such that for all $x_j \ge 0$,

$$m_j(x_j) \le l_j(x_j),\ \forall\ j \in A_0, \quad m_j(x_j) = l_j(x_j),\ \forall\ j \notin A_0,$$

and

$$C(\mathbf{x}^{WE}(R_m)) > C(\mathbf{x}^{WE}(R)).$$

In our definition we have explicitly fixed a subnetwork R_0 within which we locally "improve" performance; formally, the routing instance R' differs from R only by a reduction of the latency functions on some (or all) links. Nevertheless, in a network topology where Braess' paradox occurs, this local change can yield a higher total latency.

Similarly, the following definition captures the counterintuitive phenomenon exhibited in Example 1, where traffic engineering within some subnetwork, i.e., partially optimal routing, leads to a degradation in the overall performance compared to pure selfish routing.

Definition 2 (POR paradox) Consider a routing instance $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$, and a subnetwork $R_0 = (V_0, A_0, P_0, s_0, t_0)$. We say that the *POR paradox* (partially optimal routing paradox) occurs in R with respect to R_0 if

$$C(\mathbf{x}^{POR}(R, R_0)) > C(\mathbf{x}^{WE}(R)).$$

Intuitively, the POR paradox appears to be a form of "generalized Braess' paradox", in the following sense. Given a total flow X_0 routed through the subnetwork G_0, we define the effective Wardrop latency $\tilde{l}_0$, as follows:

$$\tilde{l}_0(X_0) = \frac{1}{X_0} \sum_{j \in A_0} x_j^{WE}(R') l_j(x_j^{WE}(R')) = \frac{C(\mathbf{x}^{WE}(R'))}{X_0}, \tag{7}$$

where $R' = (V_0, A_0, P_0, s_0, t_0, X_0, \mathbf{l})$ is a routing instance corresponding to the subnetwork R_0 with total flow X_0 [cf. Eq. (6)]. As in Lemma 1, it is straightforward to show that $\tilde{l}_0$ is strictly increasing. Furthermore, it is clear that $\tilde{l}_0(X_0) \geq l_0(X_0)$ for all $X_0 \geq 0$, since $\mathbf{x}^{WE}(R')$ is a feasible solution to problem (5). Thus when we contrast $\mathbf{x}^{POR}(R)$ and $\mathbf{x}^{WE}(R)$, it is *as if* we are lowering the effective latency of the subnetwork R_0. If this increases the total latency, then we are observing a form of Braess' paradox.

In fact, it is possible to show a stronger result: whenever the POR paradox occurs in R with respect to some $R_0 \subset R$, then Braess' paradox occurs in R centered at R_0. This result is stronger than the "generalized Braess' paradox" discussed in the preceding paragraph, because it shows that Braess' paradox occurs within the original instance R without altering the network topology.

Proposition 1 Consider a routing instance $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$ and a subnetwork $R_0 = (V_0, A_0, P_0, s_0, t_0) \subset R$. Assume that the POR paradox occurs in R with respect to R_0. Then Braess' paradox occurs in R centered at R_0.

Proof: Our approach will be to uniformly lower the latency functions in the subnetwork R_0, such that we exactly ensure at a Wardrop equilibrium the effective latency of R_0 is given by l_0, the effective latency of optimal routing within R_0. This will allow selfish routing to "replicate" the partially optimal routing of flow, and imply Braess' paradox.

Let $\mathbf{x}^{WE}(R)$ be the Wardrop equilibrium flow for the routing instance R, with corresponding path flows $\mathbf{y}^{WE}(R)$. Similarly, let $\mathbf{x}^{POR}(R, R_0)$ be the flow with partially optimal routing in R_0, with corresponding path flows $\mathbf{y}^{POR}(R, R_0)$. Let $X_0 = x_{s_0 t_0}^{POR}(R, R_0)$ represent the flow routed through the subnetwork R_0 under partially optimal routing. Note that $X_0 > 0$ since by assumption POR paradox occurs in R with respect to R_0. Let l_0 denote the effective latency of R_0 under partially optimal routing, and $\tilde{l}_0$ denote the effective latency of R_0 under selfish routing [cf. Eq. (7)].

Define a routing instance $R_0' = (V_0, A_0, P_0, s_0, t_0, X_0, \mathbf{l})$ and let $\mathbf{x}^{WE}(R_0')$ be the Wardrop equilibrium flow for the routing instance R_0'.

We define a new collection of latency functions as follows. For all $j \notin A_0$, define $m_j = l_j$. For $j \in A_0$, we choose a new strictly increasing, nonnegative latency function m_j with $m_j(x_j) \leq l_j(x_j)$ for all $x_j \geq 0$, such that

$$m_j(x_j^{WE}(R_0')) = \frac{l_0(X_0)}{\tilde{l}_0(X_0)} l_j(x_j^{WE}(R_0')).$$

Observe that such a choice is possible, since $l_0(X_0) \leq \tilde{l}_0(X_0)$.

Let $T_0 = (V_0, A_0, P_0, s_0, t_0, X_0, \mathbf{m})$; i.e., T_0 is the routing instance R_0' with latencies replaced by $\mathbf{m}$. We claim that $\mathbf{x}^{WE}(T_0) = \mathbf{x}^{WE}(R_0')$. This follows from the definition of $\mathbf{m}$: all values $m_j(x_j^{WE}(R_0'))$ are proportional to $l_j(x_j^{WE}(R_0'))$, with common constant of proportionality $l_0(X_0)/\tilde{l}_0(X_0)$. Thus if $\mathbf{x}^{WE}(R_0')$ is the Wardrop equilibrium flow with latencies $\mathbf{l}$, it must remain so with latencies $\mathbf{m}$. Furthermore, observe that for any path p with positive flow, we have

$$\sum_{j \in p} m_j(x_j^{WE}(T_0)) = \frac{l_0(X_0)}{\tilde{l}_0(X_0)} \sum_{j \in p} l_j(x_j^{WE}(R_0')) = l_0(X_0),$$

because the second summation above is equal to $\tilde{l}_0(X_0)$. Thus we conclude

$$C(\mathbf{x}^{WE}(T_0)) = \sum_{j \in A_0} x_j^{WE}(T_0) m_j(x_j^{WE}(T_0)) = X_0 l_0(X_0). \tag{8}$$

Let $T = (V, A, P, s, t, \mathbf{X}, \mathbf{m})$. Define a feasible flow $\mathbf{x} = [x_j]_{j \in A}$ as follows:

$$x_j = \begin{cases} x_j^{POR}(R, R_0), & \text{if } j \notin A_0; \\ x_j^{WE}(R_0'), & \text{if } j \in A_0. \end{cases}$$

We claim that $\mathbf{x}^{WE}(T) = \mathbf{x}$. This claim follows easily since we have already established that $\mathbf{x}^{WE}(T_0) = \mathbf{x}^{WE}(R_0')$, and (8) holds. In the flow $\mathbf{x}$ for the routing instance T, the effective latency perceived by any flow crossing the subnetwork R_0 is exactly equal to the partially optimal routing effective latency $l_0(X_0)$ (by (8)). But then since all routing outside the subnetwork R_0 is performed according to $\mathbf{x}^{POR}(R, R_0)$, we conclude that in fact $\mathbf{x}^{WE}(T) = \mathbf{x}$, as required.

Combining the preceding, we obtain

$$\begin{aligned}
&\sum_{j \in A} x_j^{WE}(T) m_j(x_j^{WE}(T)) \\
&\quad = \sum_{j \notin A_0} x_j^{POR}(R, R_0) l_j(x_j^{POR}(R, R_0)) \\
&\qquad\qquad + \sum_{j \in A_0} x_j^{WE}(R_0') m_j(x_j^{WE}(R_0')) \\
&\quad = \sum_{j \notin A_0} x_j^{POR}(R, R_0) l_j(x_j^{POR}(R, R_0)) + X_0 l_0(X_0) \\
&\quad = \sum_{j \in A'} x_j^{POR}(R, R_0) l_j(x_j^{POR}(R, R_0)) \\
&\quad = C(\mathbf{x}^{POR}(R, R_0)).
\end{aligned}$$

We assumed that the POR paradox occurs in R with respect to R_0; thus we obtain from the preceding that

$$C(\mathbf{x}^{WE}(T)) = C(\mathbf{x}^{POR}(R, R_0)) > C(\mathbf{x}^{WE}(R)),$$

implying that Braess' paradox occurs in R centered at R_0. ■

An immediate corollary of the preceding proposition is the following:

Corollary 1 Given a routing instance R, if Braess' paradox does not occur in R, then partially optimal routing with respect to any subnetwork always improves the network performance.

Since Milchtaich, [17], has shown that Braess' paradox does not occur in graphs with the *serial-parallel* structure, this corollary implies that as long as the network under consideration has a serial-parallel structure (for example, a network of parallel links), partially optimal routing *always* improves the overall network performance.

IV. Efficiency of Partially Optimal Routing

We have seen in Example 1 that partially optimal routing can actually worsen performance relative to the Wardrop equilibrium. In this section, we quantify the inefficiency of partially optimal routing. Our metric of efficiency is the *ratio* of the total cost at the social optimum to the total cost at the partially optimal routing solution, $C(\mathbf{x}^{SO})/C(\mathbf{x}^{POR})$. Throughout, we assume that all independently-operated subnetworks can be represented as subgraphs with unique entry and exit points.

We will establish two main theorems. The first provides a tight bound on the loss of efficiency when all latency functions are affine; and the second provides a tight bound on the loss of efficiency when all latency functions are polynomials of bounded degree.

We start with a simple result that compares the worst-case efficiency loss of partially optimal routing with that of selfish routing. These relations will be useful in finding tight bounds on the efficiency loss of partially optimal routing. Recall that $\mathcal{R}^{conv}$, $\mathcal{R}^{aff}$, and $\mathcal{R}^{conc}$ denote the class of all routing instances where latency functions are convex, affine, and concave, respectively.

Proposition 2 (a) For all $\mathcal{R}' \in \{\mathcal{R}^{conv}, \mathcal{R}^{aff}, \mathcal{R}^{conc}\}$, we have

$$\inf_{\substack{R\in\mathcal{R}'\\ R_0\subset R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} \le \inf_{R\in\mathcal{R}'} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))}. \quad (9)$$

(b)

$$\inf_{\substack{R\in\mathcal{R}\\ R_0\subset R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} = \inf_{R\in\mathcal{R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))}. \quad (10)$$

(c)

$$\inf_{\substack{R\in\mathcal{R}^{aff}\\ R_0\subset R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} \ge \inf_{R\in\mathcal{R}^{conc}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))}. \quad (11)$$

Proof:

(a) Given an arbitrary routing instance $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$, simply let R_0 consist of a single link $j \in A$ from the routing instance R, with the corresponding latency function l_j. Then it is clear that $\mathbf{x}^{POR}(R, R_0) = \mathbf{x}^{WE}(R)$; thus for any instance on the right hand side of (9) we have constructed an equivalent instance on the left hand side with the same objective function value, establishing the relation.

(b) The argument in part (a) establishes

$$\inf_{\substack{R\in\mathcal{R}\\ R_0\subset R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} \le \inf_{R\in\mathcal{R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))}.$$

To show the reverse inequality, let $R \in \mathcal{R}$ and $R_0 \subset \mathcal{R}$. Let R' be the equivalent POR instance for R with respect to R_0. Then it can be seen that

$$C(\mathbf{x}^{POR}(R, R_0)) = C(\mathbf{x}^{WE}(R')),$$
$$C(\mathbf{x}^{SO}(R)) = C(\mathbf{x}^{SO}(R')).$$

Hence, for every feasible solution of the optimization problem on the left-hand side of relation (10), we have a feasible solution for the problem on the left-hand side that has the same objective function value, establishing the relation.

(c) This follows by combining the argument in part (b) with Lemma 3.

■

In the remainder of this section we will prove several tight bounds on the efficiency loss of partially optimal routing. We begin by recalling the following key results in the analysis of selfish routing, due to Roughgarden and Tardos [10].

Proposition 3 (Roughgarden-Tardos (2002) [10])

(a)

$$\inf_{R\in\mathcal{R}^{conv}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} = 0.$$

(a) Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$ where l_j is an affine latency function for all $j \in A$. Then,

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} \ge \frac{3}{4}.$$

Furthermore, the bound above is tight.

The first result shows that the worst-case efficiency loss of selfish routing is unbounded, when latency functions are only known to be convex. However, if latency functions are affine, then the proposition guarantees the tight bound on efficiency loss in part (b).

Our main theorem in this section is an extension of the results in Proposition 3 to the setting of partially optimal routing.

Theorem 1

(a)

$$\inf_{\substack{R\in\mathcal{R}^{conv}\\ R_0\subset R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} = 0.$$

(a) Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$ where l_j is an affine latency function for all $j \in A$; and a subnetwork R_0 of R. Then:

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} \ge \frac{3}{4}.$$

Furthermore, the bound above is tight.

Proof: Part (a) of the theorem is an immediate corollary of Proposition 2(a) (for $\mathcal{R}' = \mathcal{R}^{conv}$) and Proposition 3(b).

The remainder of the proof establishes part (b) of the theorem by proving two lemmas. The first provides a tight bound of $3/4$ on the ratio of the optimal routing cost to the selfish routing cost for routing instances in which the latency function of each link is a *concave* function. This lemma is relevant because when all latency functions are affine, the effective latency of any subnetwork under partially optimal routing is concave, as shown in the second lemma.

The proof of the following lemma uses a geometric argument that was used in [18]. This result also follows from the analysis in [11]. Here, we provide an alternative proof, which will be useful in our subsequent analysis.

Lemma 2 Let $R \in \mathcal{R}^{conc}$ be a routing instance where all latency functions are concave. Then,

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} \geq \frac{3}{4}.$$

Furthermore, this bound is tight.

Proof: Consider a routing instance $R \in \mathcal{R}^{conc}$, with $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$. Let x^{WE} be the flow configuration at a Wardrop equilibrium. By Eq. (4), for all feasible solutions $\mathbf{x}$ of Problem (2), we have

$$\begin{aligned} C(\mathbf{x}^{WE}) &= \sum_{j\in A} x_j^{WE} l_j(x_j^{WE}) && (12)\\ &\leq \sum_{j\in A} x_j l_j(x_j^{WE}) && (13)\\ &= \sum_{j\in A} x_j l_j(x_j) + \sum_{j\in A} x_j (l_j(x_j^{WE}) - l_j(x_j)). \end{aligned}$$

We next show that for all $j \in A$, and all feasible solutions $\mathbf{x}$ of Problem (2), we have

$$x_j(l_j(x_j^{WE}) - l_j(x_j)) \leq \frac{1}{4} x_j^{WE} l_j(x_j^{WE}). \qquad (14)$$

If $x_j \geq x_j^{WE}$, then since l_j is nondecreasing, we have $l_j(x_j^{WE}) \leq l_j(x_j)$, establishing the desired relation. Assume next that $x_j < x_j^{WE}$. The term $x_j(l_j(x_j^{WE}) - l_j(x_j))$ is equal to the area of the shaded rectangle in Figure 2. Consider the triangle formed by the three points

$$(0, l_j(x_j^{WE})),\ (0, l_j(x_j) - l_j'(x_j)x_j),$$
$$\left(\frac{l_j(x_j^{WE}) - l_j(x_j) + l_j'(x_j)x_j}{l_j'(x_j)}, l_j(x_j^{WE})\right).$$

Denote this triangle by T. It can be seen that

$$x_j(l_j(x_j^{WE}) - l_j(x_j)) \leq \frac{1}{2}\text{Area}(T).$$

By the concavity of l_j, we further have

$$\text{Area}(T) \leq \int_0^{x_j^{WE}} \int_{l_j(x)}^{l_j(x_j^{WE})} dy dx \leq \frac{x_j^{WE} l_j(x_j^{WE})}{2},$$

i.e., the area of triangle T is less than or equal to the area between the curves $y = l_j(x)$ and $y = l_j(x_j^{WE})$ in the interval $x \in [0, x_j^{WE}]$, which in turn is less than or equal to half of the area $x_j^{WE} l_j(x_j^{WE})$. Combining the preceding two relations, we obtain Eq. (14), which implies $\sum_{j\in A} x_j(l_j(x_j^{WE}) - l_j(x_j)) \leq (1/4)\sum_{j\in A} x_j^{WE} l_j(x_j^{WE}) = (1/4)C(\mathbf{x}^{WE})$.

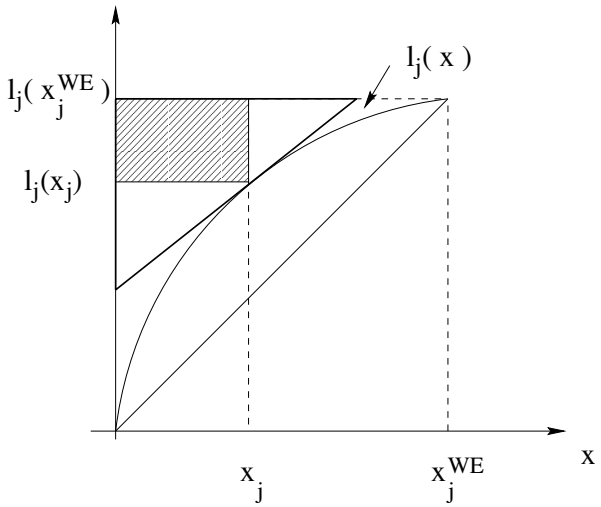

Fig. 2. Illustration of the proof of Lemma 2.

Combining with Eq. (13), we see that for all feasible solutions $\mathbf{x}$ of Problem (2), we have

$$\frac{3}{4} C(\mathbf{x}^{WE}) \leq \sum_{j\in A} x_j l_j(x_j).$$

Since the socially optimal flow configuration $\mathbf{x}^{SO}$ is a feasible solution for Problem (2), we obtain the desired result. ■

The following lemma, which establishes that the effective latency l_0 of a subnetwork under partially optimal routing is concave when the latency functions are affine, completes the proof of part (b) of the theorem.

Lemma 3 Let $R_0 = (V_0, A_0, P_0, s_0, t_0)$ be a subnetwork. Assume that the latency functions of all links in the subnetwork are nonnegative affine functions, i.e., for all $j \in A_0$, $l_j(x_j) = a_j x_j + b_j$, where $a_j \geq 0$ and $b_j \geq 0$. Let $l_0(X_0)$ denote the effective latency of partially optimal routing of X_0 units of flow in the subnetwork R_0. Then $l_0(X_0)$ is a concave function of X_0.

Proof of Lemma: Since the l_j are affine, for all $X_0 \geq 0$, we have

$$\begin{aligned} l_0(X_0) = \quad & \min_{y_p \geq 0,\ p\in P} \quad \sum_{j\in A_0} \frac{a_j x_j^2}{X_0} + \frac{b_j x_j}{X_0} \\ & \text{subject to} \quad \sum_{p\in P_0 : j\in p} y_p = x_j, \quad j \in A_0; \\ & \qquad\qquad \sum_{p\in P_0} y_p = X_0. \end{aligned}$$

Using the change of variables $\hat{y}_p = \frac{y_p}{X_0}$ for all $p \in P_0$, and $\hat{x}_j = \frac{x_j}{X_0}$ for all $j \in A_0$ in the preceding optimization problem,

we obtain

$$l_0(X_0) = \min_{\hat{y}_p \geq 0,\ p \in P_0} \sum_{j \in A_0} a_j X_0 \hat{x}_j^2 + b_j \hat{x}_j \qquad (15)$$
$$\text{subject to} \sum_{p \in P_0 : j \in p} \hat{y}_p = \hat{x}_j, \quad j \in A_0;$$
$$\sum_{p \in P_0} \hat{y}_p = 1.$$

Denote the feasible set of problem (15) by Y, i.e.,

$$Y = \left\{ \mathbf{y} \;\middle|\; y_p \geq 0,\ \forall\ p \in P_0,\ \sum_{p \in P_0} y_p = 1 \right\}.$$

Then by defining $x_j(\mathbf{y}) = \sum_{p \in P_0 : j \in p} y_p$, we can write (15) equivalently as:

$$l_0(X_0) = \inf_{\mathbf{y} \in Y} \left[\left(\sum_{j \in A_0} a_j x_j(\mathbf{y})^2 \right) X_0 + \left(\sum_{j \in A_0} b_j x_j(\mathbf{y}) \right) \right].$$

But now observe that $l_0(X_0)$ is the infimum of a collection of affine functions of X_0. By a standard result in convex analysis (see, e.g., [19], Proposition 1.2.4(c)), it follows that $l_0(X_0)$ is concave. ■

Combining Lemmas 2 and 3 with Proposition 2 completes the proof of part (b) of Theorem 1. ■

The preceding proof exploits the fact that the effective latency l_0 is concave in the subnetwork to establish a tight efficiency loss bound for partially optimal routing with respect to the social optimum, under the assumption that all latency functions are affine. We can apply a similar approach to develop bounds on the efficiency loss of partially optimal routing even when the latency functions may not be affine; our starting point is a result of Correa et al. [18], extending earlier work of Roughgarden [20], that gives bounds on the efficiency loss of selfish routing with general latency functions.

To state their result, we require the following definitions. Given a class of latency functions $\mathcal{L}$, we define $\beta(\mathcal{L})$ as:

$$\beta(\mathcal{L}) = \sup_{l \in \mathcal{L},\ x \geq 0} \beta(l, x), \qquad (16)$$

with

$$\beta(l, x) = \max_{z \geq 0} \frac{(l(x) - l(z))z}{l(x)x}, \qquad (17)$$

Intuitively β is measure of the steepness of a class of latency functions; for all the cases we will consider, it is equivalent to $1 - 1/\alpha(\mathcal{L})$, where $\alpha(\mathcal{L})$ is the steepness parameter defined by Roughgarden [20]. The following proposition was first proven by Roughgarden [20] for convex and differentiable latency functions, and then extended by Correa et al. to all classes of latency functions [18].

Proposition 4 Let $\mathcal{L}$ be a class of separable latency functions. Consider a routing instance $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$ with $l_j \in \mathcal{L}$ for all $j \in A$. Then

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} \geq (1 - \beta(\mathcal{L})).$$

Furthermore, the bound above is tight.

We emphasize that $\beta(\mathcal{L}) = 1/4$ when $\mathcal{L}$ is the class of affine latency functions, so the preceding proposition is indeed a generalization of Proposition 3.

In the spirit of Proposition 4, the following theorem generalizes the results of Theorem 1 to networks where latencies are nonnegative polynomials.

Theorem 2 Let $\mathcal{L}_d$ be a class of nonnegative separable polynomial latency functions of degree d. Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$ with $l_j \in \mathcal{L}_d$ for all $j \in A$, and a subnetwork R_0 of R. Then,

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R, R_0))} \geq (1 - \beta(\mathcal{L}_d)),$$

where $\beta(\mathcal{L}_d)$ is defined in Eqs. (16)-(17). Furthermore, the bound above is tight.

Proof: The proof proceeds as follows. First, we establish a bound on the efficiency loss of a routing instance where each latency function is the pointwise infimum of a collection of latency functions in a prespecified class. Then, we establish that if all latency functions are polynomials, the effective latency of any subnetwork under partially optimal routing is an infimum of polynomials. Combining these claims will yield the result of the proposition.

Our starting point is the following observation of Correa et al. ([18], particularly Lemma 2.3): given a class of separable latency functions $\mathcal{L}$ and a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$ with $l_j \in \mathcal{L}$ for all $j \in A$, the following inequality holds:

$$x_j l_j(x_j^{WE}(R)) \leq x_j l_j(x_j) + \beta(\mathcal{L}) x_j^{WE}(R) l_j(x_j^{WE}(R)),$$
$$\forall\ j \in A,\ \forall\ x \geq 0. \quad (18)$$

Using this fact, we can prove the following lemma.

Lemma 4 Let $\mathcal{L}_s$ be a class of nonnegative separable latency functions which is closed under scaling by a constant $k \leq 1$ (i.e., for all $l \in \mathcal{L}_s$, we have $kl \in \mathcal{L}_s$ for all scalars $k \leq 1$). Let $R = (V, A, P, s, t, \mathbf{X}, \mathbf{l})$ be a routing instance with

$$l_j(x) = \inf_{z \in \mathcal{Z}_j} \{f(x, z)\}, \qquad \forall\ j \in A, \qquad (19)$$

where: $\mathcal{Z}_j$ is a compact set; for each x, $f(x, \cdot)$ is a continuous function of z; and for each $z \in \mathcal{Z}_j$, $f(\cdot, z) \in \mathcal{L}_s$. Then:

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} \geq (1 - \beta(\mathcal{L}_s)).$$

Proof of Lemma. We use the bound in Eq. (18) together with a similar geometric argument used in the proof of Lemma 2 to prove the result.

Let $\mathbf{x}$ be a feasible solution of problem (2). By the definition of l_j [cf. Eq. (19)], for all $j \in A$, there exists some $\bar{z}_j$ such that

$$l_j(x_j) = f(x_j, \bar{z}_j), \qquad (20)$$

and

$$l_j(x_j^{WE}(R)) \leq f(x_j^{WE}(R), \bar{z}_j). \qquad (21)$$

Let $c = \frac{f(x_j^{WE}(R), \bar{z}_j)}{l_j(x_j^{WE}(R))} \geq 1$ and define

$$\bar{f}_j(y) = \frac{f(y, \bar{z}_j)}{c}, \qquad \forall\ y \geq 0.$$

Since $c \geq 1$ and $f(y, \bar{z}_j) \geq 0$, it follows that $\bar{f}_j(y) \leq f(y, \bar{z}_j)$ for all $y \geq 0$, which by Eq. (20) implies that

$$\bar{f}_j(x_j) \leq l_j(x_j).$$

In view of the assumption that the class $\mathcal{L}_s$ is closed under scaling by a constant $k \leq 1$, we have $\bar{f}_j \in \mathcal{L}_s$. Moreover, since $\bar{f}_j(x_j^{WE}(R)) = l_j(x_j^{WE}(R))$, $\mathbf{x}^{WE}(R)$ is a Wardrop equilibrium of the routing instance $\bar{R} = (V, A, P, s, t, X, \bar{\mathbf{l}})$, where $\bar{l}_j = \bar{f}_j$ for all $j \in A$. Combining the preceding, we obtain

$$\begin{aligned} x_j\Big(l_j(x_j^{WE}(R)) \quad &- \quad l_j(x_j)\Big) \\ &\leq \quad x_j\Big(\bar{f}_j(x_j^{WE}(R)) - \bar{f}_j(x_j)\Big) \\ &\leq \quad \beta(\mathcal{L}_s) C(\mathbf{x}^{WE}(R)), \end{aligned}$$

Summing over all $j \in A$ and using an argument similar to the proof of Lemma 2 [in particular Eq. (13)], we obtain the desired result. ■

The following lemma characterizes effective latencies in the special case where all latency functions are nonnegative polynomials. The proof is similar to the proof of Lemma 3, and is omitted.

Lemma 5 Let $R_0 = (V_0, A_0, P_0, s_0, t_0)$ be a subnetwork. Assume that the latency functions of all links in the subnetwork are nonnegative polynomials of degree d, i.e., for all $j \in A_0$, $l_j(x_j)$ is a polynomial of degree d such that $l_j(x_j) \geq 0$ for all $x_j \geq 0$. Let $l_0(X_0)$ denote the effective latency of partially optimal routing of X_0 units of flow in the subnetwork R_0. Then $l_0(X_0)$ is given by

$$l_0(X_0) = \inf_{y \in \mathcal{Y}} \{f(X_0, y)\},$$

where $\mathcal{Y}$ is a nonempty compact set, $f(X_0, y)$ is a continuous function of y, and for each $y \in \mathcal{Y}$, $f(\cdot, y)$ is a nonnegative polynomial of degree d.

Since the class of polynomial functions is closed under scaling by a constant, the preceding two lemmas immediately imply the conclusion of the theorem, as required. ■

The preceding theorem, together with Theorem 1, gives a tight characterization of the efficiency loss of partially optimal routing when all independently-operated subnetworks have unique entry and exit points. In particular, observe that, as long as latency functions are polynomial of bounded degree, the worst-case efficiency loss under partially optimal routing is no worse than the same worst-case value for pure selfish routing.

V. Subnetworks with Multiple Entry and Exit Points

We now turn to a discussion of the efficiency of partially optimal routing when independently-operated subnetworks can have multiple entry and exit points. Unfortunately, in this case efficiency loss may be unbounded for even in the most restrictive case where latency functions are *linear*. The next example shows that the efficiency loss may be unbounded for networks that include subnetworks with multiple entry and exit points.

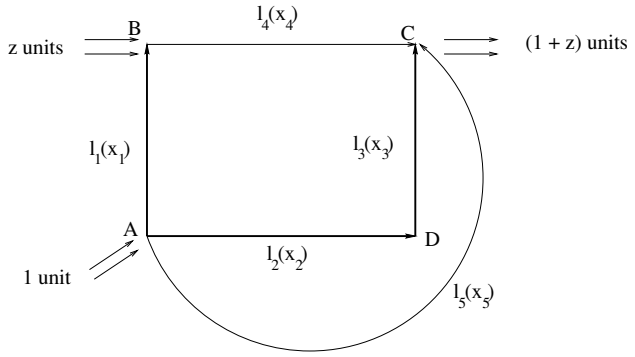

Fig. 3. A subnetwork with multiple exit points.

Example 2 Consider the general network illustrated in Figure 3.

The subnetwork consists of links 1, 2, and 3 with latency functions

$$l_1(x_1) = 0, \quad l_2(x_2) = ax_2, \quad l_3(x_3) = 0,$$

for some $a > 0$. The remaining links in the network have latency functions

$$l_4(x_4) = bx_4, \qquad l_5(x_5) = x_5,$$

for some $b > a > 0$.

The link flows at the social optimum are given by:

$$\mathbf{x}^{SO} = \left(0, \frac{1}{1+a}, \frac{1}{1+a}, z, \frac{a}{1+a}\right).$$

The cost of the optimal solution is

$$C(\mathbf{x}^{SO}) = bz^2 + \frac{a}{(1+a)^2} + \frac{a^2}{(1+a)^2}.$$

For any $a > 0$, the optimal routing for the subnetwork is to route all the incoming flow along link 1. To find the POR equilibrium, let x_1 denote the amount of traffic that is routed over the subnetwork (i.e., along link 1). Assuming that $bz < 1$, we solve for x_1 in the Wardrop conditions

$$b(z + x_1) = (1 - x_1).$$

Hence the link flows at the POR equilibrium are given by:

$$\mathbf{x}^{POR} = \left(\frac{1-bz}{1+b}, 0, 0, \frac{1+z}{1+b}, \frac{b+bz}{1+b}\right).$$

The cost of the POR equilibrium is

$$C(\mathbf{x}^{POR}) = b\left(\frac{1+z}{1+b}\right)^2 + \left(\frac{b+bz}{1+b}\right)^2.$$

For a fixed $b > 0$, taking the limit as $a \to 0$ and $z \to 0$, we obtain

$$C(\mathbf{x}^{SO}) \to 0, \qquad C(\mathbf{x}^{POR}) \to \frac{b}{1+b} > 0,$$

thus showing that the relative efficiency approaches zero.

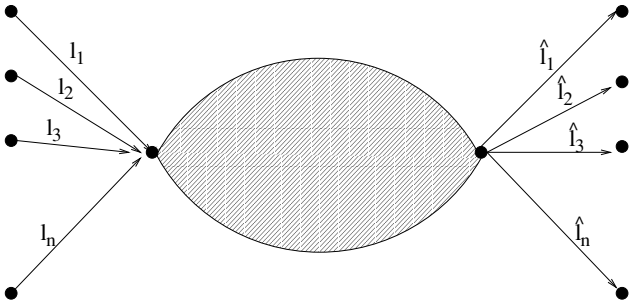

Fig. 4. A special multiple entry-exit subnetwork, where the shaded area represents a single entry-exit subnetwork with an arbitrary topology.

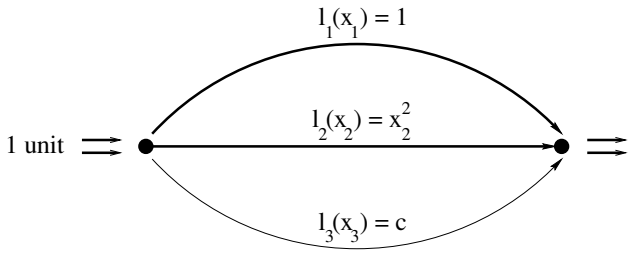

Fig. 5. A parallel link network. Links 1 and 2 form a subnetwork that is controlled by an independent administrator.

Nevertheless, there are at least two important special cases where one can show that the efficiency loss of partially optimal routing is bounded. The first is illustrated by the network topology in Figure 4. Let $s_1, \ldots, s_n$ be n entry points, and $t_1, \ldots, t_n$ be n exit points. Assume that x_i units of flow enters the network at node s_i, and is destined to node t_i, for all $i = 1, \ldots, n$. The shaded area represents a single entry-exit subnetwork with an arbitrary topology.

For this structure, it is evident that the POR effective latency between nodes s_i and t_i is given by

$$\tilde{l}_i(x_i, X) = l_i(x_i) + \hat{l}_i(x_i) + l_0(X),$$

where $X = \sum_{i=1}^{n} x_i$, and l_0 is the effective latency of optimal routing within the single entry-exit subnetwork, defined as in our previous analysis.

Given this structure, immediate corollaries of Theorems 1 or 2 imply that, under the assumptions of these theorems, exactly the same results apply to the model of Figure 4.

A bound on efficiency loss can also be provided in the case when the following three conditions are satisfied. First, we assume all latencies in the entire network are affine. Second, we assume all latencies in the subnetwork under consideration are *linear*: i.e., $\ell_j(x_j) = b_j x_j$ for all j in the subnetwork. Third, we assume that every source-destination pair has a unique entry to, and exit from, the subnetwork. Because of the latter assumption, it is straightforward to establish that a Wardrop equilibrium in the global network leads to a Wardrop equilibrium within the subnetwork. Furthermore, the Wardrop equilibrium within the subnetwork must optimize intradomain performance. The latter claim follows because when latencies are linear, the social optimum and Wardrop equilibrium coincide in view of the fact that the objective function for (2) differs only by a factor of $1/2$ from the objective function for (1). As a result, the POR equilibrium in this setting is equivalent to the Wardrop equilibrium, and thus the efficiency loss is bounded by $3/4$ (cf. Proposition 3).

VI. Partially Optimal Routing and Subnetwork Performance

In this section, we consider a model where a subnetwork chooses its routing policy to achieve the minimum (total) latency within its subnetwork. This amounts to assuming that the subnetwork ignores revenues from transmission, which is natural in this context, since we have not considered the pricing decisions of service providers (see Concluding Remarks). While optimal routing seems like the natural means to achieve this goal, end-to-end route selection may counteract any expected performance gains from this type of intradomain traffic engineering. As a result, the provider may prefer to allow traffic to route selfishly in order to reduce flow and total delay in its subnetwork. The following example illustrates this scenario.

Example 3 Consider the parallel-link network illustrated in Figure 5. The latency functions are given by $l_1(x_1) = 1, l_2(x_2) = x_2^2, l_3(x_3) = c$, for some constant c, $0 < c < 1$. Assume that links 1 and 2 form a subnetwork, denoted by G_0, which is controlled by an independent administrator. Assume that one unit of flow is to be routed over this network.

Assume first that the flow through the subnetwork G_0 is routed selfishly, i.e., according to Wardrop equilibrium. It can be seen in this case that $\sqrt{c}$ units of traffic is routed through the subnetwork, leading to a total cost of $C(\mathbf{x}^{WE}) = c$, and a subnetwork cost of $C_{G_0}(\mathbf{x}^{WE}) = c\sqrt{c}$.

Assume next that the flow through the subnetwork G_0 is routed optimally, i.e., the flow is routed through the overall network according to POR equilibrium. Assume that the constant $c \in \left[1 - \frac{2}{3\sqrt{3}}, 1\right]$. It can be seen in this case that the entire traffic is routed through the subnetwork, leading to a total and subnetwork cost of $C(\mathbf{x}^{POR}) = C_{G_0}(\mathbf{x}^{POR}) = 1 - \frac{2}{3\sqrt{3}}$. Note that for $c\sqrt{c} < 1 - \frac{2}{3\sqrt{3}}$, we have $C_{G_0}(\mathbf{x}^{POR}) > C_{G_0}(\mathbf{x}^{WE})$.

As the preceding example demonstrates, lower-layer traffic engineering may prefer selfish to optimal routing. It is equally easy to construct examples where optimal routing will be preferred. The simplest example is a situation in which the total traffic entering the subnetwork is constant, regardless of whether selfish or optimal routing is used. This will be the case in the example above when $c > 1$, and a similar analysis immediately implies that optimal routing will be preferred within the subnetwork in this case.

To gain more insights, let us next consider a "partial equilibrium" analysis of routing within a subnetwork, taking the strategies of all other subnetworks as given. To illustrate the main issues, we consider a network consisting of parallel links between a single origin-destination pair with d units of total traffic. Suppose that there are $N + 1$ providers and each network provider owns a subset of the links in the network. We represent network provider i, for $i = 1, \ldots, N$, by a single link with effective latency l_i (corresponding to the intradomain routing policy chosen by provider i, whether optimal routing

or not). We assume all these latency functions l_i are continuous and strictly increasing.

As in the preceding discussion, we assume that if provider 0 pursues an optimal intradomain routing policy, then the effective latency is given by l_0, and if provider 0 allows purely selfish routing within his network (corresponding to a Wardrop equilibrium), the effective latency is $\tilde{l}_0$. To simplify the discussion here, let us also assume that l_0 and $\tilde{l}_0$ are both continuous and strictly increasing (this will be the case, for example, when all latency functions of the links in the subnetwork are continuous and strictly increasing). As before, recall that $\tilde{l}_0(x) \geq l_0(x)$ for all $x \geq 0$. Moreover, for simplicity, let us assume that $\tilde{l}_0(x) > l_0(x)$ if $x > 0$ (though the arguments can be generalized to the case without this assumption).

We assume that the subnetwork owner can randomize between the two policies, so any convex combination of optimal and selfish routing can be achieved. In other words, the subnetwork owner chooses a $\delta \in [0,1]$ corresponding to an effective latency given by:

$$m_0(x,\delta) = (1-\delta)\, l_0(x) + \delta \tilde{l}_0(x),$$

where $\delta = 0$ corresponds to optimal routing, while $\delta = 1$ corresponds to selfish routing.

We continue to use $\mathbf{x}^{POR}$ to denote a Wardrop equilibrium with respect to the latency functions $m_0, l_1, \ldots, l_N$, so that $\mathbf{x}^{POR}$ satisfies:

$$\begin{aligned} m_0(x_0^{POR},\delta) &\geq \lambda; \\ l_i\left(x_i^{POR}\right) &\geq \lambda \text{ for } i = 1, ..., N; \\ \sum_{i=0}^{N} x_i^{POR} &= d\ ; \\ x_i^{POR} &\geq 0 \text{ for } i = 0, ..., N; \\ \lambda &= \min\left\{m_0\left(x_0^{POR},\delta\right), \right. \\ &\qquad \left. l_1\left(x_1^{POR}\right), \ldots, l_N(x_N^{POR})\right\}. \end{aligned}$$

First consider the routing of flow through the links $1, \ldots, N$. If a total flow x is routed through links $1, \ldots, N$, then the resulting flow allocation must satisfy:

$$l_i(x_i) = \min\{l_1(x_1), \ldots, l_N(x_N)\} \text{ if } x_i > 0; \tag{22}$$

$$\sum_{i=1}^{N} x_i = x; \tag{23}$$

$$x_i \geq 0,\ i = 1, \ldots, N. \tag{24}$$

In view of the assumption that $l_1, \ldots, l_N$ are strictly increasing, the preceding equations have a unique solution. We define $l_R(x)$ as the latency at this solution, i.e.,

$$l_R(x) = \min\{l_1(x_1), \ldots, l_N(x_N)\},$$

where $(x_1, \ldots, x_N)$ is the unique solution to (22)-(24). Since each l_i is strictly increasing and continuous, the function l_R is also strictly increasing and continuous.

Next consider the traffic engineering problem faced by subnetwork 0. The network provider will choose a value of δ that minimizes the total latency inside the subnetwork, given that traffic will follow the Wardrop equilibrium pattern for the resulting effective latencies. Formally, the optimization problem of subnetwork 0 is the following:

$$\min_{0 \leq x_0 \leq d, \delta \in [0,1]} \left((1-\delta)\, l_0(x_0) + \delta \tilde{l}_0(x_0)\right) x_0 \tag{25}$$

subject to

$$\begin{aligned} (1-\delta)\, l_0(0) + \delta \tilde{l}_0(0) &\geq l_R(d), \text{ if } x_0 = 0; \\ (1-\delta)\, l_0(d) + \delta \tilde{l}_0(d) &\leq l_R(0), \text{ if } x_0 = d; \\ (1-\delta)\, l_0(x_0) + \delta \tilde{l}_0(x_0) &= l_R(d - x_0), \text{ if } 0 < x_0 < d. \end{aligned}$$

Since $\tilde{l}_0(x_0) \geq l_0(x_0)$ for all $x_0 \geq 0$, and l_R is strictly increasing, as δ increases from $\delta = 0$ (purely optimal routing) to $\delta = 1$ (purely selfish routing), the flow routed through subnetwork 0 at the POR equilibrium must be nonincreasing.

Next note that when $\tilde{l}_0(0) \geq l_R(d)$, the subnetwork can achieve the minimum total latency of zero by choosing $\delta = 1$ (since the POR equilibrium will route no traffic across subnetwork 0). Similarly, if $\tilde{l}_0(d) \leq l_R(0)$, then regardless of provider 0's policy, all the flow will be routed across subnetwork 0. As a result, in this scenario the optimal strategy is $\delta = 0$ (optimal routing), as this minimizes the total latency. For the remainder of this section, we assume that $\tilde{l}_0(0) < l_R(d)$ and $\tilde{l}_0(d) > l_R(0)$. Since $\tilde{l}_0 \geq l_0$, this also implies

$$l_0(0) \leq \tilde{l}_0(0) < l_R(d). \tag{26}$$

We now proceed to define the maximum and minimum flow that will flow through subnetwork 0 over all possible choices of routing policy. The condition (26), together with the fact that $\tilde{l}_0$ and l_R are strictly increasing and continuous, ensures that the following equation has a unique solution $x_0^{MIN} > 0$:

$$\tilde{l}_0(x_0^{MIN}) = l_R(d - x_0^{MIN}).$$

Moreover, given our assumptions, x_0^{MIN} is the minimum flow that can go through subnetwork 0 (achieved exactly when $\delta = 1$, i.e., at purely selfish routing).

The maximum possible flow through subnetwork 0 will depend on the relative values of $l_0(d)$ and $l_R(0)$. Formally, we define x_0^{MAX} as follows. If $l_0(d) \leq l_R(0)$, then we let $x_0^{MAX} = d$, since choosing $\delta = 0$ (optimal routing) will lead to all traffic flowing through subnetwork 0. On the other hand, if $l_0(d) > l_R(0)$, we let x_0^{MAX} be the unique solution to the following equation:

$$l_0(x_0^{MAX}) = l_R(d - x_0^{MAX}), \quad \text{if } l_0(d) > l_R(0).$$

With these definitions, x_0^{MAX} is the maximum flow for subnetwork 0 (achieved exactly when $\delta = 0$, i.e., at optimal routing). We define $c_0(x_0^{MAX})$ as the cost to the owner of subnetwork 0 at this optimal routing; i.e.,

$$c_0(x_0^{MAX}) = \begin{cases} d l_0(d), & \text{if } l_0(d) \leq l_R(0); \\ x_0^{MAX} l_R(d - x_0^{MAX}), & \text{if } l_0(d) > l_R(0). \end{cases}$$

Clearly, any flow $x_0 \in [x_0^{MIN}, x_0^{MAX}]$ is achievable. To achieve a flow $x_0 \in [x_0^{MIN}, x_0^{MAX})$, the owner of subnetwork should choose δ such that:

$$\delta = \frac{l_R(d - x_0) - l_0(x_0)}{\tilde{l}_0(x_0) - l_0(x_0)}, \tag{27}$$

where $0 \leq \delta \leq 1$ since: (1) the definition of x_0^{MAX} ensures that $l_R(d - x_0) > l_0(x_0)$; and (2) $\tilde{l}_0(x_0) > l_0(x_0)$ for all

$x_0 > 0$. Finally, using the definition of δ in (27), observe that for all $x_0 \in [x_0^{MIN}, x_0^{MAX})$, we have the relation $m_0(x_0, \delta) = l_R(d - x_0)$. In other words, if the subnetwork owner chooses policy δ according to (27), the flow through the subnetwork will be x_0, and the resulting latency will be $l_R(d - x_0)$.

As a result, the optimization problem for the owner of subnetwork 0 becomes:

$$\min \left\{ \min_{x_0 \in [x_0^{MIN}, x_0^{MAX})} \left[x_0 l_R(d - x_0) \right], c_0(x_0^{MAX}) \right\}, \quad (28)$$

the solution of which determines the delay-minimizing routing policy of the subnetwork. If the solution yields $x_0 \in [x_0^{MIN}, x_0^{MAX})$, the subnetwork owner should choose δ in accordance with (27). If the solution yields $x_0 = x_0^{MAX}$, the subnetwork owner should choose $\delta = 0$ (pure optimal routing). If the game between service providers is one of complete information, all the latency functions are common knowledge and the owner of the subnetwork can compute x_0^{MIN}, x_0^{MAX}, and l_R, and hence the optimal flow through the subnetwork. If we assume that $l_0(d) > l_R(0)$, we can intuitively understand the solution: If $x_0 l_R(d - x_0)$ increases as x_0 increases in the neighborhood of x_0^{MIN}, the provider will (locally) prefer selfish routing. Similarly, if $x_0 l_R(d - x_0)$ decreases as x_0 decreases in the neighborhood of x_0^{MAX}, the provider prefers selfish routing.

This analysis shows that with complete information and a parallel-link network, the delay-minimizing policy of the network is straightforward to characterize. We leave the analysis of networks with more general topologies for future work.

VII. Concluding Remarks

This paper provides a model of partially optimal routing that captures the essential features of the interaction between traffic engineering, and selfish routing of end-to-end flows. While source-destination pairs transmit flows across the least cost paths, service providers controlling the subnetworks use traffic engineering to reduce delay within their own administrative domains. End-users perceive the delays resulting from the traffic engineering of the network providers. We formulate and analyze the equilibria of this global network with partially optimal routing.

Even though traffic engineering within parts of the overall network may be conjectured to reduce congestion externalities and improve overall network performance, we show this not to be the case. In particular, if the global network exhibits the Braess' paradox, traffic engineering that reduces delays within a subnetwork may worsen the performance of the overall network. More specifically, we prove that if partially optimal routing leads to an increase in overall delay relative to selfish routing over all links, Braess' paradox must occur in the global network.

Much of the paper quantifies the potential inefficiency of partially optimal routing relative to the system optimum in the case where all independently-operated subnetworks have single entry-exit points and delays can be modeled by latency functions of a specific class, such as affine or nonnegative polynomials of bounded degree. For example, with affine latency functions, we establish that the performance of partially optimal routing is no worse than 25% relative to the system optimum.

In contrast to these results that match the corresponding bounds for selfish routing throughout the whole network, when subnetworks have multiple entry-exit points, the performance of partially optimal routing can be arbitrarily bad, even with linear latencies. This result suggests that special care needs to be taken in the regulation of traffic in large-scale networks overlaying selfish source routing together with traffic engineering within subnetworks.

We also provide conditions for service providers to prefer to engage in traffic engineering rather than allowing all traffic to route selfishly within their network. The latter is a possibility because selfish routing may discourage entry of further traffic into their subnetwork, reducing total delays within the subnetwork, which may be desirable for the network provider when there are no prices per unit of transmission.

We believe that the model of partially optimal routing presented in this paper is a good approximation to the functioning of large-scale communication networks, such as the Internet, and raises a number of interesting questions for further investigation. Possible areas of further study include:

1) An average-case analysis for an appropriate stochastic model of traffic demands, rather than worst-case analysis.
2) Quantification of the loss of efficiency of partially optimal routing relative to selfish routing throughout the entire network.
3) Analysis of simple regulation schemes that can prevent realization of worst-case performance losses in networks with partially optimal routing.
4) Quantification of the loss of efficiency of partially optimal routing relative to the system optimum in specific network topologies incorporating subnetworks with multiple entry-exit points.
5) Analysis of the equilibrium of routing patterns when multiple service providers simultaneously and strategically decide the extent of traffic engineering.
6) Analysis of partially optimal routing when service providers use traffic engineering for objectives other than minimizing total delay (e.g., loss minimization or fault tolerance).
7) Analysis of partially optimal routing when service providers do not simply minimize total delay, but charge for transmission through their networks and maximize profits, taking into account the impact of delays within their network on their revenues.

Acknowledgments

The authors would like to thank Nicolas Stier-Moses for helpful discussions on the proof of Lemma 4, as well as the anonymous referees for their helpful comments on the initial draft.

References

[1] D. Acemoglu, R. Johari, and A. Ozdaglar, "Paradoxes of traffic engineering with partially optimal routing," in *Conference on Information Sciences and Systems*, 2006.

[2] J. Rexford, *Route optimization in IP networks*. Springer Science and Business Media, 2006, pp. 679–700.
[3] L. Qiu, Y. Yang, Y. Zhang, and S. Shenker, "On selfish routing in internet-like environments," in *Proceedings of ACM SIGCOMM*, 2003, pp. 151–162.
[4] H. Zhang, Y. Liu, W. Gong, and D. Towsley, "Understanding the interaction between overlay routing and traffic engineering," University of Massachusetts CMPSCI, Technical Report 04-63, 2004.
[5] ——, "On the interaction between overlay routing and underlay routing," in *Proceedings of IEEE Infocom*, 2005, pp. 2543–2553.
[6] H. Han, S. Shakkottai, C. Hollot, R. Srikant, and D. Towsley, "Overlay TCP for multi-path routing and congestion control," 2004, submitted.
[7] F. P. Kelly and T. Voice, "Stability of end-to-end algorithms for joint routing and rate control," *Computer Communication Review*, vol. 35, no. 2, pp. 5–12, 2005.
[8] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. New Haven, Connecticut: Yale University Press, 1956.
[9] A. Pigou, *The Economics of Welfare*. London, U.K.: Macmillan, 1920.
[10] T. Roughgarden and É.. Tardos, "How bad is selfish routing?" *Journal of the ACM*, vol. 49, no. 2, pp. 236–259, 2002.
[11] J. Correa, A. Schulz, and N. Stier-Moses, "Selfish routing in capacitated networks," *Mathematics of Operations Research*, vol. 29, no. 4, pp. 961–976, 2002.
[12] F. P. Kelly, "Network routing," *Philosophical Transactions: Physical Sciences and Engineering*, vol. 337, no. 1647, pp. 343–367, 1991.
[13] S. Dafermos and F. Sparrow, "The traffic assignment problem for a general network," *Journal of Research of the National Bureau of Standards-B. Mathematical Sciences*, vol. 73, no. 2, pp. 91–118, 1969.
[14] M. Smith, "The existence, uniqueness and stability of traffic equilibria," *Transportation Research*, vol. 13B, pp. 295–304, 1979.
[15] B. Fortz and M. Thorup, "Internet traffic engineering by optimizing OSPF weights," in *Proceedings of IEEE INFOCOM*, 2000, pp. 519–528.
[16] D. Acemoglu, R. Johari, and A. Ozdaglar, "Partially optimal routing," MIT Laboratory for Information and Decisions Systems, Technical Report 2703, 2006.
[17] I. Milchtaich, "Network topology and the efficiency of equilibrium," Department of Economics, Bar-Ilan University, Working Paper 12-01, 2005.
[18] J. Correa, A. Schulz, and N. Stier-Moses, "On the inefficiency of equilibria in congestion games," in *Proceedings of the 11th Conference on Integer Programming and Combinatorial Optimization*, vol. 3509, 2005, pp. 167–181.
[19] D. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*. Cambridge, Massachusetts: Athena Scientific, 2003.
[20] T. Roughgarden, "The price of anarchy is independent of the network topology," in *Proceedings of ACM Symposium on the Theory of Computing*, 2002, pp. 428–437.

Daron Acemoglu is Charles P. Kindleberger Professor of Applied Economics in the Department of Economics at the Massachusetts Institute of Technology and a member of the Economic Growth program of the Canadian Institute of Advanced Research. He is also affiliated with the National Bureau Economic Research, Center for Economic Performance, and Center for Economic Policy Research, and is a fellow of the American Academy of Arts and Sciences, the Econometric Society, the European Economic Association and the Society of Labor Economists.

Daron Acemoglu has received a BA in Economics at the University of York, 1989, M.Sc. in Mathematical Economics and Econometrics at the London School of Economics, 1990, and Ph.D. in Economics at the London School of Economics in 1992. Since 1993, he has held the academic positions of Lecturer at the London School of Economics, and Assistant Professor, Pentti Kouri Associate Professor and Professor of Economics at MIT. He has received numerous awards and fellowships, including the award for best paper published in the Economic Journal in 1996 for his paper "Consumer Confidence and Rational Expectations: Are Agents' Beliefs Consistent With the Theory?", the inaugural T. W. Shultz prize at the University of Chicago in 2004, and the inaugural Sherwin Rosen award for outstanding contribution to labor economics in 2004.

He was awarded the John Bates Clark Medal in 2005, given every two years to the best economist in the United States under the age of 40 by the American Economic Association. Daron Acemoglu is the co-editor of the Review of Economics and Statistics (until July 2007), NBER Macroannual, and Econometrica (from July 2007).

Ramesh Johari (M '05) received the A.B. degree in Mathematics from Harvard University (1998), the Certificate of Advanced Study in Mathematics from University of Cambridge (1999), and the Ph.D. in Electrical Engineering and Computer Science from M.I.T. (2004). He is currently an Assistant Professor of Management Science and Engineering, and by courtesy, Electrical Engineering, at Stanford University. His research interests include game theory, optimization, and competition and cooperation in networked systems.

Asuman Ozdaglar (M '95) received the B.S. degree in Electrical Engineering from the Middle East Technical University in 1996, and the S.M. and the Ph.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (MIT) in 1998 and 2003, respectively. Since 2003, she has been with the Department of Electrical Engineering and Computer Science at MIT, where she is currently the Class of 43 Career Development Assistant Professor. She is affiliated with the Laboratory for Information and Decision Systems, the Operations Research Center, and the Computation for Design and Optimization Program at MIT. Her research interests include game theory, optimization theory, with emphasis on nonlinear programming and convex analysis, and communication networks.