Architecting Noncooperative Networks

Yannis A. Korilis, Aurel A. Lazar and Ariel Orda
Department of Electrical Engineering
and
Center for Telecommunications Research
Columbia University, New York, NY 10027-6699

ABSTRACT

In noncooperative networks users make control decisions that optimize their own performance measure. Focusing on routing, we devise two methodologies for architecting noncooperative networks, that improve the overall network performance. These methodologies are motivated by problem settings arising in the provisioning and the run time phases of the network. For either phase, Nash equilibria characterize the operating point of the network.

The goal of the provisioning phase is to allocate link capacities that lead to systemwide efficient Nash equilibria. In general, the solution of such design problems is counterintuitive, since adding link capacity might lead to a degradation of user performance. We show that, for systems of parallel links, such paradoxes cannot occur and the optimal solution coincides with the solution in the single-user case. We derive some extensions to general network topologies.

During the run time phase, a manager controls the routing of part of the network flow. The manager is aware of the noncooperative behavior of the users and makes its routing decisions based on this information while aiming at improving the overall system performance. We obtain necessary and sufficient conditions for enforcing an equilibrium that coincides with the global systemwide optimum, and indicate that these conditions are met in many cases of interest.

CTR Technical Report: CU/CTR/TR 385-94-32

October 5, 1994.

Architecting Noncooperative Networks

Yannis A. Korilis, Aurel A. Lazar and Ariel Orda
Department of Electrical Engineering
and
Center for Telecommunications Research
Columbia University, New York, NY 10027-6699

1. Introduction

Control decisions in large scale networks are often made by each user independently, according to its own individual performance objectives. Such networks are henceforth called *noncooperative*, and game theory [MYE91, FUD92] provides the systematic framework to study and understand their behavior. Game theoretic models have been employed in the context of flow control [HSIA91, MAZ91, ZHA92, ALT93, ORD93, KOR93, ALT94], routing [ECO91, ORD93], and pricing [COC93] in modern networking. These studies mainly investigate the structure of the network operating points, i.e., the *Nash equilibria* of the respective games. Such equilibria are inherently inefficient [DUB86] and, in general, exhibit suboptimal network performance.

The goal of this paper is to demonstrate that, while users make noncooperative decisions, there is still room for improving network performance. Improvements can be achieved both during the provisioning phase, i.e., when the network parameters are sized, and during the run time phase, i.e., during the operation of the network. Focusing on routing, we give a uniform methodology for achieving such improvements. This methodology is based on architecting the network equilibria. The related analysis involves comparisons among operating points of different games. Such comparisons are scarcely attempted in the game theoretic literature, mainly due to the complex structure – or lack thereof – of the underlying game. One exception is [SHE94] which addresses the problem of designing the service discipline of a switch shared by users performing flow control.

In the *provisioning* phase, the designer allocates link capacities, i.e., architects the *capacity* configuration of the network, so that the resulting equilibrium is systemwide "efficient" or "optimal." We consider several efficiency criteria for the designer, such as the "price" (marginal cost) as seen by each user, the total cost of each user, or some combination of the above. The designer has to decide how much capacity should be allocated to each link, while satisfying lower bounds specified per link and an upper bound on the total capacity. The designer seeks an allocation of capacities that achieves the best performance, according to the chosen efficiency criterion. The immediate question that arises is whether the designer should attempt to employ all the available resources. Surprisingly, in general, the answer is no! To illustrate this counterintuitive behavior of noncooperative networks, we adapt the Braess paradox [ZAN81, COH90] to our setting and show that addition of resources may result in a *degradation* of user performance.

Example: Consider the network depicted in Figure 1. Links $(1,2)$ and $(3,4)$ have each capacity c_1. Link $(1,3)$ represents a path of n tandem links, each with capacity c_2. Similarly, links $(2,4)$ and $(2,3)$ are paths of n consecutive links each with capacities c_2 and c_3, respectively. There are I users, each with an average throughput demand r, sending flow from node 1 to node 4. Each user aims to route its demand r over the available paths, so as to minimize its total cost defined as the sum of its delays over all links. The delay per unit of flow on each link is given by the M/M/1 delay formula. Prices (marginal costs) represent derivatives of the cost with respect to user flows. In [ORD93] it has been shown that for this system there exists a unique and symmetrical Nash equilibrium, i.e., the flows (and thus, the costs and prices) of the users at equilibrium are equal. Figures 2 and 3 show, correspondingly, the user price and cost as functions of c_3 (for $c_1 = 2.7$, $c_2 = 27$, $n = 54$, $I = 10$ and $r = 0.2$). The figures indicate that, for any $c_3 > 0$, both the price and the cost of each user are higher than for $c_3 = 0$, i.e., *eliminating* the path $(2,3)$ leads to an *improvement* of performance for all users. More surprisingly, this paradoxical behavior persists even if $c_3 = \infty$.

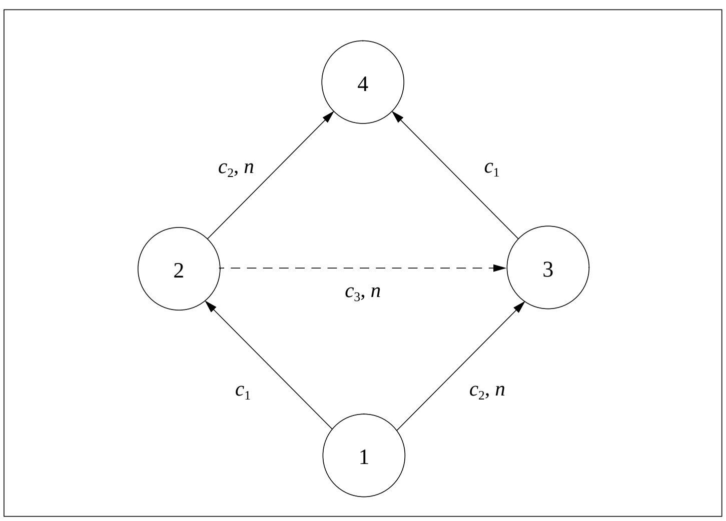

Figure 1: Network Paradox

For a system of parallel links we show that the Braess paradox cannot occur, that is,

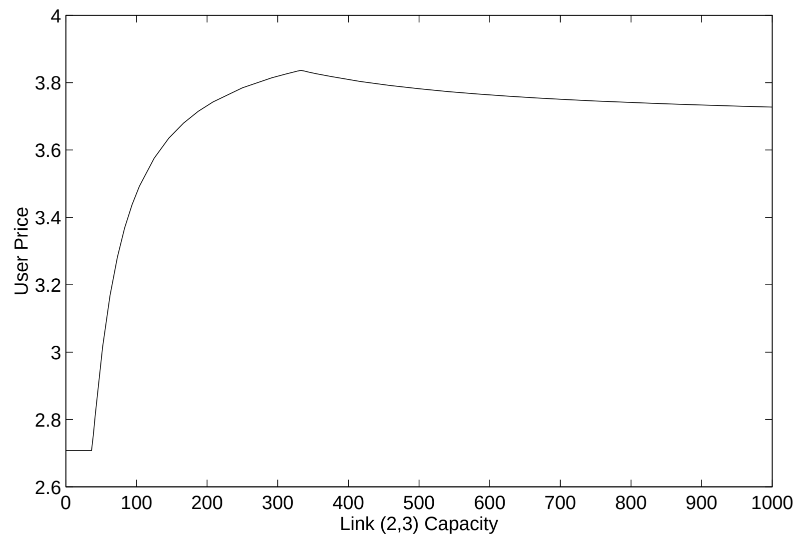

Figure 2: User price as function of the link capacity c_3

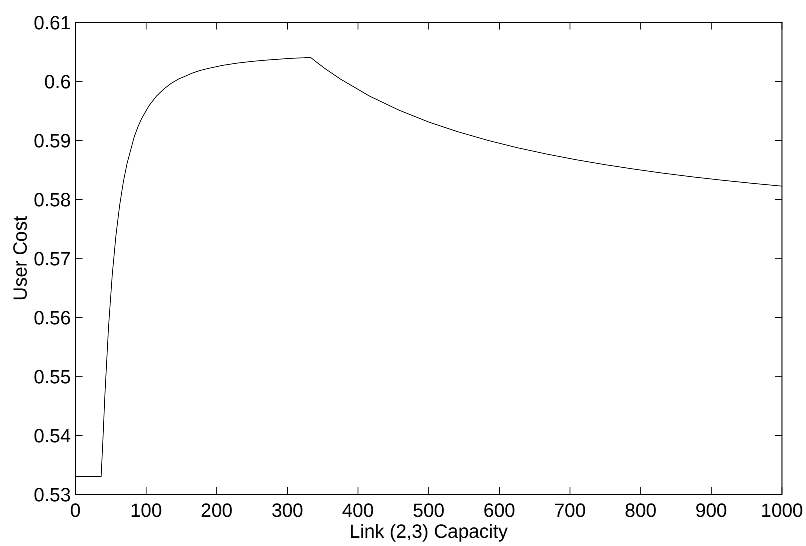

Figure 3: User cost as function of the link capacity c_3

addition of capacity improves the network performance. We then consider the problem of allocating such additional capacity to links in an optimal way. We show that the best design strategy is to allot the additional capacity exclusively to the link with the originally highest capacity. This solution coincides with the optimal capacity allocation in a network where routing is centrally controlled. We extend some of these results to general network topologies.

In the *run time* phase, we assume that, apart from the noncooperative users, there is also a manager, that attempts to optimize the system performance, by deciding upon the routing of an additional, network-controlled flow. The manager is aware of the noncooperative behavior of the users, and thus it can predict their reaction to any routing strategy that it chooses. This information enables the manager to implement a routing strategy that drives the users to the "best" Nash equilibrium in terms of system performance, architecting, this way, the *flow* configuration of the network. This is the typical scenario of a Stackelberg game [MYE91], in which the manager acts as a leader and imposes its strategy on the users which behave as followers. Stackelberg strategies have been investigated in the context of flow control in [DOU89]. In that reference, however, the leader was a selfish user concerned about its own, rather than the system's, performance.

For the parallel links model, we derive necessary and sufficient conditions that guarantee that the manager can enforce an equilibrium that coincides with the global systemwide optimum.[1] Moreover, we indicate that these conditions are met in many practical cases. In other words, the manager is often able to obtain, through limited control, the same system performance as in the case of centralized control.

The outline of the paper is the following. In Section 2, we present the noncooperative parallel links model, and explain that it is well-suited for modeling typical configurations in modern networking. The design issues arising in the provisioning phase are investigated in Section 3, which is organized as follows. After formulating the problem in Subsection 3.1, we outline the main results in Subsection 3.2. The rest of the section consists of a detailed description of the solution. In Subsection 3.3, we explore the structure of the underlying Nash equilibria. In Subsection 3.4, we establish that addition of capacity to a network of parallel links cannot degrade performance. With this result at hand, we investigate, in Subsection 3.5, the optimal strategy for adding capacity to networks of parallel links. In Subsection 3.6, we extend some of these results to general topologies. The design issues arising in the run time phase are considered in Section 4. Finally, Section 5 summarizes the main results and delineates their practical implications. Due to size constraints, most of the

[1] The global systemwide optimum is the optimal solution of the routing problem when all the flow in the network is centrally controlled.

formal proofs are omitted; for these proofs the reader is referred to [KOR94, KOR94b].

2. The Model

We consider a set $\mathcal{I} = \{1, \ldots, I\}$ of users, that share a set $\mathcal{L} = \{1, \ldots, L\}$ of communication links, interconnecting a common source to a common destination node (Figure 4). Let c_l be the capacity of link l, and $C = \sum_{l \in \mathcal{L}} c_l$ be the total capacity of the system of parallel links. Each user i has a throughput demand that is some ergodic process with average rate $r^i > 0$. We assume that $r^1 \geq r^2 \geq \ldots \geq r^I$. Let $R = \sum_{i \in \mathcal{I}} r^i$ denote the total throughput demand of the users. Throughout this paper, we consider only capacity configurations $\mathbf{c} = (c_1, \ldots, c_L)$ that can accommodate the total user demand, i.e., configurations with $C > R$.

User i ships its flow by splitting its demand r^i over the set of parallel links, according to some individual performance objective. Let f_l^i denote the expected flow that user i sends on link l. The user flow configuration $\mathbf{f}^i = (f_1^i, \ldots, f_L^i)$ is called a routing *strategy* of user i and the set $F^i = \{\mathbf{f}^i \in I\!R^L : f_l^i \geq 0,\ l \in \mathcal{L};\ \sum_{l \in \mathcal{L}} = r^i\}$ of strategies that satisfy the user's demand is called the strategy space of user i. The system flow configuration $\mathbf{f} = (\mathbf{f}^1, \ldots, \mathbf{f}^I)$ is called a routing *strategy profile* and takes values in the product strategy space $F = \otimes_{i \in \mathcal{I}} F^i$.

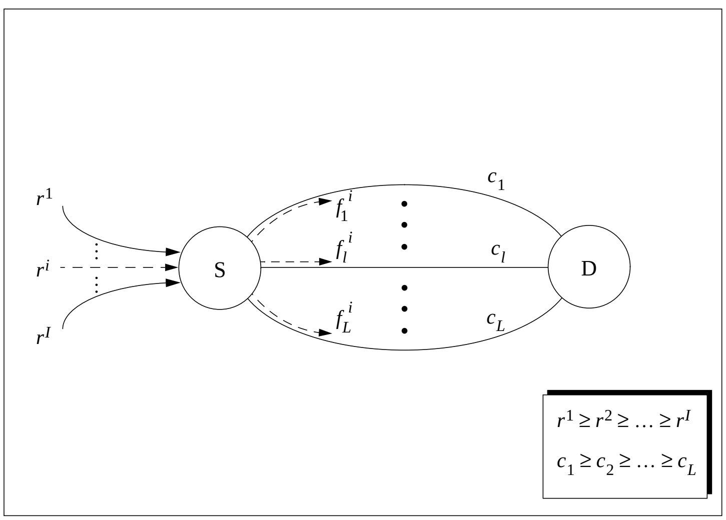

Figure 4: The system of parallel links

The performance objective of user i is quantified by means of a cost function $J^i(\mathbf{f})$. The user aims to find a strategy $\mathbf{f}^i \in F^i$ that minimizes its cost. This optimization problem depends on the routing decisions of the other users, described by the strategy profile $\mathbf{f}^{-i} = (\mathbf{f}^1, \ldots, \mathbf{f}^{i-1}, \mathbf{f}^{i+1}, \ldots, \mathbf{f}^I)$, since J^i is a function of the system flow configuration $\mathbf{f}$. A *Nash equilibrium* of the routing game is a strategy profile from which no user finds it beneficial to unilaterally deviate. Hence, $\mathbf{f} \in F$ is a Nash equilibrium if:

$$\mathbf{f}^i \in \arg\min_{\mathbf{g}^i \in F^i} J^i(\mathbf{g}^i, \mathbf{f}^{-i}), \quad i \in \mathcal{I}. \tag{2.1}$$

The problem of existence and uniqueness of equilibria of the routing game has been investigated in [ORD93] for certain general classes of cost functions. In the present paper, we consider cost functions that are the sum of link cost functions:

$$J^i(\mathbf{f}) = \sum_{l \in \mathcal{L}} J_l^i(\mathbf{f}_l), \quad J_l^i(\mathbf{f}_l) = f_l^i T_l(f_l), \quad l \in \mathcal{L},\ i \in \mathcal{I}, \tag{2.2}$$

where $\mathbf{f}_l = (f_l^1, \ldots, f_l^I)$, and $T_l(f_l)$ is the average delay per unit of flow on link l and depends only on the total flow $f_l = \sum_{i \in \mathcal{I}} f_l^i$ on that link. In particular, we concentrate on the M/M/1 delay function:

$$T_l(f_l) = \begin{cases} \dfrac{1}{c_l - f_l}, & f_l < c_l \\ \infty, & f_l \geq c_l \end{cases} . \tag{2.3}$$

Note that the stability constraint $f_l < c_l$ of link l is manifested through the definition of T_l. In particular, since the total user demand R does not exceed the total capacity C of the network, eqs. (2.1) and (2.3) guarantee that at any Nash equilibrium $f_l < c_l$, for all $l \in \mathcal{L}$ and the costs of all users are finite.

Given a strategy profile $\mathbf{f}^{-i}$ of the other users, the cost of user i, as defined by eqs. (2.2) and (2.3), is a convex function of its strategy $\mathbf{f}^i$. Hence, the minimization problem in (2.1) has a unique solution. The Kuhn-Tucker optimality conditions [LUE84], then, imply that $\mathbf{f}^i$ is the optimal response of user i to $\mathbf{f}^{-i}$ if and only if there exists a (Lagrange multiplier) λ^i, such that:

$$\lambda^i = \frac{\partial J^i}{\partial f_l^i}(\mathbf{f}), \text{ if } f_l^i > 0, \quad l \in \mathcal{L} \tag{2.4}$$

$$\lambda^i \leq \frac{\partial J^i}{\partial f_l^i}(\mathbf{f}), \text{ if } f_l^i = 0, \quad l \in \mathcal{L} \tag{2.5}$$

$$\textstyle\sum_{l \in \mathcal{L}} f_l^i = r^i, \quad f_l^i \geq 0, \tag{2.6}$$

i.e., a strategy profile $\mathbf{f} \in F$ is a Nash equilibrium, if and only if there exist λ^i, $l \in \mathcal{L}$, such that the optimality conditions (2.4)–(2.6) are satisfied for all $i \in \mathcal{I}$. The above conditions imply that the Lagrange multiplier λ^i is in fact the marginal cost of user i at the optimality point. In accordance with the economics terminology, λ^i will be referred to as the *price* of user i.

For the cost function $J^i(\mathbf{f})$ given by eqs. (2.2) and (2.3), we have:

$$\frac{\partial J^i}{\partial f_l^i}(\mathbf{f}) = f_l^i T_l'(f_l) + T_l(f_l) = \frac{c_l - f_l^{-i}}{(c_l - f_l)^2}, \tag{2.7}$$

where T_l' is the derivative of T_l with respect to f_l, and $f_l^{-i} \equiv \sum_{j \neq i} f_l^j$ is the total flow that all users except the i-th send on link l.

In [ORD93] it has been shown that the routing game described above has a *unique* Nash equilibrium.

2.1 Special Cases

At times we will concentrate on special types of users, defined in the following.

Definition 2.1 *Identical users: we say that users are identical if their demands are all equal, that is, $r^i = r^j$ for all $i, j \in \mathcal{I}$.*

The Nash equilibrium of identical users is symmetrical, i.e., $f_l^i = f_l^j = f_l/I$ for all $i, j \in \mathcal{I}$ [ORD93].

Definition 2.2 *Simple users: we say that a user is simple if it routes all of its flow through links (or paths) of minimal delay.*

In general, routing of simple users does not satisfy the optimality conditions (2.4)–(2.6). However, many times users route according to the "simple" scheme due to practical reasons. In fact, many typical routing algorithms send flow through shortest paths, without getting into the trouble of accounting for derivatives (T_l') and bifurcating flow. The Nash equilibrium of simple users in a system of parallel links is unique with respect to the *total* link flows [ORD93], and the corresponding necessary and sufficient conditions are, the existence of some λ, such that:

$$\lambda = T_l, \text{ if } f_l > 0, \quad l \in \mathcal{L} \tag{2.8}$$

$$\lambda \leq T_l, \text{ if } f_l = 0, \quad l \in \mathcal{L} \tag{2.9}$$

$$\textstyle\sum_{l \in \mathcal{L}} f_l = R, \quad f_l \geq 0. \tag{2.10}$$

We shall refer to the value of λ as the price of the simple users. From (2.8)–(2.10), it is easy to see that users that route according to the optimality conditions (2.4)–(2.6) become simple as their population grows to infinity and their individual demands become infinitesimally small, while their total demand remains R. This is the typical scenario in a transportation network.

Definition 2.3 *Consistent users: we say that users are consistent (for a given capacity configuration) if, at the Nash equilibrium, they all use the same set of links.*

Due to the structure of their Nash equilibrium, identical users are consistent. It is easy to verify that simple users are also consistent. Finally, consistent users are typical of systems with heavy traffic, i.e., when R gets close to C, in which case all users will use all links in the network.

2.2 Validity of the Model

We note that systems of parallel links represent an appropriate model for seemingly unrelated networking problems. Consider, for example, a network in which resources are pre-allocated to various routing paths that do not interfere. Such scenaria are common in modern networking. In broadband networks bandwidth is separated among different virtual paths, resulting effectively in a system of parallel and non-interfering "links" between source/destination pairs. Another example is that of internetworking, in which each "link" models a different sub-network.

3. Architecting the Capacity Configuration in the Provisioning Phase

3.1 Problem Formulation

Consider a network of parallel links with initial capacity configuration $\mathbf{c}^0$ and total capacity $C^0 > R$. We assume, without loss of generality, that $c_1^0 \geq \ldots \geq c_L^0 > 0$. Suppose that there exists some additional capacity allowance of at most Δ, which the network designer can distribute among the network links. The aim of the designer is to implement a capacity configuration $\mathbf{c}$, with $c_l \geq c_l^0$ for all links $l \in \mathcal{L}$, that results in a network, with a total capacity of at most $C^0 + \Delta$, that is "efficient" at the corresponding Nash equilibrium. Without loss of generality, we can concentrate on capacity configurations $\mathbf{c}$ that preserve the initial link order, i.e., configurations with $c_1 \geq \ldots \geq c_L$.[2] Therefore, the set of all capacity configurations that can be implemented by the designer can be described by:

$$\mathcal{C}_\Delta = \{\mathbf{c} \in I\!R_+^L : c_1 \geq \ldots \geq c_L;\ c_l \geq c_l^0,\ l \in \mathcal{L};\ \textstyle\sum_{l\in\mathcal{L}}(c_l - c_l^0) \leq \Delta\}.$$

[2]The properties of the Nash equilibrium in a system of parallel links with capacity configuration $\mathbf{c}$ depends on the actual link capacities and not on the link "labels," that are determined by the initial configuration $\mathbf{c}^0$. Hence, renaming the links, so that $c_1 \geq \ldots \geq c_L$, does not affect the characteristics of the resulting equilibrium.

Each capacity configuration in $\mathcal{C}_\Delta$ induces a routing game that has a unique Nash equilibrium. Therefore, we can define a function $\mathcal{N} : \mathcal{C}_\Delta \to I\!R^{IL}$, that assigns to each $\mathbf{c} \in \mathcal{C}_\Delta$ the Nash equilibrium $\mathcal{N}(\mathbf{c})$ of its respective routing game. $\mathcal{N}$ will be referred to as the *Nash mapping*. The set $\mathcal{C}_\Delta$ will be called the *space of routing games*.

The designer may have different measures to characterize the efficiency of a capacity configuration. In this work, we shall concentrate on measures that are expressed by means of either the user prices or costs. We mention that, although the user's cost is a direct measure of its level of satisfaction, prices may be a more important measure from a system's point of view, since they account for the level of congestion as seen by users and are the direct indication of how each user could accommodate fluctuations in the system's state. The designer can consider various ways of combining either the prices or the costs of the users. We shall consider the following two: userwise optimization, i.e., trying to reduce the price or cost of each and every user, and total optimization, i.e., trying to reduce the sum of all prices or costs.

The various performance measures of the designer are formally stated in the following definitions:

Definition 3.1 *Userwise/total efficiency: Consider two capacity configurations* $\mathbf{c}$ *and* $\hat{\mathbf{c}}$ *and let* λ^i *and* $\hat{\lambda}^i$ *(correspondingly,* J^i *and* $\hat{J}^i$*) be the price (correspondingly, cost) of user* i *at the respective equilibria. Then:*

1. *Configuration* $\hat{\mathbf{c}}$ *is said to be userwise price (correspondingly, cost) efficient relative to configuration* $\mathbf{c}$*, if, for all* $i \in \mathcal{I}$*, it holds that* $\hat{\lambda}^i \leq \lambda^i$ *(correspondingly,* $\hat{J}^i \leq J^i$*).*

2. *Configuration* $\hat{\mathbf{c}}$ *is said to be totally price (correspondingly, cost) efficient relative to configuration* $\mathbf{c}$*, if it holds that* $\sum_{i\in\mathcal{I}} \hat{\lambda}^i \leq \sum_{i\in\mathcal{I}} \lambda^i$ *(correspondingly,* $\sum_{i\in\mathcal{I}} \hat{J}^i \leq \sum_{i\in\mathcal{I}} J^i$*).*

Definition 3.2 *Userwise/total optimality: Given a set of capacity configurations* $\mathcal{C}$*, a capacity configuration* $\mathbf{c}^* \in \mathcal{C}$ *is said to be:*

1. *userwise price (correspondingly, cost) optimal, if it is userwise price (correspondingly, cost) efficient relative to any* $\mathbf{c} \in \mathcal{C}$*,*

2. *totally price (correspondingly, cost) optimal, if it is totally price (correspondingly, cost) efficient relative to any* $\mathbf{c} \in \mathcal{C}$*.*

Obviously, userwise efficiency (optimality) implies total efficiency (optimality). However, price and cost efficiency (optimality) do not imply each other in either direction. Note also that, in general, existence of userwise optima cannot be guaranteed even if total optima do exist.

The optimal capacity allocation problem, corresponding to the various designer's performance measures, is described as follows:

> Given a system of parallel links $\mathcal{L}$ with users $\mathcal{I}$, an (initial) capacity configuration $\mathbf{c}^0$ and an additional capacity allowance Δ, find a capacity configuration $\mathbf{c}^*$ that is userwise/totally price/cost optimal with respect to $\mathcal{C}_\Delta$.

By definition , the initial capacity c_l^0 of every link l is positive, in other words, the designer can only add capacity to existing links. Nonetheless, as shown in [KOR94], the results of the following subsections can be easily extended to the case where $c_l^0 = 0$ for some links $l \in \mathcal{L}$, i.e., when the designer is also allowed to add a (finite) number of links to the network. Although the problem is formulated as allocating additional capacity to an existing network, this formulation is equivalent to the typical capacity allocation problem, where the capacity of each link has to be higher than a lower bound.

Solving the optimal capacity allocation problem in a network shared by noncooperative users amounts to comparing the Nash equilibria of the routing games induced by different capacity configurations in $\mathcal{C}_\Delta$. Comparing the outcomes of different games is, in general, a highly complex task and is feasible only if an explicit characterization of the respective equilibria is available. The structure of the unique Nash equilibrium of the routing game is investigated in subsection 3.3. Before we proceed, let us first summarize the main results of this section.

3.2 Outline of Results

Following is an informal summary of the main results on the design problem. Unless otherwise stated, the results apply to the model formulated in Section 2.

1. Addition of (any amount of) capacity to any link decreases the price of all users, i.e., it results in a configuration that is userwise (thus, totally) price efficient.

2. Addition of capacity to the link with the initially highest capacity (i.e., to link 1) results in a totally cost efficient configuration.

3. For consistent users (thus, in particular, for identical or simple users), addition of capacity to any link results with a userwise cost efficient configuration.

4. The capacity configuration that results from allocating the entire additional capacity allowance Δ to the link with the initially highest capacity (i.e., to link 1) is userwise (thus, totally) price optimal in $\mathcal{C}_\Delta$.

5. Cost optimality of the above configuration is shown when users are consistent (thus, in particular, when they are identical or simple), and also in the case of two users.

6. Considering general topologies (i.e., not necessarily systems of parallel links), we obtain methods for adding capacity to links so the Braess paradox does not occur.

3.3 Structure of the Nash Equilibrium

In this subsection we study the structure of the Nash equilibrium of the routing game in a network of parallel links with capacity configuration $\mathbf{c}$. In [ORD93] it has been shown that the Nash equilibrium of the routing game exhibits a number of intuitive monotonicity properties that are summarized in the following:

Lemma 3.3 *Let* $\mathbf{f}$ *be the unique Nash equilibrium of the routing game in a network of parallel links with capacity configuration* $\mathbf{c}$*. Then:*

1. *For every user* $i \in \mathcal{I}$*, we have* $f_1^i \geq f_2^i \geq \ldots \geq f_L^i$*. In particular, for* $f_l^i > 0$*, we have* $f_l^i = f_m^i$ *if and only if* $c_l = c_m$*.*

2. *For every link* $l \in \mathcal{L}$*, we have* $f_l^1 \geq f_l^2 \geq \ldots \geq f_l^I$*. In particular, for* $f_l^i > 0$*, we have* $f_l^i = f_l^j$ *if and only if* $r^i = r^j$*.*

3. $c_1 - f_1 \geq c_2 - f_2 \geq \ldots \geq c_L - f_L$*, or equivalently,* $T_1 \leq T_2 \leq \ldots \leq T_L$*. In particular,* $T_l = T_m$*, if and only if* $c_l = c_m$*.*

4. *For every user* $i \in \mathcal{I}$*, define* $c_l^i = c_l - f_l^{-i}$*,* $l \in \mathcal{L}$*. Then,* $c_1^i \geq c_2^i \geq \ldots \geq c_L^i$*. In particular,* $c_l^i = c_m^i$*, if and only if* $c_l = c_m$*.*

Note that $c_l^i = c_i - f_l^{-i}$ is the residual capacity of link l as seen by user i, given that the strategy profile of the other users is $\mathbf{f}^{-i}$.

Let $\mathcal{L}^i$ denote the set of links that receive some flow from user i, and $\mathcal{I}_l$ denote the set of users that send flow over link l. The first statement in Lemma 3.3 implies that for every user i, there exists some link L^i, such that $f_l^i > 0$ for all $l \leq L^i$, and $f_l^i = 0$ for $l > L^i$, that is, $\mathcal{L}^i = \{1, 2, \ldots, L^i\}$. Similarly, the second statement in the Lemma implies that for every link l, there exists some user I_l, such that $f_l^i > 0$ for all $i \leq I_l$, and $f_l^i = 0$ for $i > I_l$, that is, $\mathcal{I}_l = \{1, 2, \ldots, I_l\}$.

Consider now the best reply $\mathbf{f}^i$ of user i to a fixed strategy profile $\mathbf{f}^{-i}$ of the other users. This is the unique solution to the optimal routing problem in a network of parallel links with capacities c_l^i, $l \in \mathcal{L}$, and is determined by the Kuhn-Tucker optimality conditions (2.4)–(2.6). Note that conditions (2.4) and (2.5) can be written as:

$$\lambda^i = \frac{c_l^i}{(c_l^i - f_l^i)^2}, \quad \text{if } f_l^i > 0, \tag{3.1}$$

$$\lambda^i \leq \frac{c_l^i}{(c_l^i - f_l^i)^2}, \quad \text{if } f_l^i = 0, \tag{3.2}$$

for any link $l \in \mathcal{L}$.

In the sequel, we will give an explicit characterization of the structure of the user's equilibrium strategy $\mathbf{f}^i$, as a function of $\mathbf{c}^i = (c_1^i, \dots, c_L^i)$, which depends on the capacity configuration $\mathbf{c}$ and the strategy profile $\mathbf{f}^{-i}$ of the other users. To this end, let us define:

$$G_l^i = \sum_{m=1}^{l-1} c_m^i - \sqrt{c_l^i} \sum_{m=1}^{l-1} \sqrt{c_m^i}, \quad l = 2, \dots, L, \quad i \in \mathcal{I}, \tag{3.3}$$

$$G_1^i = 0, \quad G_{L+1}^i = \sum_{m=1}^{L} c_m^i = C - R^{-i}, \quad i \in \mathcal{I},$$

where $R^{-i} \equiv \sum_{j \neq i} r^j$ is the total demand of all users except the i-th. Note that $C - R^{-i}$ is the total (residual) capacity of the network as seen by user i. Then it is easy to verify that $c_l^i \geq c_{l+1}^i$ (see Lemma 3.3) implies:

$$G_l^i \leq G_{l+1}^i, \quad l = 2, \dots, L-1, \tag{3.4}$$

with equality holding if and only if $c_l = c_{l+1}$. We are now ready to state the following:

Proposition 3.4 *The Nash equilibrium* $\mathbf{f}$ *of the routing game in a system of parallel links with capacity configuration* $\mathbf{c}$ *is described by:*

$$f_l^i = \begin{cases} c_l^i - (\sum_{m=1}^{L^i} c_m^i - r^i) \frac{\sqrt{c_l^i}}{\sum_{m=1}^{L^i} \sqrt{c_m^i}}, & 1 \leq l \leq L^i \\ 0, & L^i < l \leq L \end{cases}, \quad i \in \mathcal{I}, \tag{3.5}$$

where, for every user $i \in \mathcal{I}$, *the threshold* L^i *is determined by:*

$$G_{L^i}^i < r^i \leq G_{L^i+1}^i. \tag{3.6}$$

The equilibrium price and the equilibrium cost for user i are respectively:

$$\lambda^i = \left\{ \frac{\sum_{l=1}^{L^i} \sqrt{c_l^i}}{\sum_{l=1}^{L^i} c_l^i - r^i} \right\}^2, \quad J^i = \lambda^i \sum_{l=1}^{L^i} (c_l - f_l) - L^i = \frac{\left\{ \sum_{l=1}^{L^i} \sqrt{c_l^i} \right\}^2}{\sum_{l=1}^{L^i} c_l^i - r^i} - L^i. \tag{3.7}$$

□

From Proposition 3.4, and especially the expressions for the equilibrium prices and costs of the users, it is clear that the set of links over which each user sends its flow has a prominent role in the properties of the Nash equilibrium of the routing game. In studying the capacity allocation problem, we need to compare the equilibria of games that are induced by different capacity configurations in $\mathcal{C}_\Delta$. If the resulting equilibria are such that the sets of links over which each user sends its flow do not coincide at both equilibria, such comparisons are extremely complex, if possible at all. In [KOR94], we exploit the structure of the Nash equilibrium of the routing game, given by Proposition 3.4, to show that the Nash mapping $\mathcal{N}$, that assigns to each capacity configuration in $\mathcal{C}_\Delta$ the unique equilibrium of the induced routing game, is continuous. In the same reference, we show that this fundamental property allows us to investigate the optimal capacity allocation problem based on comparisons of capacity configurations whose "distance" in $\mathcal{C}_\Delta$ is sufficiently small, so that each user sends its flow over the same links under both configurations. The related results from [KOR94] are summarized in the Appendix.

3.4 Efficiency of Capacity Addition

In Section 1 we demonstrated by way of examples that, in general, addition of capacity to a network may increase both the user prices and costs. In this subsection we investigate the addition of capacity to systems of parallel links and show that, under various conditions, the paradoxical behavior of general topology networks cannot occur in this setting.

We say that a capacity configuration $\hat{\mathbf{c}}$ is an *augmentation* of a capacity configuration $\mathbf{c}$ if $\hat{c}_l \geq c_l$ for all l and $\sum_l \hat{c}_l > \sum_l c_l$. Throughout this subsection we shall compare the Nash equilibrium of a capacity configuration $\mathbf{c}$ to that of some augmentation $\hat{\mathbf{c}}$. The "hat" values will refer to configuration $\hat{\mathbf{c}}$, while the "non-hat" values to $\mathbf{c}$. For example, $\hat{\lambda}^i$ and λ^i are the prices of user i under $\hat{\mathbf{c}}$ and $\mathbf{c}$, respectively.

The first lemma shows that addition of capacity is always efficient as with respect to prices.

Lemma 3.5 *If a capacity configuration $\hat{\mathbf{c}}$ is an augmentation of a capacity configuration $\mathbf{c}$ then $\hat{\mathbf{c}}$ is userwise price efficient relative to $\mathbf{c}$, that is, $\hat{\lambda}^i \leq \lambda^i$, for all $i \in \mathcal{I}$. Moreover, the* *equilibrium delay of each link l is lower (not higher) under configuration $\hat{\mathbf{c}}$, i.e., $\hat{T}_l \leq T_l$, for all $l \in \mathcal{L}$.*

Proof: See the Appendix.

□

Due to the results presented in the next subsection, we are particularly interested in augmentations in which capacity is added to link 1 solely. The second lemma shows that, for such augmentations, the resulting configuration is totally cost efficient.

Lemma 3.6 *Let $\mathbf{c}$ and $\hat{\mathbf{c}}$ be two capacity configurations such that $\hat{c}_l = c_l$ for all $l > 1$ and $\hat{c}_1 > c_1$. Then $\hat{\mathbf{c}}$ is totally cost efficient relative to $\mathbf{c}$.*

Proof: See the Appendix.

□

The following two lemmata establish userwise cost efficiency of capacity addition in some cases of special interest.

Lemma 3.7 *Let $\mathbf{c}$ and $\hat{\mathbf{c}}$ be two capacity configurations such that $\hat{\mathbf{c}}$ is an augmentation of $\mathbf{c}$. Assume that users are consistent with respect to both $\hat{\mathbf{c}}$ and $\mathbf{c}$. Then $\hat{\mathbf{c}}$ is userwise cost efficient relative to $\mathbf{c}$, that is, $\hat{J}^i \leq J^i$, for all $i \in \mathcal{I}$.*

□

The above result applies, in particular, both to identical users and to simple users, since they belong to the class of consistent users, under all capacity configurations. The next lemma establishes userwise cost efficiency in the case of two users ($I = 2$) and for adding capacity exclusively to link 1.

Lemma 3.8 *Let $\mathbf{c}$ and $\hat{\mathbf{c}}$ be two capacity configurations such that $\hat{c}_l = c_l$ for all $l > 1$ and $\hat{c}_1 > c_1$. Then, for $I = 2$, $\hat{\mathbf{c}}$ is userwise cost efficient relative to $\mathbf{c}$.*

□

3.5 Optimal Capacity Allocation

Let us now proceed to investigate the optimal capacity allocation problem in a system of parallel links, according to the various performance measures defined in Section 3.1. The main result of this section, namely Theorem 3.14, asserts that the capacity configuration $\mathbf{c}^* = \mathbf{c}^0 + \Delta \mathbf{e}_1$,[3] that results from allocating the entire additional capacity to the link with the initially highest capacity is userwise price optimal in $\mathcal{C}_\Delta$. Furthermore, $\mathbf{c}^*$ will be shown to be userwise cost optimal for a number of special cases of interest.

While this is a simple and intuitive result, its proof requires systematic analysis that establishes some "order" in the complex structure of the underlying game. Although most of the formal proofs are excluded from the main text, the lemmata presented in this section delineate the methodology through which that task has been achieved. We note that, while most of these results have a clear, intuitive appeal and a simple proof in the typical single-user case, the proofs for the multi-user (game) case, considered here, demand a cautious and rather tedious analysis.

We start by considering two capacity configurations $\mathbf{c}$ and $\hat{\mathbf{c}}$, such that $\hat{\mathbf{c}}$ is derived from $\mathbf{c}$ by a "transfer" of capacity from some link q with $c_q < c_1$ to link 1, and show that $\hat{\mathbf{c}}$ is userwise price efficient with respect to $\mathbf{c}$. Hence, if an additional capacity of exactly Δ is to be allocated to the system, then $\mathbf{c}^*$ is the optimal capacity configuration. By virtue of the price efficiency of capacity addition (Lemma 3.5), $\mathbf{c}^*$ is also userwise price optimal in the entire space of games $\mathcal{C}_\Delta$, that allows for addition of capacity not necessarily equal, but also less than Δ.

Consider a system of parallel links with initial capacity configuration $\mathbf{c} \in \mathcal{C}_\Delta$. Let $\hat{\mathbf{c}} = \mathbf{c} + \Delta_q(\mathbf{e}_1 - \mathbf{e}_q)$ be the configuration that results by transferring capacity $\Delta_q > 0$ from some link $q > 1$, with $c_q^0 < c_q < c_1$, to link 1. As before, the "hat" values will refer to configuration $\hat{\mathbf{c}}$, while the "non-hat" values to the initial configuration $\mathbf{c}$. Recall that $\mathcal{L}^i = \{1, \ldots, L^i\}$ is the set of links that receive flow from user i, while $\mathcal{I}_l = \{1, \ldots, I_l\}$ is the set of users that send flow to link l, both under configuration $\mathbf{c}$. By virtue of the continuity properties presented in the Appendix (Lemma A.2), we choose Δ_q sufficiently small, so that this transfer of capacity does not force any user to change the set of links over which it sends its flow, i.e., $\hat{\mathcal{L}}^i = \mathcal{L}^i$, for all $i \in \mathcal{I}$. Then, $\hat{\mathcal{I}}_l = \mathcal{I}_l$, for all links $l \in \mathcal{L}$.

The comparison of capacity configurations $\hat{\mathbf{c}}$ and $\mathbf{c}$ is carried out in the sequel, under the assumption $\hat{\mathcal{L}}^i = \mathcal{L}^i$, $i \in \mathcal{I}$, in a series of four lemmata. Lemma 3.9, examines the effect of the transfer of capacity from link q to link 1 on the equilibrium delays of these two links.

[3] $\mathbf{e}_l$ is the vector in $\mathbb{R}^L$ with the l-th component equal to 1 and all other components equal to 0.

Lemmata 3.10 and 3.11 show that the transfer of capacity affects the prices of the users and the equilibrium delays of the links in $\mathcal{L}\setminus\{1,q\}$ in an "ordered" way – in a sense that will be explained in the lemmata – and play a key role in the proof of the main result of this section. Finally, userwise price efficiency of $\hat{\mathbf{c}}$ with respect to $\mathbf{c}$ will be established in Lemma 3.12. These results will be extended to the case of an arbitrary capacity transfer Δ_q, that might lead to $\hat{\mathcal{L}}^i \neq \mathcal{L}^i$ for some user i, in Theorem 3.13.

The following lemma shows that, under capacity configuration $\hat{\mathbf{c}}$, the delay on link 1 is lower, while the delay on link q is higher.

Lemma 3.9 *Consider two capacity configurations* $\mathbf{c}, \hat{\mathbf{c}} \in \mathcal{C}_\Delta$ *with* $\hat{\mathbf{c}} = \mathbf{c} + \Delta_q(\mathbf{e}_1 - \mathbf{e}_q)$. *Then,* $\hat{T}_1 \leq T_1$ *and* $\hat{T}_q > T_q$.

Proof: See the Appendix.

□

The next lemma asserts that the transfer of capacity from link q to link 1 affects the prices of all users that send flow to link q in the same way, that is, either $\hat{\lambda}^i > \lambda^i$ for all $i \in \mathcal{I}_q$, or $\hat{\lambda}^i \leq \lambda^i$ for all $i \in \mathcal{I}_q$. Similarly, either all links with capacity lower than link q increase their equilibrium delays with $\hat{\mathbf{c}}$, or else all of them decrease their equilibrium delays.

Lemma 3.10 *Consider two capacity configurations* $\mathbf{c}, \hat{\mathbf{c}} \in \mathcal{C}_\Delta$ *with* $\hat{\mathbf{c}} = \mathbf{c} + \Delta_q(\mathbf{e}_1 - \mathbf{e}_q)$, *where* $\Delta_q > 0$ *is such that* $\hat{\mathcal{L}}^i = \mathcal{L}^i$, *for all* $i \in \mathcal{I}$. *Suppose that* $\mathcal{I}_q \neq \emptyset$. *Then:*

1. *If* $\hat{\lambda}^1 > \lambda^1$, *then* $\hat{\lambda}^i > \lambda^i$, *for all* $i \in \mathcal{I}_q$ *and* $\hat{T}_l > T_l$, *for all links* l *with* $q < l \leq L^1$.

2. *If* $\hat{\lambda}^1 \leq \lambda^1$, *then* $\hat{\lambda}^i \leq \lambda^i$, *for all* $i \in \mathcal{I}_q$ *and* $\hat{T}_l \leq T_l$, *for all links* l *with* $q < l \leq L^1$.

Proof: See the Appendix.

□

In the following lemma we show that if the delay of some link l in $\{2, \ldots, q-1\}$ is higher under configuration $\hat{\mathbf{c}}$, then the same is true for all links in $\{l+1, \ldots, q-1\}$.

Lemma 3.11 *Consider two capacity configurations* $\mathbf{c}, \hat{\mathbf{c}} \in \mathcal{C}_\Delta$ *with* $\hat{\mathbf{c}} = \mathbf{c} + \Delta_q(\mathbf{e}_1 - \mathbf{e}_q)$, *where* $\Delta_q > 0$ *is such that* $\hat{\mathcal{L}}^i = \mathcal{L}^i$, *for all* $i \in \mathcal{I}$. *For any link* $l < q-1$, *if* $\hat{T}_l > T_l$ *then* $\hat{T}_n > T_n$ *for all links* $n \in \{l+1, \ldots, q-1\}$.

Proof: See the Appendix.

□

Let us now proceed by showing that $\hat{\mathbf{c}}$ is userwise price efficient compared to $\mathbf{c}$, under the assumption $\hat{\mathcal{L}}^i = \mathcal{L}^i$, for all $i \in \mathcal{I}$. The proof is given in the following lemma, which asserts also that the equilibrium delays on all links except link q are lower under configuration $\hat{\mathbf{c}}$.

Lemma 3.12 *Consider two capacity configurations $\mathbf{c}, \hat{\mathbf{c}} \in \mathcal{C}_\Delta$ with $\hat{\mathbf{c}} = \mathbf{c} + \Delta_q(\mathbf{e}_1 - \mathbf{e}_q)$, where $\Delta_q > 0$ is such that $\hat{\mathcal{L}}^i = \mathcal{L}^i$, for all $i \in \mathcal{I}$. Then, for every user $i \in \mathcal{I}$, we have $\hat{\lambda}^i \leq \lambda^i$. Furthermore, for every link $l \in \mathcal{L}\backslash\{q\}$, we have $\hat{T}_l \leq T_l$.*

Proof: If no user sends flow to link q, i.e., if $L^1 < q$,[4] transferring capacity from link q to link 1 is, in fact, equivalent to adding capacity to the system of parallel links $\mathcal{L}' = \{1, \ldots, L^1\}$, and the result is immediate from the efficiency of capacity addition in Lemma 3.5. Thus, we have to consider only the case $\mathcal{I}_q \neq \emptyset$, i.e., $L^1 \geq q$. In particular, without loss of generality, we will assume that user 1 (the one with the highest throughput demand) sends flow on all links in the network, i.e., that $L^1 = L$.

Let us first show that $\hat{\lambda}^1 \leq \lambda^1$. Assume by contradiction that $\hat{\lambda}^1 > \lambda^1$. Then, by Lemma 3.10, $\hat{T}_l > T_l$, for all links $l \in \{q+1, \ldots, L\}$. Observe that an immediate consequence of Lemma 3.11 is that there exists some link l_0, $1 \leq l_0 \leq q$, such that $\hat{T}_l \leq T_l$ for all $l \in \{1, \ldots, l_0\}$, and $\hat{T}_l > T_l$ for all $l \in \{l_0 + 1, \ldots, q\}$. Therefore, the delay of all links in $\{1, \ldots, l_0\}$ is lower under configuration $\hat{\mathbf{c}}$, while the delay of all links in $\{l_0 + 1, \ldots, L\}$ is higher. Let us now define: $y_l = |(\hat{c}_l - \hat{f}_l) - (c_l - f_l)|$, $l \in \mathcal{L}$. Note that:

$$\sum\nolimits_{l=1}^{l_0} y_l = \sum\nolimits_{l=l_0+1}^{L} y_l, \tag{3.8}$$

since $\sum_{l\in\mathcal{L}}(\hat{c}_l - \hat{f}_l) = \sum_{l\in\mathcal{L}}(c_l - f_l) = C - R$, and for any link $l \in \{1, \ldots, l_0\}$, we have $\hat{c}_l - \hat{f}_l \geq c_l - f_l$, while for $l \in \{l_0 + 1, \ldots, L\}$ the opposite inequality holds. Recalling (from Lemma 3.3) that $c_l - f_l \geq c_{l+1} - f_{l+1}$, $l \in \mathcal{L}$, we have:

$$\begin{aligned}
\sum_{l=1}^{L}(\hat{c}_l - \hat{f}_l)^2 - \sum_{l=1}^{L}(c_l - f_l)^2 &= \sum_{l=1}^{l_0}\{(c_l - f_l) + y_l\}^2 + \sum_{l=l_0+1}^{L}\{(c_l - f_l) - y_l\}^2 - \sum_{l=1}^{L}(c_l - f_l)^2 \\
&= \sum_{l=1}^{L} y_l^2 + 2\left\{\sum_{l=1}^{l_0}(c_l - f_l)y_l - \sum_{l=l_0+1}^{L}(c_l - f_l)y_l\right\} \\
&\geq \sum_{l=1}^{L} y_l^2 + 2\left\{(c_{l_0} - f_{l_0})\sum_{l=1}^{l_0} y_l - (c_{l_0+1} - f_{l_0+1})\sum_{l=l_0+1}^{L} y_l\right\} \\
&= \sum_{l=1}^{L} y_l^2 + 2\{(c_{l_0} - f_{l_0}) - (c_{l_0+1} - f_{l_0+1})\}\sum_{l=1}^{l_0} y_l > 0,
\end{aligned} \tag{3.9}$$

[4]Recall that $L^1 \geq L^i$, for all $i \in \mathcal{I}$.

where the last equality is obtained using eq. (3.8). From eq. (3.1) and (3.9), we have:[5]

$$\hat{\lambda}^1 = \frac{\sum_{l\in\mathcal{L}^1} c_l - R^{-1}}{\sum_{l\in\mathcal{L}^1}(\hat{c}_l - \hat{f}_l)^2} < \frac{\sum_{l\in\mathcal{L}^1} c_l - R^{-1}}{\sum_{l\in\mathcal{L}^1}(c_l - f_l)^2} = \lambda^1, \tag{3.10}$$

since $\mathcal{L}^1 = \mathcal{L}$. But this contradicts the assumption $\hat{\lambda}^1 > \lambda^1$. Therefore, $\hat{\lambda}^1 \leq \lambda^1$. Lemma 3.10, then, implies that $\hat{T}_l \leq T_l$ for all $l > q$, and $\hat{\lambda}^i \leq \lambda^i$ for all $i \in \mathcal{I}_q$. Thus, $\sum_{i\in\mathcal{I}_q} \hat{\lambda}^i \leq \sum_{i\in\mathcal{I}_q} \lambda^i$. As explained in the remark following the proof of Lemma 3.11 in the Appendix, this implies that $\hat{T}_{q-1} \leq T_{q-1}$. Applying Lemma 3.11 inductively for $l = q-2,\ldots,2$, it follows that $\hat{T}_l \leq T_l$, for every link l in $\{2,\ldots,q-1\}$. In view of $\hat{T}_1 \leq T_1$ (Lemma 3.9), this concludes the proof of the second statement in the lemma.

It remains to be shown that $\hat{\lambda}^i \leq \lambda^i$, for all users $i \in \mathcal{I}$. Assume by contradiction that there exists some user j, such that $\hat{\lambda}^j > \lambda^j$. Then, $j \in \mathcal{I}\backslash\mathcal{I}_q$, since $\hat{\lambda}^i \leq \lambda^i$ for all $i \in \mathcal{I}_q$. Therefore, $\hat{T}_l \leq T_l$, for all $l \in \mathcal{L}^j$. Since $f_l^j T_l' + T_l = \lambda^j < \hat{\lambda}^j = \hat{f}_l^j \hat{T}_l' + \hat{T}_l$, this implies that $\hat{f}_l^j > f_l^j$ for all $l \in \mathcal{L}^j$. Thus, $r^j = \sum_{l\in\mathcal{L}^j} \hat{f}_l^j > \sum_{l\in\mathcal{L}^j} f_l^j = r^j$, which is a contradiction. Therefore, for all $i \in \mathcal{I}$, we have $\hat{\lambda}^i \leq \lambda^i$, and this completes the proof.

□

Let us now generalize the results in Lemma 3.12 to the case where the transfer of capacity from link q to link 1 forces some users to change the set of links over which they send their flow. The general idea is to envision the capacity Δ_q as being transferred in small "steps" $\delta_q(1)$, $\delta_q(2),\ldots$ According to the continuity properties presented in the Appendix (Lemma A.2), each capacity transfer $\delta_q(n)$ can be chosen such that it does not cause any user to change its set of links, but brings some user i exactly to the point where it is about to either stop sending flow to its last link, or to start sending flow to the next one. At that point, we appropriately redefine the set of links over which user i sends its flow, and choose the next transfer of capacity $\delta_q(n+1)$, such that it has the same properties with $\delta_q(n)$. Proceeding this way, the entire capacity Δ_q is eventually transferred after a (possibly infinite) number of steps. Applying Lemma 3.12 to each step, the result follows.

Theorem 3.13 *Consider two capacity configurations* $\mathbf{c}, \hat{\mathbf{c}} \in \mathcal{C}_\Delta$ *with* $\hat{\mathbf{c}} = \mathbf{c} + \Delta_q(\mathbf{e}_1 - \mathbf{e}_q)$, *where* $0 < \Delta_q \leq c_q - c_q^0$. *Then:*

1. *Configuration* $\hat{\mathbf{c}}$ *is userwise price efficient compared to* $\mathbf{c}$, *i.e,* $\hat{\lambda}^i \leq \lambda^i$, *for all* $i \in \mathcal{I}$.
2. *The delay of all links except link* q *is lower under* $\hat{\mathbf{c}}$, *while the delay of link* q *is higher, that is,* $\hat{T}_l \leq T_l$ *for all* $l \in \mathcal{L} \setminus \{q\}$ *and* $\hat{T}_q > T_q$.

[5] Writing (3.1) as $\lambda^1(c_l - f_l)^2 = c_l - f_l + f_l^1$, and summing over $l \in \mathcal{L}^1$, the equalities in (3.10) follow.

Proof: See the Appendix.

□

We are now ready to prove the main result of this section, namely that the capacity configuration that is obtained by allocating the entire additional capacity Δ to link 1 is userwise and, therefore, totally price efficient in $\mathcal{C}_\Delta$.

Theorem 3.14 *Consider a system of parallel links with initial capacity configuration* $\mathbf{c}^0$*, shared by* I *noncooperative users, and an additional capacity allowance* Δ*. The capacity configuration* $\mathbf{c}^* = \mathbf{c}^0 + \Delta\mathbf{e}_1$*, that results from allocating the entire additional capacity to the link with the initially highest capacity (i.e., to link 1), is userwise (thus, totally) price optimal in* $\mathcal{C}_\Delta$*.*

Proof: Let $\mathcal{D}_\delta$ denote the subspace of routing games that is generated by allocating an additional capacity of exactly δ, $0 \le \delta \le \Delta$, to a system of parallel links with initial capacity configuration $\mathbf{c}^0$. Then $\mathcal{C}_\Delta = \cup_{0\le\delta\le\Delta}\mathcal{D}_\delta$. For every δ, define $\mathbf{c}^*(\delta) = \mathbf{c}^0 + \delta\mathbf{e}_1$. Theorem 3.13, then, implies that $\mathbf{c}^*(\delta)$ is userwise price efficient in $\mathcal{D}_\delta$. To see this consider any $\mathbf{c} \in \mathcal{D}_\delta$. From Theorem 3.13, the capacity configuration $\mathbf{c} + (c_L - c_L^0)(\mathbf{e}_1 - \mathbf{e}_L)$, that is obtained by reducing the capacity of link L to its lower bound and adding the excess capacity $(c_L - c_L^0)$ to link 1, is userwise price efficient compared to $\mathbf{c}$. Proceeding inductively, for every $m > 1$, the configuration $\mathbf{c} + \sum_{l=m}^{L}(c_l - c_l^0)(\mathbf{e}_1 - \mathbf{e}_l)$ is userwise price efficient compared to $\mathbf{c} + \sum_{l=m+1}^{L}(c_l - c_l^0)(\mathbf{e}_1 - \mathbf{e}_l)$. Hence, $\mathbf{c}^0 + \delta\mathbf{e}_1 = \mathbf{c} + \sum_{l=2}^{L}(c_l - c_l^0)(\mathbf{e}_1 - \mathbf{e}_l)$ is userwise price efficient with respect to $\mathbf{c}$, that is, $\mathbf{c}^*(\delta)$ is userwise price optimal in $\mathcal{D}_\delta$. From Lemma 3.5, $\mathbf{c}^* = \mathbf{c}^*(\Delta)$ is userwise price efficient with respect to any $\mathbf{c}^*(\delta)$ with $0 \le \delta < \Delta$. Therefore, $\mathbf{c}^*$ is userwise price optimal in $\mathcal{C}_\Delta$.

□

The following two propositions establish that the userwise price optimal capacity configuration $\mathbf{c}^*$ is also userwise cost optimal in some special cases of interest.

Proposition 3.15 *Consider a system of parallel links with initial capacity configuration* $\mathbf{c}^0$*, shared by* I *users, consistent at all capacity configurations in* $\mathcal{C}_\Delta$*, and an additional capacity allowance* Δ*. The capacity configuration* $\mathbf{c}^* = \mathbf{c}^0 + \Delta\mathbf{e}_1$*, that results from allocating the entire additional capacity to the link with the initially highest capacity (i.e., to link 1), is userwise (thus, totally) cost optimal in* $\mathcal{C}_\Delta$*.*

□

Note that the above result applies to the special cases of simple and of identical users.

Proposition 3.16 *Consider a system of parallel links with initial capacity configuration* $\mathbf{c}^0$*, shared by two users, and an additional capacity allowance* Δ*. The capacity configuration* $\mathbf{c}^* = \mathbf{c}^0 + \Delta \mathbf{e}_1$*, that results from allocating the entire additional capacity to the link with the initially highest capacity (i.e., to link 1), is userwise (thus, totally) cost optimal in* $\mathcal{C}_\Delta$*.*

□

3.6 General Topologies

The example presented in the Introduction shows that adding capacity to a network, even in infinite amounts, may result in an increase of both the price and cost of each and every user. This indicates that an upgrade of a general network, in terms of capacity and link addition, should be carried out in a cautious way. In this subsection we establish conditions for the Braess paradox not to occur in any network topology. The terminology introduced in previous subsections for the parallel links model, readily extends to the general case. Due to space limits, we omit the details, which can be found in [KOR94].

We consider now a network $\mathcal{G}(\mathcal{V}, \mathcal{L})$, where $\mathcal{V}$ is a finite set of nodes and $\mathcal{L} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of directed links. A set $\mathcal{I} = \{1, 2, \ldots, I\}$ of users share the network $\mathcal{G}$. We shall assume that all users ship flow from a common source s to a common destination t. As before, each user i has a throughput demand that is some ergodic process with average rate r^i (and $r^1 \geq r^2 \geq \ldots \geq r^I$). User i ships its flow by splitting this demand through the various paths connecting the source to the destination, according to its performance objective. The terms of user flow f_l^i, user routing strategy $\mathbf{f}^i$, user strategy space F^i and system flow configuration $\mathbf{f}$, originally defined in the context of parallel links, readily apply to general topologies, except that now the strategy space F^i of user i should account for the conservation of flow at nodes [KOR94]. The cost function J^i of user i is the sum of link cost functions J_l^i, taken over all network links $l \in \mathcal{L}$. The concepts of Nash equilibria, Nash mapping and optimality conditions are derived similarly as for parallel links. The various versions of the design problem, as previously stated, apply also for the case of general topologies. We note that, for a general topology, the price of a user i is its Lagrange multiplier (implied by the optimality conditions) at the source node.

The class of problems investigated in this paper is well defined if the Nash equilibrium, of any capacity configuration, is unique. Whether this property holds in general topologies is an open question. Thus, we shall concentrate on cases for which uniqueness has been established, such as those of identical users and simple users [ORD93].

Consider an upgrade that is achieved by multiplying the capacity of each link by some constant factor $\alpha > 1$. That is, out of a capacity configuration $\mathbf{c} = (c_l)_{l \in \mathcal{L}}$ we obtain a capacity configuration $\hat{\mathbf{c}} = (\hat{c}_l)_{l \in \mathcal{L}}$, such that, for all $l \in \mathcal{L}$, $\hat{c}_l = \alpha c_l$. We say that $\hat{\mathbf{c}}$ is an *α-product* of $\mathbf{c}$. We then have:

Proposition 3.17 *In a general topology, consider two capacity configurations $\hat{\mathbf{c}}$ and $\mathbf{c}$, such that $\hat{\mathbf{c}}$ is an α-product of $\mathbf{c}$, $\alpha > 1$. Then:*

1. *If the users are simple, then $\hat{\mathbf{c}}$ is userwise price and cost efficient relative to $\mathbf{c}$.*

2. *If users are identical, then $\hat{\mathbf{c}}$ is userwise price efficient relative to $\mathbf{c}$; moreover, for $\alpha > I$, $\hat{\mathbf{c}}$ is also userwise cost efficient.*

□

What the above result suggests is that the potential danger of degradation in performance, as manifested in the Braess paradox, can be avoided by upgrading the network uniformly.

Consider now an upgrade that is achieved by adding capacity to a link between the source s and the destination t (and, as a special case, adding a new link between s and t). Denote by $\mathbf{c}$ and $\hat{\mathbf{c}}$, respectively, the capacity configurations before and after this addition. We say that $\hat{\mathbf{c}}$ is a *direct augmentation* of $\mathbf{c}$. We then have:

Proposition 3.18 *In a general topology, consider two capacity configurations $\hat{\mathbf{c}}$ and $\mathbf{c}$, such that $\hat{\mathbf{c}}$ is a direct augmentation of $\mathbf{c}$. Then:*

1. *If the users are simple, then $\hat{\mathbf{c}}$ is userwise price and cost efficient relative to $\mathbf{c}$.*

2. *If users are identical, then $\hat{\mathbf{c}}$ is userwise price efficient relative to $\mathbf{c}$.*

□

This result suggests that yet another way to avoid the paradox is to upgrade the network through direct connections between source and destination. In fact, this result extends that of Lemma 3.5, which has been obtained for parallel links. We conclude that upgrading direct connections is always beneficial (at least for the classes of users considered), independently of the topology and configuration of other possible connections.

4. Architecting the Flow Configuration in the Run Time Phase

As explained in the Introduction, improvement of the systemwide performance of a noncooperative network can be performed not only in the provisioning phase, but also during the actual operation of the network. In this section we demonstrate this approach based on the noncooperative routing model described in Section 2. We assume that, apart from the flow generated by the self-optimizing users, there is also some flow whose routing is controlled by a central network entity, that will be referred to as the "manager." Typical examples of such flows are the traffic generated by signaling and/or control mechanisms, as well as traffic of users that belong to virtual networks. The manager has the following goals and capabilities: (i) it aims at optimizing the system performance, i.e., the average delay of *all* flow in the network, and (ii) it is cognizant of the user throughput demands (r^i) and of the noncooperative structure of their routing. The first property makes the manager just another user, whose cost function corresponds to the system (rather than its own) performance. The second property, however, enables the manager to predict the response of the noncooperative users to any strategy that it chooses, and hence to determine a strategy of its own flow that would pilot them to a Nash equilibrium that minimizes the manager's cost. Therefore, instead of *reacting* to the routing strategies of the users, the manager *fixes* this strategy and lets them converge to their respective equilibrium.

This is the typical scenario of a Stackelberg game [MYE91] in which the manager plays the role of the "leader," and the noncooperative users play the role of the "followers".[6] An optimal strategy of the leader together with the respective Nash equilibrium of the followers is a *Stackelberg equilibrium.* The presence of sophisticated users that can acquire information about the demands and the cost functions of the other users and become Stackelberg leaders, in order to optimize their own performance, is in general undesirable [SHE94]. However, in the problem considered here, the cost function of the manager is that of the system, and therefore it plays a social, rather than a selfish role.

In this section we investigate the optimal strategy of the leader. In particular, we address the following question: is it possible for the leader to impose a strategy that drives the system into the global optimum, i.e., to the point that corresponds to the solution of a routing problem, in which the leader has full control over the entire flow? Intuitively, one would expect that the leader cannot enforce the global optimum, since it controls only part of the flow, while the rest is controlled by noncooperative users. Rather surprisingly, the results reported in the sequel show that in most cases the leader does have such capability. Due to

[6]The terms "manager" and "leader," as well as "users" and "followers," will be used interchangeably in this section.

space limits, we confine ourselves to a general and brief overview of the results; details can be found in [KOR94b]. The analysis of the special case of a leader and a single follower is presented in the Appendix. We begin with an informal statement of the results.

1. In the special case of a single follower, the leader can always enforce the global optimum, and we specify its optimal strategy.

2. In the general case of any (finite) number of followers, the leader can enforce the global optimum if and only if its demand is larger than some specified threshold $\underline{r}^0$, in which case we specify the leader's optimal strategy.

3. The threshold $\underline{r}^0$ is feasible, in the sense that the total demand of the followers plus $\underline{r}^0$ is lower than the total capacity of the network (assuming, of course, that the total demand of followers itself is less than the total capacity). Thus, for every set of (feasible) followers, there are feasible leaders that can enforce the global optimum.

4. In heavy loaded networks it is "easy" for the leader to enforce the global optimum (i.e., the threshold $\underline{r}^0$ is small).

5. As the number of users increases, it becomes harder for the leader to enforce the global optimum (i.e., the threshold $\underline{r}^0$ increases).

6. The higher the difference in the throughput demand of any two followers, the easier it becomes for the leader to enforce the global optimum. Conversely, the value of the threshold of the leader is highest when all followers are identical.

7. In the case of an infinite number of followers (i.e., the case of simple followers), the leader cannot, in general, enforce the global optimum. For this case, we derive the structure of its optimal strategy and a simple algorithm to compute it.

We proceed with a more detailed description of these results. Consider a system of parallel links $\mathcal{L} = \{1, \ldots, L\}$ shared by a set $\mathcal{I} = \{1, \ldots, I\}$ of noncooperative users (the followers) and a leader, labeled as user 0, who aims at minimizing the total cost of the system:

$$J(\mathbf{f}) = \sum\nolimits_{i=0}^{I} J^i(\mathbf{f}) = \sum\nolimits_{l=1}^{L} \frac{f_l}{c_l - f_l}, \tag{4.1}$$

where we use the notation introduced in Section 2. Each follower $i \in \mathcal{I}$ tries to minimize its individual cost function given by eq. (2.2). As in Section 2, let r^i and F^i denote, respectively, the throughput demand and the strategy space of user $i \in \mathcal{I} \cup \{0\}$. Let $r = \sum_{i \in \mathcal{I}} r^i$ denote the total demand of the followers and $R = r^0 + r$ the total demand of all users. We assume that $R < C$, so that the network is stable.

The leader has knowledge of the noncooperative behavior of the followers and makes its routing decision based on this knowledge. More precisely, let $\mathcal{N}^0(\mathbf{f}^0)$ denote the unique Nash equilibrium of the followers, when the leader employs strategy $\mathbf{f}^0$.[7] The leader seeks a strategy $\mathbf{f}^0 \in F^0$ that minimizes $J(\mathbf{f}^0, \mathcal{N}^0(\mathbf{f}^0))$. It is worth mentioning that this optimization problem is similar to the optimal capacity allocation problem studied in the previous sections. Indeed, the two problems are similar, in the sense that the manager modifies the link capacities that are available to the users. They are different, in the sense that its routing decisions incur a cost for the manager's flow that has to be accounted for in these decisions.

Let $\mathbf{f}^*$ denote the unique solution to the problem of optimally routing the total demand R over the set of parallel links. $\mathbf{f}^*$ will be referred to as the global optimum, and it is easy to see that is determined from Proposition 3.4, by replacing f_l^i with f_l^*, and c_l^i with c_l, $l \in \mathcal{L}$. In the sequel, we consider the problem of finding a strategy $\mathbf{f}^0$ of the leader, such that, if $\mathbf{f} = \mathcal{N}^0(\mathbf{f}^0)$, then $\sum_{i=0}^{I} f_l^i = f_l^*$, $l \in \mathcal{L}$. Clearly, if such a strategy exists, it is an optimal strategy of the leader.

Before we present the results of this section, let us first define:

$$H_l = \sum_{n=1}^{l-1} f_n^* - \frac{f_l^*}{c_l} \sum_{n=1}^{l-1} c_n, \quad l = 2, \ldots, L, \tag{4.2}$$

$$H_0 = 0, \quad H_{L+1} = \sum_{n=1}^{L} f_n^* = R.$$

Then, as shown in the Appendix, we have $H_l \leq H_{l+1}$, for all $l \in \mathcal{L}$.

Consider first the case of a single follower. Except for being the simplest case of the general Stackelberg problem, this case is of interest since it represents practical scenaria, in which different types of traffic (say, "system" and "data") are routed by different entities, one of which is cognizant of the operation of the other, hence the leader-follower setting.

Theorem 4.1 *In a single-follower Stackelberg routing game, consider the strategy* $\mathbf{f}^0$ *of the leader described by:*

$$f_l^0 = \begin{cases} c_l \frac{\sum_{n=1}^{N} f_n^* - r^1}{\sum_{n=1}^{N} c_n}, & l = 1, \ldots, N \\ f_l^*, & l = N+1, \ldots, L \end{cases}, \tag{4.3}$$

where N *is determined by:*

$$H_N < r^1 \leq H_{N+1}. \tag{4.4}$$

[7] $\mathcal{N}^0(\mathbf{f}^0)$ can be determined from Proposition 3.4, by replacing $\mathbf{c}$ with $\mathbf{c} - \mathbf{f}^0$.

Denoting by $\mathbf{f}^1$ *the best reply of the follower to* $\mathbf{f}^0$*, we have that* $f_l^0 + f_l^1 = f_l^*$*, for every link* $l \in \mathcal{L}$*, i.e.,* $\mathbf{f}^0$ *is an optimal strategy of the leader, that achieves the global optimum. Moreover,* $\mathbf{f}^0$ *is the unique optimal strategy of the leader.*

Proof: See the Appendix.

□

The above theorem indicates that, for a single follower, the leader can enforce the global optimum, independently of the relative sizes, in terms of demands, of the leader and the follower. In other words, it is enough to have control on just a nonzero portion of flow in order to "tame" a single selfish user.

We now proceed to the general case of any (finite) number of users. The following lemma describes the strategies of the leader and the followers, assuming that the leader can enforce the global optimum. Later, we will present necessary and sufficient conditions that guarantee that the leader can force the global optimum.

Lemma 4.2 *In a multi-follower Stackelberg routing game, if the leader can enforce the global optimum* $\mathbf{f}^*$*, then its optimal strategy* $\mathbf{f}^0$ *is unique and is given by:*

$$f_l^0 = c_l \sum_{i \in \mathcal{K}_l} \frac{\sum_{n=1}^{N^i} f_n^* - r^i}{\sum_{n=1}^{N^i} c_n} - (K_l - 1) f_l^*, \quad l \in \mathcal{L}, \tag{4.5}$$

where, for every $i \in \mathcal{I}$*,* N^i *is determined by:*

$$H_{N^i} < r^i \leq H_{N^i+1}, \tag{4.6}$$

and for every $l \in \mathcal{L}$*,* $\mathcal{K}_l = \{i \in \mathcal{I} : l \leq N^i\}$ *and* $K_l = |\mathcal{K}_l|$*. In that case, the equilibrium strategy* $\mathbf{f}^i$ *of user* $i \in \mathcal{I}$ *is described by:*

$$f_l^i = \begin{cases} f_l^* - c_l \frac{\sum_{n=1}^{N^i} f_n^* - r^i}{\sum_{n=1}^{N^i} c_n} & , \quad l = 1, \ldots, N^i \\ 0 & , \quad l = N^i + 1, \ldots, L \end{cases}, \tag{4.7}$$

□

Note that if the leader employs strategy $\mathbf{f}^0$, then eq. (4.7) implies that the set of links used by follower i is precisely $\{1, \ldots, N^i\}$, thus $\mathcal{K}_l$ is the set of followers that send flow on link l. In general, $\mathbf{f}^0$ might fail to be an admissible strategy of the leader; however, if it *is* admissible, then it follows from the theorem that it is also the optimal strategy. In

[KOR94b], we show that $\mathbf{f}^0$ is admissible, if and only if the demand of the leader is higher than a threshold $\underline{r}^0$.[8] Therefore, we have the following:

Theorem 4.3 *There exists some $\underline{r}^0$, with $0 \leq \underline{r}^0 < C - r$, such that the leader in a multi-user Stackelberg routing game can enforce the global optimum $\mathbf{f}^*$, if and only if its throughput demand r^0 satisfies $\underline{r}^0 \leq r^0 < C - r$.*

□

From the theorem, it follows that, for any set of followers for which $r < C$, there is a (feasible) leader, with $\underline{r}^0 \leq r^0 < C - r$, that can enforce the global optimum. Moreover, when $r \to C$, we have $\underline{r}^0 \to 0$, meaning that in heavily loaded networks it suffices to control just a small portion of the flow in order to drive the network into the global optimum. It is worth mentioning that even though this behavior might seem surprising, it has a rather intuitive explanation. In the heavy load region, the average delay increases rapidly to infinity, thus small changes in the flow configuration result in drastic changes of the average delay. Therefore, although the leader controls only a small part of the total flow, it has the power to steer the network to the desired global optimum. This result is quite encouraging, because it is in heavily loaded networks where the presence of a manager/leader is particularly important.

The minimum throughput demand $\underline{r}^0$ of the leader that guarantees it can enforce the global optimum depends on the number and the throughput demands of the followers. This dependence is summarized in the following two propositions. The first gives the dependence of $\underline{r}^0$ on the number of followers when their total throughput demand r is fixed. To simplify the formulation of the problem, we concentrate on the case of identical users. The proposition shows that as the number of users increases, the harder it becomes for the leader to enforce $\mathbf{f}^*$.

Proposition 4.4 *Suppose that the followers are identical and their total throughput demand r is fixed. Then, the minimum throughput demand $\underline{r}^0$ that enables the leader to enforce the global optimum $\mathbf{f}^*$ is nondecreasing with the number of followers.*

□

Let us now concentrate on the dependence of $\underline{r}^0$ on differences of the demands of the followers, when their total throughput demand r is fixed. The following proposition shows that the higher the difference in the throughput demand of any two followers, the easier it becomes for the leader to enforce $\mathbf{f}^*$.

[8] The expression for determining $\underline{r}^0$ can be found in [KOR94b].

Proposition 4.5 *Suppose that the total throughput demand r of the followers is fixed. Then, for any two followers j and k, the minimum throughput demand $\underline{r}^0$ that enables the leader to enforce the global optimum $\mathbf{f}^*$ is nonincreasing with $|r^j - r^k|$. Therefore, $\underline{r}^0$ attains its maximum value when all followers are identical.*

□

Let us now demonstrate the properties of $\underline{r}^0$, established in the previous propositions, by means of a numerical example. We consider a system of parallel links with capacity configuration $\mathbf{c} = (12, 7, 5, 3, 2, 1)$, shared by I identical followers with total demand r. Note that the case of identical followers corresponds to the worst case scenario for the leader, according to Proposition 4.5. The threshold $\underline{r}^0$ of the leader is depicted in Figure 5 as a function of r, for various values of I. We concentrate on total follower demands that exceed half the total capacity of the network. In the same figure, we also show the saturation line "$\underline{r}^0 + r = C$". From the figure, one can see that $\underline{r}^0$ always lies below the saturation line, in accordance with Theorem 4.3. Furthermore, $\underline{r}^0$ increases with the number of users. An important observation from the figure is that $\underline{r}^0$ decreases as the total demand of the followers increases, not only in the heavy load region, but also for moderate loads.

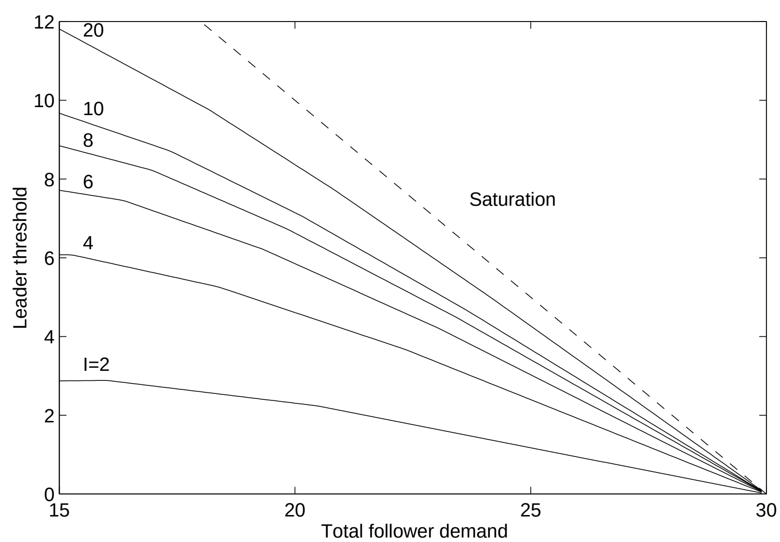

Figure 5: Leader threshold as a function of total follower demand

Finally, let us consider an infinite number of followers, i.e., the case of simple followers. In [KOR94b], we explain that in this case, the leader cannot enforce, in general, the global optimum, i.e., its optimal strategy results with a cost that is higher than the global minimum. Furthermore, we derive the structure of the leader's optimal strategy and specify a simple algorithm to compute it.

5. Conclusions

In this paper we considered design strategies for improving the performance of noncooperative networks. A practical implication of this work is that, design rules for noncooperative networks may follow the same simple patterns that apply to centrally controlled networks, and limited controllability can be as powerful as full controllability.

Our first strategy called for devising proper design rules during the provisioning phase of the network. The problem was formulated as one of allocating additional capacity to an existing noncooperative network. In addition to being prohibitively complex and hard to analyze, this problem exhibits, in general, paradoxical behavior, according to which added resources might degrade the user performance.

For a system of parallel links we established that the addition of capacity guarantees improved performance for all users. Given this result, we showed that the capacity allocation problem has a simple and intuitive solution: the optimal allocation assigns the additional capacity exclusively to the link with the (initially) highest capacity. It is worth noting that, although the noncooperative setting makes the analysis complex and tedious, this solution coincides with the optimal capacity allocation when routing is centrally controlled. For general network topologies, we also derived a set of sufficient conditions that guarantee that the Braess Paradox does not occur.

The second strategy called for improving the performance of the network during its actual operation. This can be achieved by a management entity, that has control on only part of the network flow, and is cognizant of the presence of noncooperative users. Specifically, we considered a network manager that acts as a Stackelberg leader. Considering a system of parallel links, we showed that, in a wide range of cases, by controlling just a small portion of the network flow, the network operating point can be driven into the global optimum. This result suggests that, even with limited controllability of network flows, proper run-time actions can diminish considerably, or even avoid altogether, the inefficiency implicated by noncooperative users.

Some conclusions can also be derived from our investigation of general topologies, for which we provided ways to overcome the Braess paradox. One indication from our results is that capacity should be added across the network, rather than on a local (e.g., single link) scale. This fits well with common engineering practice, where common folklore suggests that local improvement may only result in transferring the problem somewhere else in the system. Another indication is that upgrades should be aimed at direct connections between the source and the destination. Indeed, we have seen that, for any topology, the addition of capacity to direct “links” is always efficient. This is yet a further indication of the potential benefit of decoupling complex structures in the network, so that the corresponding controllers (in our context, the routing controllers) are presented with simple choices.

Acknowledgments

The research reported here was supported by the Office of Naval Research under Contract # N00014–90–J–1289. The authors would like to thank Raphael Rom and Nahum Shimkin for helpful discussions.

APPENDIX

A. Continuity of the Nash Mapping

In this appendix we present the continuity properties of the Nash mapping that have been established in [KOR94]. To this end, let us augment the definition of the Nash mapping, so that to each capacity configuration $\mathbf{c} \in \mathcal{C}_\Delta$ it assigns the Nash equilibrium of the routing game and the corresponding Lagrange multipliers λ^i, $i \in \mathcal{I}$, i.e., $\mathcal{N}(\mathbf{c}) = (\mathbf{f}, (\lambda^i)_{i\in\mathcal{I}})$. In [KOR94], we prove the following:

Theorem A.1 *The Nash mapping $\mathcal{N} : \mathcal{C}_\Delta \to I\!R^{I(L+1)}$ is continuous. Furthermore, let $\Phi_1 : \mathcal{C}_\Delta \to I\!R^I_+$ and $\Phi_2 : \mathcal{C}_\Delta \to I\!R^L_+$ be such that for every $\mathbf{c} \in \mathcal{C}_\Delta$, $\Phi_1(\mathbf{c}) = (J^1, \ldots, J^I)$, where J^i is the equilibrium cost of user $i \in \mathcal{I}$ under the capacity configuration $\mathbf{c}$, and $\Phi_2(\mathbf{c}) = (T_1, \ldots, T_L)$, where T_l is the equilibrium delay of link $l \in \mathcal{L}$ under $\mathbf{c}$. Then, Φ_1 and Φ_2 are continuous.*

The continuity of the Nash mapping is employed to establish that the analysis of the optimal capacity allocation problem can be carried based on comparisons of capacity configurations $\mathbf{c}, \hat{\mathbf{c}} \in \mathcal{C}_\Delta$, which are such that each user sends its flow over the same set of links under both configurations. More specifically, we compare configurations $\mathbf{c}$ and $\hat{\mathbf{c}}$, such that $\hat{\mathbf{c}}$ results from $\mathbf{c}$ by a transfer of capacity Δ_q from some link q to a link l, i.e., $\hat{\mathbf{c}} = \mathbf{c} + \Delta_q(\mathbf{e}_l - \mathbf{e}_q)$. In [KOR94], we show that if this transfer of capacity is sufficiently small, i.e., if the distance of $\mathbf{c}$ and $\hat{\mathbf{c}}$ in $\mathcal{C}_\Delta$ is small, then each user sends its flow over the same set of links under both configurations. The results are given in the following:

Lemma A.2 *For any capacity configuration $\mathbf{c} \in \mathcal{C}_\Delta$ with $c_q > c_q^0$ and $c_l < c_{l-1}$, there exists some η, $0 < \eta \leq \min\{c_q - c_q^0, c_{l-1} - c_l\}$, such that if $\hat{\mathbf{c}}$ is the configuration that results from a transfer of capacity $\Delta_q \leq \min\{c_q - c_q^0, c_{l-1} - c_l\}$ from link q to link l, i.e., if $\hat{\mathbf{c}} = \mathbf{c} + \Delta_q(\mathbf{e}_l - \mathbf{e}_q)$, then:*

1. *If $\Delta_q \leq \eta$, then each user $i \in \mathcal{I}$ sends its traffic over the same links under both capacity configurations $\mathbf{c}$ and $\hat{\mathbf{c}}$.*

2. *If $\Delta_q > \eta$, then there exists some user $i \in \mathcal{I}$ that changes the set of links over which it sends its flow.*

B. Proof of results in Subsection 3.4

Proof of Lemma 3.5

Assume by contradiction that the set $\mathcal{T}^+$, defined as $\mathcal{T}^+ = \{l : \hat{T}_l > T_l\}$, is nonempty. Since the flow in each $l \in \mathcal{T}^+$ is higher under configuration $\hat{\mathbf{c}}$, there must be a user i and links $l \in \mathcal{T}^+$ and $n \notin \mathcal{T}^+$, such that $\hat{f}_l^i > f_l^i$ and $\hat{f}_n^i < f_n^i$. Since $\hat{T}_l > T_l$, $\hat{T}_n \le T_n$, and $\hat{f}_l^i > f_l^i \ge 0$, the optimality conditions (2.4)–(2.6) imply that $\hat{f}_n^i \hat{T}_n + \hat{T}_n \ge \hat{f}_l^i \hat{T}_l' + \hat{T}_l$, and similarly, since $f_n^i > \hat{f}_n^i \ge 0$, we have $f_n^i T_n' + T_n \le f_l^i T_l' + T_l$, which, combined with $\hat{T}_n \le T_n$ and $\hat{f}_n^i < f_n^i$, yields $\hat{f}_l^i \hat{T}_l' + \hat{T}_l \le f_l^i T_l' + T_l$. Since $\hat{f}_l^i > f_l^i$ and $\hat{T}_l > T_l$, this is a contradiction. We, thus, conclude that the set $\mathcal{T}^+$ is empty, i.e., $\hat{T}_l \le T_l$ for all links l. Since the demand of each user i is r^i in both capacity configurations, it cannot increase its flow on all links, thus there must be a link l for which $f_l^i > 0$ and $\hat{f}_l^i \le f_l^i$. Therefore:

$$\hat{\lambda}^i \le \hat{f}_l^i \hat{T}_l' + \hat{T}_l \le f_l^i T_l' + T_l = \lambda^i,$$

thus concluding the proof.

□

Proof of Lemma 3.6

From Lemma 3.5 we have that $\hat{T}_l \le T_l$ for all links l. This means that, for all $l > 1$, we have $\hat{f}_l \le f_l$. Therefore, when "moving" from $\mathbf{c}$ to $\hat{\mathbf{c}}$ we observe flow being "transferred" from all links $l > 1$ to link 1. Since $\hat{T}_l \le T_l$ for all links l, the flow that remains (under configuration $\hat{\mathbf{c}}$) in a link $l > 1$, experiences a delay which is not higher than the previous one. Moreover, since $\hat{T}_1 \le T_1 \le T_l$, we conclude that also the flow that moved to link 1 experiences a delay which is not higher than the previous one, and the lemma follows.

□

C. Proofs of results in Subsection 3.5

We begin by deriving an alternative expression for the user prices that will be used throughout the following proofs. Noting that eq. (3.1) can be written as $\lambda^i (c_l - f_l)^2 = c_l - f_l^{-i}$, $l \in \mathcal{L}^i$, and summing over any set of links $A \subseteq \mathcal{L}^i$ that receive some flow from user i, we get:

$$\lambda^i = \frac{\sum_{l \in A}(c_l - f_l^{-i})}{\sum_{l \in A}(c_l - f_l)^2} = \frac{\sum_{l \in A} c_l - R^{-i} + \sum_{l \in \mathcal{L}^1 \setminus A} f_l^{-i}}{\sum_{l \in A}(c_l - f_l)^2}, \quad A \subseteq \mathcal{L}^i, \tag{C.1}$$

since $\sum_{l \in A} f_l^{-i} = R^{-i} - \sum_{l \in \mathcal{L} \setminus A} f_l^{-i}$. Note that, in the last summation, $\mathcal{L} \setminus A$ can be replaced by $\mathcal{L}^1 \setminus A$, because no user sends flow over links in $\mathcal{L} \setminus \mathcal{L}^1$. Taking $A = \mathcal{L}^i$ in eq. (C.1), we have:

$$\lambda^i = \frac{\sum_{l\in\mathcal{L}^i} c_l - R^{-i} + \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^i} f_l}{\sum_{l\in\mathcal{L}^i}(c_l - f_l)^2}, \tag{C.2}$$

since, for all $l \in \mathcal{L}^1\setminus\mathcal{L}^i$, we have $f_l^i = 0$ and, therefore, $f_l^{-i} = f_l$.

From eqs. (2.4) and (2.7) it is also easy to see that:

$$\sum_{i\in\mathcal{I}_l} \lambda^i = f_l T_l' + I_l T_l, \quad l \in \mathcal{L}. \tag{C.3}$$

Let $\mathcal{T}^+$ (respectively, $\mathcal{T}^-$) denote the set of links in $\mathcal{L}\setminus\{1,q\}$ whose equilibrium delay is higher (respectively, is not higher) under $\hat{\mathbf{c}}$, i.e., $\mathcal{T}^+ = \{l \in \mathcal{L}\setminus\{1,q\} : \hat{T}_l > T_l\}$ and $\mathcal{T}^- = \{l \in \mathcal{L}\setminus\{1,q\} : \hat{T}_l \le T_l\}$. Since the capacity of any link in $l \in \mathcal{L}\setminus\{1,q\}$ does not change ($c_l = \hat{c}_l$), eq. (C.3) implies that:

$$\text{For any } l \in \mathcal{L}\setminus\{1,q\}: \quad l \in \mathcal{T}^+ \;\Leftrightarrow\; \hat{f}_l > f_l \;\Leftrightarrow\; \sum_{i\in\mathcal{I}_l} \hat{\lambda}^i > \sum_{i\in\mathcal{I}_l} \lambda^i, \tag{C.4}$$

$$\text{For any } l \in \mathcal{L}\setminus\{1,q\}: \quad l \in \mathcal{T}^- \;\Leftrightarrow\; \hat{f}_l \le f_l \;\Leftrightarrow\; \sum_{i\in\mathcal{I}_l} \hat{\lambda}^i \le \sum_{i\in\mathcal{I}_l} \lambda^i. \tag{C.5}$$

We proceed with the following technical lemma:

Lemma C.1 *Consider two capacity configurations* $\mathbf{c}, \hat{\mathbf{c}} \in \mathcal{C}_\Delta$. *For any user* $i \in \mathcal{I}$:

1. *If* $\hat{\lambda}^i > \lambda^i$, *then* $\hat{f}_l^i > f_l^i$, *for all* $l \in \mathcal{L}^i$ *such that* $\hat{T}_l \le T_l$.
2. *If* $\hat{\lambda}^i \le \lambda^i$, *then* $\hat{f}_l^i < f_l^i$, *for all* $l \in \mathcal{L}^i$ *such that* $\hat{T}_l > T_l$.
3. *There cannot be two links* $m, n \in \mathcal{L}^i$, *such that* $\hat{T}_n > T_n$, $\hat{T}_m \le T_m$, $\hat{f}_n^i \ge f_n^i$ *and* $\hat{f}_m^i \le f_m^i$.

Proof: Assume that there is a user i with $\hat{\lambda}^i > \lambda^i$ and a link $l \in \mathcal{L}^i$ such that $\hat{T}_l \le T_l$ and $\hat{f}_l^i \le f_l^i$. Then, $\hat{\lambda}^i = \hat{f}_l^i \hat{T}_l' + \hat{T}_l \le f_l^i T_l' + T_l = \lambda^i$, which contradicts $\hat{\lambda}^i > \lambda^i$. The proof of part (2) is symmetric. For part (3), assume that there are such links $n, m \in \mathcal{L}^i$. Since $m \in \mathcal{L}^i$, we have $\lambda^i = f_m^i T_m' + T_m \ge \hat{f}_m^i \hat{T}_m' + T_m \ge \hat{\lambda}^i$. Then part (2) implies that $\hat{f}_n^i < f_n^i$.

□

We are now ready to prove the claims of Subsection 3.5.

Proof of Lemma 3.9

Assume by contradiction that $\hat{T}_1 > T_1$. We have to consider two cases:

(i) $\hat{T}_q > T_q$

For any link $l \in \mathcal{T}^+ \cup \{1,q\}$, we have $\hat{T}_l > T_l$. Thus:

$$\sum_{l \in \mathcal{T}^+ \cup \{1,q\}} (\hat{c}_l - \hat{f}_l) < \sum_{l \in \mathcal{T}^+ \cup \{1,q\}} (c_l - f_l) \quad \Rightarrow \quad \sum_{l \in \mathcal{T}^+ \cup \{1,q\}} \hat{f}_l > \sum_{l \in \mathcal{T}^+ \cup \{1,q\}} f_l,$$

since $\hat{c}_1 + \hat{c}_q = c_1 + c_q$ and $\hat{c}_n = c_n$, for all $n \in \mathcal{T}^+$. This implies that there must be a user j whose total flow in $\mathcal{T}^+ \cup \{1,q\}$ is higher under $\hat{\mathbf{c}}$, i.e.:

$$\sum_{l \in \mathcal{T}^+ \cup \{1,q\}} \hat{f}_l^j > \sum_{l \in \mathcal{T}^+ \cup \{1,q\}} f_l^j \quad \Rightarrow \quad \sum_{l \in \mathcal{T}^-} \hat{f}_l^j < \sum_{l \in \mathcal{T}^-} f_l^j$$

Thus, there must be links $n \in \mathcal{T}^+ \cup \{1,q\}$ and $m \in \mathcal{T}^-$, such that $\hat{f}_n^j > f_n^j$ and $\hat{f}_m^j < f_m^j$. Since $\hat{T}_n > T_n$ and $\hat{T}_m \leq T_m$, this is a contradiction to part (3) of Lemma C.1.

(ii) $\hat{T}_q \leq T_q$

In this case $\hat{f}_q < f_q$, since $\hat{c}_q < c_q$. Using (C.5), we get $\sum_{l \in \mathcal{T}^- \cup \{q\}} \hat{f}_l < \sum_{l \in \mathcal{T}^- \cup \{q\}} f_l$, thus there must be a user j, such that:

$$\sum_{l \in \mathcal{T}^- \cup \{q\}} \hat{f}_l^j < \sum_{l \in \mathcal{T}^- \cup \{q\}} f_l^j \quad \Rightarrow \quad \sum_{l \in \mathcal{T}^+ \cup \{1\}} \hat{f}_l^j > \sum_{l \in \mathcal{T}^+ \cup \{1\}} f_l^j.$$

This implies that there must be links $m \in \mathcal{T}^- \cup \{q\}$ and $n \in \mathcal{T}^+ \cup \{1\}$, such that $\hat{f}_m^j < f_m^j$ and $\hat{f}_n^j > f_n^j$. Since $\hat{T}_n > T_n$ and $\hat{T}_m \leq T_m$, this is a contradiction to part (3) of Lemma C.1.

Therefore, the delay on link 1 cannot be higher under capacity configuration $\hat{\mathbf{c}}$, i.e., $\hat{T}_1 \leq T_1$. Let us now proceed to show the second part of the lemma, i.e., that $\hat{T}_q > T_q$.

Suppose that $\hat{T}_q \leq T_q$. Let us first show that this implies that $\mathcal{T}^+ = \emptyset$. Assume by contradiction that $\mathcal{T}^+$ is nonempty. Then, the total flow sent over links in $\mathcal{T}^+$ is higher under $\hat{\mathbf{c}}$ – since $\hat{c}_l = c_l$ for all $l \in \mathcal{T}^+$ – and there must be a user j, such that:

$$\sum_{l \in \mathcal{T}^+} \hat{f}_l^j > \sum_{l \in \mathcal{T}^+} f_l^j \quad \Rightarrow \quad \sum_{l \in \mathcal{T}^- \cup \{1,q\}} \hat{f}_l^j < \sum_{l \in \mathcal{T}^- \cup \{1,q\}} f_l^j.$$

Thus, there exist links $n \in \mathcal{T}^+$ and $m \in \mathcal{T}^- \cup \{1,q\}$, such that $\hat{f}_n^j > f_n^j$ and $\hat{f}_m^j < f_m^j$. Since $\hat{T}_n > T_n$ and $\hat{T}_m \leq T_m$, this contradicts part (3) of Lemma C.1. Therefore, $\mathcal{T}^+ = \emptyset$, and:

$$\hat{c}_l - \hat{f}_l \geq c_l - f_l, \quad l \in \mathcal{L}. \tag{C.6}$$

Recalling that $\sum_{l \in \mathcal{L}} (\hat{c}_l - \hat{f}_l) = \sum_{l \in \mathcal{L}} (c_l - f_l) = C - R$, (C.6) must hold as an equality for all $l \in \mathcal{L}$, or equivalently:

$$\hat{T}_l = T_l, \quad l \in \mathcal{L}. \tag{C.7}$$

Note that this implies that no user can change its flow on any link in the network. To see this, assume that there exists a user j such that $\hat{f}_n^j > f_n^j$ and $\hat{f}_m^j < f_m^j$, for some links $m, n \in \mathcal{L}^j$. Then, using eq. (C.7), we have:

$$\hat{\lambda}^j = \hat{f}_m^j \hat{T}'_m + \hat{T}_m < f_m^j T'_m + T_m = f_n^j T'_n + T_n < \hat{f}_n^j \hat{T}'_n + \hat{T}_n = \hat{\lambda}^j.$$

Thus, no user modifies its flow configuration, and $\hat{f}_l = f_l$, for all $l \in \mathcal{L}$. But this contradicts $\hat{T}_q \leq T_q$, since $\hat{c}_q < c_q$ and $\hat{f}_q = f_q$, and the result follows.

□

Remark: Lemma 3.9 does not rely on the assumption $\hat{\mathcal{L}}^i = \mathcal{L}^i$, for all $i \in \mathcal{I}$, as can be seen from the above proof.

Proof of Lemma 3.10

First we consider the case $\hat{\lambda}^1 > \lambda^1$. Then eq. (C.2) for $i = 1$ gives:

$$\frac{\sum_{l \in \mathcal{L}^1} \hat{c}_l - R^{-1}}{\sum_{l \in \mathcal{L}^1} (\hat{c}_l - \hat{f}_l)^2} > \frac{\sum_{l \in \mathcal{L}^1} c_l - R^{-1}}{\sum_{l \in \mathcal{L}^1} (c_l - f_l)^2}.$$

Since $\sum_{l=1}^{m} \hat{c}_l = \sum_{l=1}^{m} c_l$, for all $m \geq q$, we have:

$$\sum_{l \in \mathcal{L}^1} (\hat{c}_l - \hat{f}_l)^2 < \sum_{l \in \mathcal{L}^1} (c_l - f_l)^2. \tag{C.8}$$

Let us start by proving that $\hat{\lambda}^2 > \lambda^2$. If $\mathcal{L}^2 = \mathcal{L}^1$, the result is immediate from (C.8):

$$\hat{\lambda}^2 = \frac{\sum_{l \in \mathcal{L}^1} \hat{c}_l - R^{-2}}{\sum_{l \in \mathcal{L}^1} (\hat{c}_l - \hat{f}_l)^2} > \frac{\sum_{l \in \mathcal{L}^1} c_l - R^{-2}}{\sum_{l \in \mathcal{L}^1} (c_l - f_l)^2} = \lambda^2.$$

Therefore, we concentrate on the case $\mathcal{L}^1 \backslash \mathcal{L}^2 \neq \emptyset$. Since $\mathcal{L}^I \subseteq \ldots \mathcal{L}^2 \subseteq \mathcal{L}^1$, user 1 is the only user that sends flow on any link in $\mathcal{L}^1 \backslash \mathcal{L}^2$. For any such link l, we have $f_l = f_l^1$, $\hat{f}_l = \hat{f}_l^1$, and in view of eqs. (2.4) and (2.7), $\hat{\lambda}^1 > \lambda^1$ implies that:

$$\hat{f}_l = \hat{f}_l^1 > f_l^1 = f_l, \quad l \in \mathcal{L}^1 \backslash \mathcal{L}^2, \tag{C.9}$$

From eq. (C.1) with $A = \mathcal{L}^2$ and $i = 1$, we have:

$$\frac{\sum_{l\in\mathcal{L}^2} c_l - R^{-1}}{\sum_{l\in\mathcal{L}^2}(\hat{c}_l - \hat{f}_l)^2} = \hat{\lambda}^1 > \lambda^1 = \frac{\sum_{l\in\mathcal{L}^2} c_l - R^{-1}}{\sum_{l\in\mathcal{L}^2}(c_l - f_l)^2} \;\Rightarrow\; \sum_{l\in\mathcal{L}^2}(\hat{c}_l - \hat{f}_l)^2 < \sum_{l\in\mathcal{L}^2}(c_l - f_l)^2.$$

Therefore, eqs. (C.2) and (C.9) give:

$$\hat{\lambda}^2 = \frac{\sum_{l\in\mathcal{L}^2} c_l - R^{-2} + \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^2}\hat{f}_l}{\sum_{l\in\mathcal{L}^2}(\hat{c}_l - \hat{f}_l)^2} > \frac{\sum_{l\in\mathcal{L}^2} c_l - R^{-2} + \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^2} f_l}{\sum_{l\in\mathcal{L}^2}(c_l - f_l)^2} = \lambda^2,$$

which completes the proof for $i = 2$.

Proceeding inductively, let us assume that $\hat{\lambda}^i > \lambda^i$, for all $i \le k < I_q$, and show that the same holds for $i = k+1$. If $\mathcal{L}^{k+1} = \mathcal{L}^1$, the proof is immediate from (C.2) and (C.8):

$$\hat{\lambda}^{k+1} = \frac{\sum_{l\in\mathcal{L}^1} c_l - R^{-(k+1)}}{\sum_{l\in\mathcal{L}^1}(\hat{c}_l - \hat{f}_l)^2} > \frac{\sum_{l\in\mathcal{L}^1} c_l - R^{-(k+1)}}{\sum_{l\in\mathcal{L}^1}(c_l - f_l)^2} = \lambda^{k+1},$$

thus, we only have to consider the case $\mathcal{L}^1\setminus\mathcal{L}^{k+1} \neq \emptyset$. Let $\mathcal{I}_0$ denote the set of users who send flow on some link in $\mathcal{L}^1\setminus\mathcal{L}^{k+1}$, i.e., $\mathcal{I}_0 = \cup_{\mathcal{L}^1\setminus\mathcal{L}^{k+1}}\mathcal{I}_l$. Note that user $k+1$ does not send flow on any link in $\mathcal{L}^1\setminus\mathcal{L}^{k+1}$. By Lemma 3.3, the same is true for all users $i > k+1$. Thus, $\mathcal{I}_0 \subseteq \{1,\ldots,k\}$ and by the inductive hypothesis we have: $\hat{\lambda}^i > \lambda^i$, for all users $i \in \mathcal{I}_0$. Since $\mathcal{I}_l \subseteq \mathcal{I}_0$, for all $l \in \mathcal{L}^1\setminus\mathcal{L}^{k+1}$, this implies:

$$\sum_{i\in\mathcal{I}_l}\hat{\lambda}^i > \sum_{i\in\mathcal{I}_l}\lambda^i, \quad l \in \mathcal{L}^1\setminus\mathcal{L}^{k+1}, \tag{C.10}$$

and (C.4) gives:

$$\hat{f}_l > f_l, \quad l \in \mathcal{L}^1\setminus\mathcal{L}^{k+1}. \tag{C.11}$$

Since $k+1 > i$, for all $i \in \mathcal{I}_0$, we have that $\mathcal{L}^{k+1} \subseteq \mathcal{L}^i$, for all $i \in \mathcal{I}_0$, i.e., any user that sends flow on some link in $\mathcal{L}^1\setminus\mathcal{L}^{k+1}$, also sends flow on all links in $\mathcal{L}^{k+1}$. Hence, taking $A = \mathcal{L}_{k+1}$ in eq. (C.1), we get:

$$\lambda^i = \frac{\sum_{l\in\mathcal{L}^{k+1}} c_l - R^{-i} + \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}}(f_l - f_l^i)}{\sum_{l\in\mathcal{L}^{k+1}}(c_l - f_l)^2}, \quad i \in \mathcal{I}_0. \tag{C.12}$$

From eq. (C.2), we have:

$$\lambda^{k+1} \;=\; \frac{\sum_{l\in\mathcal{L}^{k+1}} c_l - R^{-(k+1)} + \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} f_l}{\sum_{l\in\mathcal{L}^{k+1}}(c_l - f_l)^2} \tag{C.13}$$

$$= \frac{\sum_{l\in\mathcal{L}^{k+1}} c_l - R^{-i} + \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}}(f_l - f_l^i)}{\sum_{l\in\mathcal{L}^{k+1}}(c_l - f_l)^2} - \frac{r^i - r^{k+1} - \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} f_l^i}{\sum_{l\in\mathcal{L}^{k+1}}(c_l - f_l)^2},\ i\in\mathcal{I}_0,$$

where we have used $R^{-(k+1)} = R^{-i} + r^i - r^{k+1}$. Eqs. (C.12) and (C.13) give:

$$\lambda^{k+1} = \lambda^i - \frac{\sum_{l\in\mathcal{L}^{k+1}} f_l^i - r^{k+1}}{\sum_{l\in\mathcal{L}^{k+1}}(c_l - f_l)^2},\quad i\in\mathcal{I}_0. \tag{C.14}$$

Let us now assume that:

$$\hat{\lambda}^{k+1} \le \lambda^{k+1}. \tag{C.15}$$

Since $q\in\mathcal{L}_{k+1}$, we have $\sum_{l\in\mathcal{L}_{k+1}} \hat{c}_l = \sum_{l\in\mathcal{L}_{k+1}} c_l$. Eqs. (C.13) and (C.15), then, imply:

$$\frac{\sum_{l\in\mathcal{L}^{k+1}} c_l - R^{-(k+1)} + \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} \hat{f}_l}{\sum_{l\in\mathcal{L}^{k+1}}(\hat{c}_l - \hat{f}_l)^2} \le \frac{\sum_{l\in\mathcal{L}^{k+1}} c_l - R^{-(k+1)} + \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} f_l}{\sum_{l\in\mathcal{L}^{k+1}}(c_l - f_l)^2},$$

and since $\sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} \hat{f}_l > \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} f_l$, according to (C.11), we have:

$$\sum_{l\in\mathcal{L}^{k+1}} (\hat{c}_l - \hat{f}_l)^2 > \sum_{l\in\mathcal{L}^{k+1}} (c_l - f_l)^2. \tag{C.16}$$

Using eq. (C.14) and (C.15) we have, for $i\in\mathcal{I}_0$:

$$\frac{\sum_{l\in\mathcal{L}^{k+1}} \hat{f}_l^i - r^{k+1}}{\sum_{l\in\mathcal{L}^{k+1}}(\hat{c}_l - \hat{f}_l)^2} - \frac{\sum_{l\in\mathcal{L}^{k+1}} f_l^i - r^{k+1}}{\sum_{l\in\mathcal{L}^{k+1}}(c_l - f_l)^2} \ge \hat{\lambda}^i - \lambda^i > 0,$$

since $\hat{\lambda}^i > \lambda^i$, for all $i\in\mathcal{I}_0$. In view of (C.16), this implies:

$$\sum_{l\in\mathcal{L}^{k+1}} \hat{f}_l^i > \sum_{l\in\mathcal{L}^{k+1}} f_l^i \ \Rightarrow\ \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} \hat{f}_l^i < \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} f_l^i,\quad i\in\mathcal{I}_0.$$

By the definition of set $\mathcal{I}_0$, summing the last inequality over all $i\in\mathcal{I}_0$, we get:

$$\sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} \hat{f}_l < \sum_{l\in\mathcal{L}^1\setminus\mathcal{L}^{k+1}} f_l,$$

which is a contradiction to (C.11). Hence $\hat{\lambda}^{k+1} > \lambda^{k+1}$, and by induction:

$$\hat{\lambda}^i > \lambda^i,\ \text{for all}\ i\in\mathcal{I}_q.$$

For any link $l > q$, $\hat{\lambda}^i > \lambda^i$, for all $i\in\mathcal{I}_l$, and (C.4) implies $\hat{T}_l > T_l$. This completes the proof for case $\hat{\lambda}^1 > \lambda^1$. The proof for the case $\hat{\lambda}^1 \le \lambda^1$ is symmetric.

□

Proof of Lemma 3.11

It suffices to show that for any link $l < q-1$, if $l \in \mathcal{T}^+$, then $l+1 \in \mathcal{T}^+$. Assume by contradiction that there exists a link $l < q-1$, such that $l \in \mathcal{T}^+$ and $l+1 \in \mathcal{T}^-$. Then (C.4) and (C.5) give:

$$\hat{f}_{l+1} \leq f_{l+1} \text{ and } \sum_{i \in \mathcal{I}_{l+1}} \hat{\lambda}^i \leq \sum_{i \in \mathcal{I}_{l+1}} \lambda^i, \tag{C.17}$$

$$\hat{f}_l > f_l \text{ and } \sum_{i \in \mathcal{I}_l} \hat{\lambda}^i > \sum_{i \in \mathcal{I}_l} \lambda^i. \tag{C.18}$$

If $\mathcal{I}_{l+1} = \mathcal{I}_l$, (C.17) and (C.18) lead to a contradiction. Thus, we need to consider only the case $\mathcal{I}_l \backslash \mathcal{I}_{l+1} \neq \emptyset$. Note that this is the set of users that send flow on link l and do not send flow on link $l+1$. For any such user i, $L^i = l$. Summing eq. (2.4) for link l over all $i \in \mathcal{I}_{l+1} \subset \mathcal{I}_l$ and using eq. (2.7), we get:

$$\sum_{i \in \mathcal{I}_{l+1}} \hat{f}_l^i \hat{T}_l' + I_{l+1} \hat{T}_l = \sum_{i \in \mathcal{I}_{l+1}} \hat{\lambda}^i \leq \sum_{i \in \mathcal{I}_{l+1}} \lambda^i = \sum_{i \in \mathcal{I}_{l+1}} f_l^i T_l' + I_{l+1} T_l,$$

and since $l \in \mathcal{T}^+$, this implies:

$$\sum_{i \in \mathcal{I}_{l+1}} \hat{f}_l^i < \sum_{i \in \mathcal{I}_{l+1}} f_l^i. \tag{C.19}$$

Recalling that $\hat{f}_l > f_l$, (C.19) implies:

$$\sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} \hat{f}_l^i = \hat{f}_l - \sum_{i \in \mathcal{I}_{l+1}} \hat{f}_l^i > f_l - \sum_{i \in \mathcal{I}_{l+1}} f_l^i = \sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} f_l^i,$$

and since for any $i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}$, we have $r^i = \sum_{m=1}^{l} f_m^i = \sum_{m=1}^{l} \hat{f}_m^i$, this implies:

$$\sum_{m<l} \sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} \hat{f}_m^i < \sum_{m<l} \sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} f_m^i. \tag{C.20}$$

From (C.17) and (C.18), we have that $\sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} \hat{\lambda}^i > \sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} \lambda^i$, and summing eq. (2.4) over all $i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}$, for any link $m \leq l$, it is easy to see that this implies:

$$\sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} \hat{f}_m^i \hat{T}_m' + (I_l - I_{l+1}) \hat{T}_m > \sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} f_m^i T_m' + (I_l - I_{l+1}) T_m, \quad m \leq l. \tag{C.21}$$

Consider now any link $m < l$, such that $m \in \mathcal{T}^- \cup \{1\}$. Since $\hat{T}_m' > T_m$ and $I_l - I_{l+1} =$

$|\mathcal{I}_l \backslash \mathcal{I}_{l+1}| > 0$, (C.21) implies that $\sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} \hat{f}_m^i > \sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} f_m^i$. Hence, by (C.20), we have:

$$\sum_{\substack{m \in \mathcal{T}^+ \\ m<l}} \sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} \hat{f}_m^i < \sum_{\substack{m \in \mathcal{T}^+ \\ m<l}} \sum_{i \in \mathcal{I}_l \backslash \mathcal{I}_{l+1}} f_m^i. \tag{C.22}$$

Using a similar argument as in the proof of (C.19), one can see that (C.17) implies:

$$\sum_{i \in \mathcal{I}_{l+1}} \hat{f}_m^i < \sum_{i \in \mathcal{I}_{l+1}} f_m^i, \quad m \in \mathcal{T}^+ \text{ and } m < l.$$

Summing this inequality over all $m \in \mathcal{T}^+$ and $m < l$, and adding it to (C.22), we have:

$$\sum_{\substack{m \in \mathcal{T}^+ \\ m<l}} \sum_{i \in \mathcal{I}_l} \hat{f}_m^i < \sum_{\substack{m \in \mathcal{T}^+ \\ m<l}} \sum_{i \in \mathcal{I}_l} f_m^i. \tag{C.23}$$

For any link $m \in \mathcal{T}^+$, $\hat{f}_m > f_m$. Therefore, the total flow sent to the set of links $\mathcal{T}^+ \cap \{1, \ldots, l-1\}$ is higher under configuration $\hat{\mathbf{c}}$, and (C.23) implies:

$$\sum_{\substack{m \in \mathcal{T}^+ \\ m<l}} \sum_{i \in \mathcal{I} \backslash \mathcal{I}_l} \hat{f}_m^i > \sum_{\substack{m \in \mathcal{T}^+ \\ m<l}} \sum_{i \in \mathcal{I} \backslash \mathcal{I}_l} f_m^i.$$

Therefore, there exists some user $j \in \mathcal{I} \backslash \mathcal{I}_l$, such that:

$$\sum_{\substack{m \in \mathcal{T}^+ \\ m<l}} \hat{f}_m^j > \sum_{\substack{m \in \mathcal{T}^+ \\ m<l}} f_m^j,$$

or equivalently:

$$\sum_{m \in \mathcal{T}^+} \hat{f}_m^j > \sum_{m \in \mathcal{T}^+} f_m^j, \tag{C.24}$$

since $j \in \mathcal{I} \backslash \mathcal{I}_l$, i.e., user j does not send any flow to links $m \geq l$. Note that:

$$\hat{\lambda}^j > \lambda^j, \tag{C.25}$$

since (C.24) implies that there exists some link $m' \in \mathcal{T}^+$, such that $\hat{f}_{m'}^j > f_{m'}^j$. Hence, by Lemma C.1, we have:

$$\hat{f}_m^j > f_m^j, \quad m \in \mathcal{T}^- \cap \mathcal{L}^j. \tag{C.26}$$

Since $\hat{f}_q^j = f_q^j = 0$, from (C.24) and (C.26) we have:

$$\hat{f}_1^j - f_1^j = \sum_{m \in \mathcal{T}^+ \cup \mathcal{T}^-} f_m^j - \sum_{m \in \mathcal{T}^+ \cup \mathcal{T}^-} \hat{f}_m^j < 0,$$

which together with $\hat{T}_1 \leq T_1$ (Lemma 3.9) implies that $\hat{\lambda}^j < \lambda^j$. But this is a contradiction to (C.25). Hence, it must be $l+1 \in \mathcal{T}^+$.

We have, thus, shown that for any link $l < q-1$, if $l \in \mathcal{T}^+$, then $l+1 \in \mathcal{T}^+$. Proceeding inductively, $n \in \mathcal{T}^+$ for any link $n \in \{l+1, \ldots, q-1\}$.

□

Remark: In the proof of the lemma, we assumed that there exists a link $l < q-1$, such that $l \in \mathcal{T}^+$ and $l+1 \in \mathcal{T}^-$, and arrived at a contradiction. Note, however, that the only implication of the assumption $l+1 \in \mathcal{T}^-$ that was used to arrive at the contradiction was $\sum_{i \in \mathcal{I}_{l+1}} \hat{\lambda}^i \leq \sum_{i \in \mathcal{I}_{l+1}} \lambda^i$. Thus, the same proof can be used to show that if $q-1 \in \mathcal{T}^+$, then $\sum_{i \in \mathcal{I}_q} \hat{\lambda}^i > \sum_{i \in \mathcal{I}_q} \lambda^i$.

Proof of Theorem 3.13

Note that $\hat{T}_q > T_q$ has been established in Lemma 3.9, that does not rely on the assumption $\hat{\mathcal{L}}^i = \mathcal{L}^i$, for all $i \in \mathcal{I}$, thus we only have to prove the remaining statements in the theorem.

We will construct inductively a sequence of capacity transfers $\{\delta_q(n)\}$ from link q to link 1, such that during each transfer $\delta_q(n)$ no user changes its set of links. Define $\delta_q(0) = 0$, $\mathbf{c}(0) = \mathbf{c}$ and $\Delta_q(1) = \Delta_q$. Let $\eta(1)$, $0 < \eta(1) \leq c_q(0) - c_q^0$, be as in Lemma A.2. If $\eta(1) \geq \Delta_q(1)$, we can transfer the entire capacity Δ_q to link 1 and terminate the process; the result is immediate from Lemma 3.12. If $\eta(1) < \Delta_q(1)$, we transfer capacity $\delta_q(1) = \eta(1)$ from link q to link 1, set $\Delta_q(2) = \Delta_q(1) - \delta_q(1)$ as the capacity that remains to be transferred and advance to the next step.

Proceeding inductively, consider the n-th step of the process, where the total capacity that has already been transferred to link 1 is $\sum_{k=1}^{n-1} \delta_q(k)$ and the resulting capacity configuration is $\mathbf{c}(n) = \mathbf{c}(0) + \sum_{k=1}^{n-1} \delta_q(k)(\mathbf{e}_1 - \mathbf{e}_q)$. Let $\eta(n)$ be as in Lemma A.2. If $\eta(n) \geq \Delta_q(n)$, we can transfer the entire remaining capacity $\Delta_q(n)$ and terminate the process. If, on the other hand, $\eta(n) < \Delta_q(n)$, we transfer capacity $\delta_q(n) = \eta(n)$ from link q to link 1, define $\Delta_q(n+1) = \Delta_q(n) - \delta_q(n)$ and proceed to step $n+1$. Since $\delta_q(n) \leq \eta(n)$, Lemma A.2 implies that no user changes its flow during the n-th step and by Lemma 3.12 we have:

$$\lambda^i(\mathbf{c}(n)) \leq \lambda^i(\mathbf{c}(n-1)), \quad i \in \mathcal{I}, \tag{C.27}$$

$$T_l(\mathbf{c}(n)) \leq T_l(\mathbf{c}(n-1)), \quad l \in \mathcal{L} \setminus \{q\}. \tag{C.28}$$

Suppose that the capacity transfer process terminates after a finite number of steps, i.e., that there exists some $n_0 \in I\!N$, such that $\sum_{k=1}^{n_0} \delta_q(k) = \Delta_q$. Then, $\hat{\mathbf{c}} = \mathbf{c}(n_0)$ and, since

(C.27)–(C.28) hold for all $1 \leq n \leq n_0$, we have:

$$\hat{\lambda}^i = \lambda^i(\mathbf{c}(n_0)) \leq \lambda^i(0) = \lambda^i, \quad i \in \mathcal{I},$$

$$\hat{T}_l = T_l(\mathbf{c}(n_0)) \leq T_l(0) = T_l, \quad l \in \mathcal{L} \setminus \{q\},$$

and the result follows.

Let us now consider the case where the transfer process does not terminate after a finite number of steps, i.e., $\sum_{k=1}^{\infty} \delta_q(k) \equiv \bar{\Delta}_q \leq \Delta_q$. Note that in this case $\delta_q(n) = \eta(n)$, for all $n \geq 1$. Hence, the convergence of the infinite series implies that $\eta(n) \to 0$, as $n \to \infty$. If $\bar{\Delta}_q = \Delta_q$, then $\hat{\mathbf{c}} = \lim_n \mathbf{c}(n)$. Replacing $\mathbf{c}(n-1)$ in the right-hand-side of each inequality in (C.27)–(C.28) with $\mathbf{c}(0) = \mathbf{c}$ and then taking the limit as $n \to \infty$ on the left-hand-side, the result follows, due to the continuity of the Nash mapping and the equilibrium link delays (Theorem A.1). If, on the other hand, $\bar{\Delta}_q < \Delta_q$, consider the capacity configuration $\tilde{\mathbf{c}} = \mathbf{c} + \bar{\Delta}_q(\mathbf{e}_1 - \mathbf{e}_q)$, that results by transferring (in a single step) capacity $\bar{\Delta}_q$ from link q to link 1. Since $\bar{\Delta}_q < \Delta_q \leq c_q - c_q^0$, we have $\tilde{c}_q = c_q - \bar{\Delta}_q > c_q^0$ and there is some remaining capacity that can be transferred from link q to link 1. Note that $\lim_n \eta(n) = 0$ implies that if we start from capacity allocation $\tilde{\mathbf{c}}$, any (nonzero) transfer of capacity will cause at least one user to change its set of links. This is a contradiction to Lemma A.2, hence we must have $\bar{\Delta}_q = \Delta_q$.

□

D. Single-Follower Stackelberg Routing Game

In this appendix we present the proof of Theorem 4.1 that gives the optimal strategy of the leader in the single-follower case. Let us first note that the global optimum $\mathbf{f}^*$ is, in fact, the Nash equilibrium of a degenerate single-user routing game, thus the analysis of Section 3.4 – Proposition 3.4 and Lemma 3.3 – can be readily applied to determine its explicit structure. In particular, the Lagrange multiplier λ^* associated with the global optimum is determined by (3.1)–(3.2), by replacing c_l^i by c_l and f_l^i by f_l^*. Also, there exists some link L^*, such that $\mathbf{f}^*$ routes flow only over the links in $\{1, \ldots, L^*\}$, and is determined by an expression similar to (3.6), where R replaces r^i and for every link l, G_l is defined as in eq. (3.3).

Before presenting the proof of the theorem, we need to establish that, for every link l, we have $H_l \leq H_{l+1}$. Using eq. (3.1) in eq. (4.2), we get:

$$H_l = \begin{cases} \dfrac{G_l}{\sqrt{\lambda^*}\sqrt{c_l}}, & l = 1, \ldots, L^* \\ R, & l = L^*+1, \ldots, L \end{cases}, \tag{D.1}$$

and the result is immediate, since $G_l \leq G_{l+1}$, and $c_l \geq c_{l+1}$, for all $l \in \mathcal{L}$.

Since $H_{L^*+1} = R \geq r^1$, (4.4) implies that $N \leq L^*$. Note that, in view of eq. (4.3), Theorem 4.1 implies that N is such that if the leader employs the strategy $\mathbf{f}^0$, then the follower sends its flow over the links in $\{1, \ldots, N\}$. Let us now proceed with the proof of the theorem.

Proof of Theorem 4.1

In order to prove optimality of $\mathbf{f}^0$, we have to show the following:

(i) The strategy $\mathbf{f}^0$ of the leader is an admissible flow configuration, i.e., that $f_l^0 \geq 0$ for all $l \in \mathcal{L}$ and that $\sum_{l=1}^{L} f_l^0 = r^0$.

(ii) The set of links over which the follower sends its flow is precisely $\{1, \ldots, N\}$.

(iii) For every link $l \in \mathcal{L}$: $f_l^0 + f_l^1 = f_l^*$.

We begin by establishing property (i) above. Since $H_{N+1} \geq r^1$, eq. (4.2) gives $\sum_{n=1}^{N} f_n^* \geq r^1$, therefore eq. (4.3) implies that $f_l^0 \geq 0$, for $l = 1, \ldots, N$. Nonnegativity of f_l^0 for $l = N+1, \ldots, L$ is immediate. Furthermore:

$$\sum_{l=1}^{L} f_l^0 = \sum_{l=1}^{N} c_l \frac{\sum_{n=1}^{N} f_n^* - r^1}{\sum_{n=1}^{N} c_n} + \sum_{l=N+1}^{L} f_l^* = \sum_{n=1}^{N} f_n^* - r^1 + \sum_{l=N+1}^{L} f_l^* = r^0,$$

since $\sum_{l=1}^{L} f_l^* = R = r^0 + r^1$. Thus $\mathbf{f}^0$ is an admissible routing strategy for the leader.

In the sequel we will establish property (ii) above. Let $c_l^1 = c_l - f_l^0$ be the residual capacity of link $l \in \mathcal{L}$ as seen by user 1. Let us first show that:

$$c_l^1 \geq c_{l+1}^1, \quad l = 1, \ldots, L-1. \tag{D.2}$$

For all links $l > N$, we have $c_l^1 = c_l - f_l^*$, since $f_l^0 = f_l^*$. According to Lemma 3.3, $c_l - f_l^* \geq c_{l+1} - f_{l+1}^*$, for all $l = 1, \ldots, L-1$. Thus, (D.2) holds for all $l = N+1, \ldots, L-1$.

Moreover, from eq. (4.3), we have:

$$c_l^1 = c_l \frac{\sum_{n=1}^N (c_n - f_n^*) + r^1}{\sum_{n=1}^N c_n}, \quad l = 1, \ldots, N, \tag{D.3}$$

therefore, (D.2) holds for $l = 1, \ldots, N-1$. Finally:

$$c_N^1 = c_N - f_N^0 \geq c_N - f_N^* \geq c_{N+1} - f_{N+1}^* = c_N - f_{N+1}^0 = c_{N+1}^1,$$

where the first inequality follows from $f_N^0 \leq f_N^*$, which can be easily derived from $H_N < r^1$. Thus, inequality (D.2) holds for all $l = 1, \ldots, L-1$. Hence, the best response $\mathbf{f}^1$ of the follower to $\mathbf{f}^0$ can be determined by means of Proposition 3.4. In particular, $f_l^1 \geq f_{l+1}^1$, and there exists a link $L^1 \leq L$, such that $f_l^1 > 0$ for $l \leq L^1$ and $f_l^1 = 0$ for $l > L^1$. The threshold L^1 is determined by equations (3.6) and (3.3).

We are now ready to prove the following:

Lemma D.1 *If the leader implements the strategy* $\mathbf{f}^0$ *described by eq. (4.3), then* $L^1 = N$, *i.e., the follower sends its flow precisely over the links in* $\{1, \ldots, N\}$.

Proof: Let start by showing that:

$$r^1 \leq G_{N+1}^1. \tag{D.4}$$

Using eqs. (D.3) and (3.3), this is equivalent to showing:

$$r^1 \leq \sum_{l=1}^N c_l \frac{\sum_{n=1}^N (c_n - f_n^*) + r^1}{\sum_{n=1}^N c_n} - \sqrt{c_{N+1} - f_{N+1}^*} \sum_{l=1}^N \sqrt{c_l} \frac{\sqrt{\sum_{n=1}^N (c_n - f_n^*) + r^1}}{\sqrt{\sum_{n=1}^N c_n}}$$

$$\Leftrightarrow \sqrt{c_{N+1} - f_{N+1}^*} \sum_{l=1}^N \sqrt{c_l} \frac{\sqrt{\sum_{n=1}^N (c_n - f_n^*) + r^1}}{\sqrt{\sum_{n=1}^N c_n}} \leq \sum_{n=1}^N (c_n - f_n^*)$$

$$\Leftrightarrow r^1 \leq \frac{1}{c_{N+1} - f_{N+1}^*} \left\{ \frac{\sum_{n=1}^N (c_n - f_n^*)}{\sum_{n=1}^N \sqrt{c_n}} \right\}^2 \sum_{n=1}^N c_n - \sum_{n=1}^N (c_n - f_n^*). \tag{D.5}$$

Since $N \leq L^*$, eq. (3.1) implies that:

$$\sqrt{\lambda^*} = \frac{\sum_{n=1}^N \sqrt{c_n}}{\sum_{n=1}^N (c_n - f_n^*)}, \tag{D.6}$$

therefore, (D.5) is equivalent to:

$$r^1 \leq \frac{1}{c_{N+1} - f_{N+1}^*} \frac{1}{\lambda^*} \sum_{n=1}^N c_n - \frac{\sum_{n=1}^N \sqrt{c_n}}{\sqrt{\lambda^*}}. \tag{D.7}$$

Suppose that $f^*_{N+1} > 0$. Then, eq. (3.1) gives $c_{N+1} - f^*_{N+1} = \sqrt{c_{N+1}/\lambda^*}$, and (D.7) is equivalent to:

$$r^1 \leq \frac{1}{\sqrt{\lambda^*}\sqrt{c_{N+1}}}\left\{\sum_{n=1}^{N} c_n - \sqrt{c_{N+1}}\sum_{n=1}^{N}\sqrt{c_n}\right\} = \frac{G_{N+1}}{\sqrt{\lambda^*}\sqrt{c_{N+1}}} = H_{N+1}, \tag{D.8}$$

which holds true in view of (4.4).

If $f^*_{N+1} = 0$, (3.2) implies $c_{N+1} - f^*_{N+1} = c_{N+1} \leq \sqrt{c_{N+1}/\lambda^*}$, and to prove (D.7), it suffices to show (D.8), which holds true. Thus, we have proven (D.4).

Let us now proceed to show that:

$$G^1_N < r^1. \tag{D.9}$$

Using eqs. (3.3) and (D.3), this is equivalent to showing:

$$r^1 > \left\{\sum_{n=1}^{N-1} c_n - \sqrt{c_N}\sum_{n=1}^{N-1}\sqrt{c_n}\right\}\frac{\sum_{n=1}^{N}(c_n - f^*_n) + r^1}{\sum_{n=1}^{N} c_n}$$

$$\Leftrightarrow r^1\sqrt{c_N}\sum_{n=1}^{N}\sqrt{c_n} > \left\{\sum_{n=1}^{N-1} c_n - \sqrt{c_N}\sum_{n=1}^{N-1}\sqrt{c_n}\right\}\sum_{n=1}^{N}(c_n - f^*_n).$$

Using eq. (D.6), this is equivalent to:

$$r^1 > \frac{1}{\sqrt{\lambda^*}\sqrt{c_N}}\left\{\sum_{n=1}^{N-1} c_n - \sqrt{c_N}\sum_{n=1}^{N-1}\sqrt{c_n}\right\} = \frac{G_N}{\sqrt{\lambda^*}\sqrt{c_N}} = H_N, \tag{D.10}$$

which holds true in view of (4.4). Therefore, we have proven that $G^1_N < r^1 \leq G^1_{N+1}$, which implies that $L^1 = N$ and the lemma follows.

□

In order to prove optimality of $\mathbf{f}^0$, it remains to be shown that for every link $l \in \mathcal{L}$, we have $f^1_l + f^0_l = f^*_l$. In view of eq. (4.3), we have to show that the best reply $\mathbf{f}^1$ of the follower to $\mathbf{f}^0$ is such that:

$$f^1_l = f^*_l - f^0_l = f^*_l - c_l\frac{\sum_{n=1}^{N} f^*_n - r^1}{\sum_{n=1}^{N} c_n}, \quad l = 1, \ldots, N. \tag{D.11}$$

It is easy to verify that $\sum_{l=1}^{N} f^1_l = r^1$, and $f^1_l \geq 0$ for all $l = 1, \ldots, N$. Therefore, it suffices to show that the strategy given by eq. (D.11) satisfies the optimality conditions (3.1)–(3.2) of the follower. In view of Lemma D.1, we only need to show (3.1), i.e., that:

$$\frac{c_l^1}{(c_l^1 - f_l^1)^2} = \frac{c_m^1}{(c_m^1 - f_m^1)^2}, \quad l, m \in \{1, \ldots, N\}. \tag{D.12}$$

Using eqs. (D.3) and (D.11), this is equivalent to showing:

$$\frac{c_l}{(c_l - f_l^*)^2} = \frac{c_m}{(c_m - f_m^*)^2}, \quad l, m \in \{1, \ldots, N\},$$

which holds due to the global optimality of $\mathbf{f}^*$ and $L^* \geq N$.

Thus, we have established that $\mathbf{f}^0$ is an optimal strategy for the leader, that achieves the global optimum. To conclude the proof of Theorem 4.1, we have to show the following:

Lemma D.2 *The optimal strategy of the leader in a single-follower Stackelberg routing game is unique.*

Proof: It suffices to show that any strategy $\mathbf{f}^0$ of the leader that enforces the network optimal flow configuration $\mathbf{f}^*$ has the structure described by eqs. (4.3) and (4.4). Let $\mathbf{f}^0$ be such a strategy and $\mathbf{f}^1$ the best reply of the follower. Then:

$$f_l^0 + f_l^1 = f_l^*, \quad l \in \mathcal{L}. \tag{D.13}$$

Let us first show that for every link l, $f_l^1 \geq f_{l+1}^1$. Assume by contradiction that, for some l, $0 \leq f_l^1 < f_{l+1}^1$. Then, the optimality conditions (2.4)–(2.5) for $\mathbf{f}^1$ imply that:

$$\frac{1}{c_{l+1} - f_{l+1}^*} + \frac{f_{l+1}^1}{(c_{l+1} - f_{l+1}^*)^2} \leq \frac{1}{c_l - f_l^*},$$

which is a contradiction, since $c_l - f_l^* \geq c_{l+1} - f_{l+1}^*$ and $f_{l+1}^1 > 0$. Therefore, there exists some link L^1, such that $f_l^1 > 0$ for $l \leq L^1$ and $f_l^1 = 0$ for $l > L^1$. Then, it is easy to see that the optimality conditions for $\mathbf{f}^1$ imply (D.2), i.e., the residual link capacities as seen by user 1 preserve the order of the link capacities themselves. Hence, the threshold L^1 is determined by (3.6). Evidently, $L^1 \leq L^*$, otherwise eq. (D.13) would be violated.

The optimality conditions for $\mathbf{f}^1$ and $\mathbf{f}^*$ imply:

$$\frac{c_l^1}{c_m^1} = \frac{c_l}{c_m} = \left\{ \frac{c_l - f_l^*}{c_m - f_m^*} \right\}^2, \quad l, m \in \{1, \ldots, L^1\},$$

and taking $m = 1$, this gives:

$$f_l^1 = f_l^* - \frac{c_l}{c_1}(f_1^* - f_1^1), \quad l = 1, \ldots, L^1, \tag{D.14}$$

which, together with $\sum_{l \in \mathcal{L}} f_l^1 = r^1$, gives:

$$f_l^1 = f_l^* - c_l \frac{\sum_{n=1}^{L^1} f_n^* - r^1}{\sum_{n=1}^{L^1} c_n}, \quad l = 1, \ldots, L^1, \tag{D.15}$$

and in view of eq. (D.13), $\mathbf{f}^0$ is given by an expression similar to eq. (4.3), with N replaced by L^1. Since N is uniquely determined by (4.4), in order to prove the lemma it suffices to show that $L^1 = N$, i.e., that:

$$H_{L^1} < r^1 \leq H_{L^1+1}. \tag{D.16}$$

Following precisely the proof of Lemma D.1, with N replaced by L^1, one can see that (D.16) holds true and this concludes the proof.

□

References

[ALT93] Eitan Altman and Nahum Shimkin, "Worst-Case and Nash Routing Policies in Parallel Queues with Uncertain Service Allocations," IMA Preprint Series 1120, Institute for Mathematics and Applications, University of Minnesota, Minneapolis, USA, 1993.

[ALT94] Eitan Altman, "Flow Control Using The Theory of Zero-Sum Markov Games," *IEEE Transactions on Automatic Control*, 1994. To appear.

[COC93] R. Cocchi, D. Estrin, S. Shenker, and L. Zhang, "Pricing in Computer Networks: Motivation, Formulation and Example," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 614–627, December 1993.

[COH90] J. E. Cohen and F. P. Kelly, "A Paradox of Congestion in a Queuing Network," *Journal of Applied Probability*, vol. 27, pp. 730–734, 1990.

[DOU89] Christos Douligeris and Ravi Mazumdar, "Multilevel Flow Control of Queues," in *Proceedings of the Johns Hopkins Conference on Information Sciences*, Baltimore, MD, March 1989.

[DUB86] Pradeep Dubey, "Inefficiency of Nash Equilibria," *Mathematics of Operations Research*, vol. 11, pp. 1–8, February 1986.

[ECO91] A. A. Economides and J. A. Silvester, "Multi-Objective Routing in Integrated Services Networks: A Game Theory Approach," in *Proceedings of INFOCOM'91*, pp. 1220–1225, 1991.

[FUD92] Drew Fudenberg and Jean Tirole, *Game Theory*. Cambridge, Massachusetts: MIT Press, 1992.

[HSIA91] Man-Tung T. Hsiao and Aurel A. Lazar, "Optimal Decentralized Flow Control of Markovian Queueing Networks with Multiple Controllers," *Performance Evaluation*, vol. 13, no. 3, pp. 181–204, 1991.

[KOR93] Yannis A. Korilis and Aurel A. Lazar, "On the Existence of Equilibria in Noncooperative Optimal Flow Control," in *Proceedings of the ITC Sponsored Seminar*, pp. 243–251, Bangalore, India, November 1993. Indian Institute of Science. Under revision for the *Journal of the ACM*.

[KOR94] Yannis A. Korilis, Aurel A. Lazar, and Ariel Orda, "Noncooperative Networks: The Designer's Perspective," CTR Technical Report CU/CTR/TR 372-94-19, Center for Telecommunications Research, Columbia University, 1994.

[KOR94b] Yannis A. Korilis, Aurel A. Lazar, and Ariel Orda, "Stackelberg Routing Strategies," CTR Technical Report CU/CTR/TR 384-94-31, Center for Telecommunications Research, Columbia University, 1994.

[LUE84] David G. Luenberger, *Linear and Nonlinear Programming.* Addison Wesley, 2nd edition, 1984.

[MAZ91] Ravi Mazumdar, Lorne Mason, and Christos Douligeris, "Fairness in Network Optimal Flow Control: Optimality of Product Forms," *IEEE Transactions on Communications*, vol. 39, pp. 775–781, May 1991.

[MYE91] Roger B. Myerson, *Game Theory: Analysis of Conflict.* Cambridge, Massachusetts: Harvard University Press, 1991.

[ORD93] Ariel Orda, Raphael Rom, and Nahum Shimkin, "Competitive Routing in Multiuser Communication Networks," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 510–521, October 1993.

[SHE94] Scott Shenker, "Making Greed Work in Networks: A Game-Theoretic Analysis of Switch Service Disciplines," in *Proceedings of the ACM SIGCOM '94*, September 1994.

[ZAN81] W. I. Zangwill and C. B. Garcia, *Pathways to Solutions, Fixed Points and Equilibria.* Englewood Cliffs, NJ: Prentice-Hall, Inc., 1981.

[ZHA92] Zhensheng Zhang and Christos Douligeris, "Convergence of Synchronous and Asynchronous Greedy Algorithms in a Multiclass Telecommunications Environment," *IEEE Transactions on Communications*, vol. 40, pp. 1277–1281, August 1992.