Pricing Network Edges for Heterogeneous Selfish Users

Richard Cole* Yevgeniy Dodis† Tim Roughgarden‡

November 6, 2002

Abstract

We study the negative consequences of selfish behavior in a congested network and economic means of influencing such behavior. We consider the model of selfish routing defined by Wardrop [30] and studied in a computer science context by Roughgarden and Tardos [26]. In this model, the latency experienced by network traffic on an edge of the network is a function of the edge congestion, and network users are assumed to selfishly route traffic on minimum-latency paths. The quality of a routing of traffic is measured by the sum of travel times (the *total latency*).

It is well known that the outcome of selfish routing (a *Nash equilibrium*) does not minimize the total latency and can be improved upon with coordination. An ancient strategy for improving the selfish solution is the principle of *marginal cost pricing*, which asserts that on each edge of the network, each network user on the edge should pay a tax offsetting the congestion effects caused by its presence. By pricing network edges according to this principle, the inefficiency of selfish routing can always be eradicated.

This result, while fundamental, assumes a very strong *homogeneity* property: all network users are assumed to trade off time and money in an identical way. The guarantee also ignores both the algorithmic aspects of edge pricing and the unfortunate possibility that an efficient routing of traffic might only be achieved with exorbitant taxes. Motivated by these shortcomings, we extend this classical work on edge pricing in several different directions and prove the following results.

- We prove that the edges of a single-commodity network can always be priced so that an optimal routing of traffic arises as a Nash equilibrium, even for very general heterogeneous populations of network users.
- When there are only finitely many different types of network users and all edge latency functions are convex, we show how to compute such edge prices efficiently.
- We prove that an easy-to-check mathematical condition on the population of heterogeneous network users is both necessary and sufficient for the existence of edge prices that induce an optimal routing while requiring only moderate taxes, and mention consequences.

*Department of Computer Science, New York University, 251 Mercer Street, New York, NY 10012. Supported by NSF grant CCR0105678. Email: `cole@cs.nyu.edu`.

†Department of Computer Science, New York University, 251 Mercer Street, New York, NY 10012. Supported by an NSF CAREER Award. Email: `dodis@cs.nyu.edu`.

‡Department of Computer Science, Cornell University, Ithaca NY 14853. Supported by ONR grant N00014-98-1-0589. This work was done in part while visiting NYU. Email: `timr@cs.cornell.edu`.

1 Introduction

Selfish Routing and Edge Pricing. We study the negative consequences of selfish behavior in networks and economic means of influencing such behavior. We focus on a simple model of *selfish routing*, defined by Wardrop [30] and first studied from a computer science perspective by Roughgarden and Tardos [26]. In this model, we are given a directed network in which each edge possesses a latency function, describing the common delay experienced by all traffic on the edge as a function of the edge congestion. There is a fixed amount of traffic wishing to travel from a source vertex s to a sink vertex t; as in earlier works, we assume that the traffic comprises a very large population of users, so that the actions of a single individual have negligible effect on network congestion. We measure the quality of an assignment of traffic to s-t paths by the resulting sum of all travel times—the *total latency*. We assume that each network user, when left to its own devices, acts selfishly and routes itself on a minimum-latency path, given the network congestion due to the other users. In general such a "selfishly motivated" assignment of traffic to paths (a *Nash equilibrium*) does not minimize the total latency; put differently, the outcome of selfish behavior can be improved upon with coordination.

The inefficiency of selfish routing (and of Nash equilibria more generally) motivates strategies for *coping with selfishness*—methods for ensuring that noncooperative behavior results in a socially desirable outcome. For selfish routing, an ancient strategy—discussed informally as early as 1920 [22]—is *marginal cost pricing*. The principle of marginal cost pricing asserts that on each edge, each network user on the edge should pay a tax equal to the additional delay its presence causes for the other users on the edge. Several decades later, researchers showed that this principle leads to the following rigorous guarantee [4]: assuming all network users choose routes to minimize the sum of the latency experienced and taxes paid, it is possible to levy a tax on each network edge so that the resulting Nash equilibrium achieves the minimum-possible total latency. Briefly, the inefficiency of selfish routing can always be eradicated by pricing network edges appropriately.

This guarantee, while fundamental, is unsatisfying in several respects. First, it assumes a very strong *homogeneity* property: even though the model assumes a very large number of network users, all users are assumed to trade off time and money in an identical way. How should edges be priced with heterogeneous network users? Second, the guarantee ignores the *algorithmic* aspect of edge pricing: how can edge prices be efficiently computed? When many different sets of edge prices induce a minimum-latency Nash equilibrium, can we efficiently compute the "best" one? Finally, this result is single-minded in its pursuit of a routing minimizing the total latency, and ignores the possibility that exorbitant taxes may be required to achieve such a routing. When can the edges of a network be priced so that an efficient routing results and, in addition, the disutility of network users due to taxes is small?

Our Results. In this paper, we address these three issues and prove several results about pricing networks with heterogeneous users. We model heterogeneity in the following simple way: we associate to each network user a a scalar $\alpha(a)$ and assume that a seeks to minimize the latency it experiences plus $\alpha(a)$ times the money it is required to pay. In the classical setup, $\alpha(a) = 1$ for all users a. Introducing the function α allows the full spectrum of time vs. money tradeoffs, with a small value of $\alpha(a)$ corresponding to a large sensitivity to time, and a large value indicating sensitivity to money. In this model, we prove the following.

- We show that for an arbitrary (single-commodity) network and a heterogeneous community of traffic in which all network users have some sensitivity to taxes ($\alpha(a) > 0$ for all a), network edges can be priced so that a minimum-latency routing of traffic arises as a Nash equilibrium. Thus, even with heterogeneous traffic, optimally chosen edge taxes are sufficiently powerful to eliminate all of the inefficiency of selfish routing. We remark that this result does *not* follow from the principle of marginal cost pricing and is instead proved with a nonconstructive fixed-point theorem.

- Under the additional assumptions that the function α takes on only finitely many distinct values and that all network latency functions are convex, we show how to compute a set of optimal edge prices. Such taxes need not be unique, and we prove that the set of all optimal taxes can be explicitly described by a small set of linear inequalities; this description permits optimization of a secondary linear objective function, such as minimizing the taxes paid by network users.

- We give a precise mathematical characterization of the functions α for which a minimum-latency routing of traffic can always be induced with small disutility to network users due to taxes. As a consequence, we find that in many settings, including the classical model where $\alpha(a) = 1$ for all a, inducing the minimum-latency routing can require extremely costly taxes.

Related Work. The problems of analyzing and influencing selfish behavior have been extensively studied by many different communities; even restricting attention to the computer science literature, there are far too many works to enumerate here. In what follows, we will discuss only the results on pricing and on selfish routing that are closest to the present work.

As discussed above, the model of selfish routing studied here was first defined by Wardrop [30] and has been extensively studied ever since (see [25] for many more references). Beckmann et al. [4] showed that marginal cost taxes produce a minimum-latency routing of traffic when $\alpha(a) = 1$ for all users a. This is accomplished by a simple application of the Karush-Kuhn-Tucker theorem (see e.g. [21]) to an appropriate convex program. Marginal cost pricing has also been studied with heterogeneous network users; unfortunately, extending the techniques of [4] to this more general setting requires users with different α-values to pay *different taxes* on the same edge [10, 28]. This solution is obviously unsatisfying as the number of different user types grows large (indeed, we will in general consider functions α taking on infinitely many different values); moreover, this scheme requires knowledge of a user's type when it traverses an edge, to charge it an appropriate tax. By contrast, we are seeking a solution with a *single* tax placed on each edge; as we will see, implementing this more ambitious goal requires a mathematical approach quite different from that of marginal cost pricing.

The problem of optimally pricing shared resources has also motivated a large body of work in both the welfare economics literature (see e.g. [19]) and by the networking community (see e.g. [17]). While work in the networking literature has not considered our model of selfish routing, the scope of previous work in economics is sufficiently broad to encompass the heterogeneous traffic model considered here. However, the general techniques of welfare economics for optimizing prices provide no method to ensure that such prices can induce a selfish solution that is as socially desirable as the best cooperative outcome. By contrast, our central existence theorem asserts that optimal taxes always suffice to induce the best coordinated routing of traffic.

Researchers have also studied how to compute optimal taxes. Efficiently computing marginal cost prices in the classical setup of homogeneous network traffic turns out to be straightforward (via convex programming) [4], but more general models and goals have also been studied. The emphasis of work in the economics literature has unfortunately been on general-purpose algorithms that need not run in polynomial time [19], even when specialized to the edge pricing problems considered in this paper. The transportation science community has recently made significant progress in efficiently computing and characterizing taxes that induce a minimum-latency routing of traffic [5, 13, 14, 16]. This work considers only the case of homogeneous traffic, but forms the basis of the algorithmic results in this paper.

The disutility to network users of edge pricing has apparently been largely ignored in previous work. Our approach is similar in spirit to work of Archer and Tardos [2, 3] on "frugal mechanisms"—mechanisms (such as auctions) that solve an optimization problem in an incentive-compatible way and also make use of only moderate incentives. In the context of selfish routing, in previous work [8] we studied the optimization problem of computing taxes that induce a Nash equilibrium with minimum-possible user disutility (accounting for disutility due to both latency and taxes), for the case of homogeneous traffic. In [8], we showed that this optimization problem essentially reduces to the network design problem of avoiding Braess's Paradox [23] and is hard to approximate to within any reasonable factor.

Organization. Section 2 formally introduces our model and states some preliminaries. In Section 3 we show that edge taxes always suffice to induce a minimum-latency routing of traffic, even when network users are heterogeneous. In Section 4 we show how to compute such taxes in networks with convex latency functions when there are only finitely many different types of users. In Section 5 we study the disutility caused by these taxes, and characterize the user distributions for which this disutility is small.

2 Preliminaries

Congested Networks and Flows. We consider a directed graph $G = (V, E)$ with source s and sink t. We denote the set of simple s-t paths in G by $\mathcal{P}$, which we assume is nonempty. We allow parallel edges but have no use for self-loops. There is one unit of traffic wishing to travel from s and t, modeled as the unit interval $[0, 1]$ endowed with Lebesgue measure λ.[1] Each point $a \in [0, 1]$ will be called an *agent*, and is thought of as a noncooperative and infinitesimal unit of traffic.

By a *flow*, we mean a Lebesgue-measurable function $f : [0, 1] \to \mathcal{P}$ describing who goes where. There are two ways to ignore some of the information provided by a flow to recover more familiar combinatorial objects. A flow naturally induces a *flow on paths*, which we define to be the vector $\{f_P\}_{P\in\mathcal{P}}$ indexed by s-t paths, with $f_P = \lambda(\{a \in [0, 1] : f(a) = P\})$ the amount of traffic assigned to the path P by f. A flow on paths then induces a *flow on edges*, defined as a vector $\{f_e\}_{e\in E}$ on edges with $f_e = \sum_{P:e\in P} f_P$ the amount of traffic using edge e en route from s to t. A flow on edges may correspond to many different flows on paths, and a flow on paths corresponds to many different flows.

The network G suffers from congestion effects; to model this, we assume each edge e possesses a nonnegative, continuous, nondecreasing *latency function* ℓ_e that describes the delay incurred by traffic on e as a function of the edge congestion f_e. The latency of a path P in G with respect to a flow f is then given by $\ell_P(f) = \sum_{e\in P} \ell_e(f_e)$. We measure the quality of a flow by its *total latency* $L(f)$, defined by $L(f) = \sum_{P\in\mathcal{P}} \ell_P(f)f_P$ or, equivalently, by $L(f) = \sum_{e\in E} \ell_e(f_e)f_e$. Evidently, any two flows inducing the same flow on edges have equal total latency. We will call a flow minimizing $L(\cdot)$ *optimal* or *minimum-latency*; such a flow always exists, for the set of flows on edges is a compact set and $L(\cdot)$ is continuous.

Finally, we will allow a set of nonnegative *taxes* $\{\tau_e\}_{e\in E}$ to be placed on the edges of a network G. We denote the resulting network with taxes by G^τ. We will call a triple of the form (G, ℓ, α) or (G^τ, ℓ, α) an *instance*.

Nash equilibria. We now discuss how agents react to edge pricing. We assume that agent a has a money/time valuation ratio of $\alpha(a)$. Thus, if a set τ of taxes are placed on the edges of a network, agent a seeks a shortest s-t path relative to edge lengths $\ell_e(f_e) + \alpha(a)\tau_e$. We will assume that agents are sorted in order of money-sensitivity, so that $\alpha : [0, 1] \to [0, \infty]$ in a nondecreasing function. We call α a *distribution function*. We do not want to assume that distribution functions are bounded, and therefore permit functions α with $\alpha(1) = +\infty$; however, we will always assume that α is finite on $[0, 1)$.

We assume that noncooperative behavior results in a Nash equilibrium—a "stable point" in which no agent has an incentive to unilaterally alter its strategy (i.e., its route from s to t). To make this precise, let $c^a_P(f, \tau) = \ell_P(f) + \alpha(a)\tau_P$ denote agent a's evaluation of path P relative to taxes τ and latencies with respect to the flow f, where $\tau_P = \sum_{e\in P} \tau_e$.

[1] Allowing an arbitrary rate $r > 0$ of traffic requires only cosmetic changes to this paper.

Definition 2.1 A flow $f : [0,1] \to \mathcal{P}$ is *at Nash equilibrium* or is a *Nash flow* for instance (G^{τ}, ℓ, α) if for every agent $a \in [0,1]$ and path $P \in \mathcal{P}$,

$$c^a_{f(a)}(f, \tau) \leq c^a_P(f, \tau). \tag{1}$$

If (1) holds for agent a, we will say that a is routed on a c^a-minimum path (with the dependence on f and τ understood). Existence of Nash flows follows from, for example, the quite general results of Schmeidler [27, Thm 2].

Proposition 2.2 *Any instance (G^{τ}, ℓ, α) admits a flow at Nash equilibrium.*

Proposition 2.2 is reassuring, but gives us little insight into how Nash flows differ from arbitrary measurable functions on $[0,1]$. We intuitively expect Nash flows to exhibit the following nice structure. Agents $a \in [0,1]$ with a near 0, who value time far more than money, should be routed on a path with very small latency but possibly very large tax; agents slightly farther away from 0 should route themselves on a path with somewhat larger latency and somewhat smaller tax; at the other extreme, agents near 1—the biggest misers of all—should be routed on a path that may have large latency but possesses near-zero taxes. Motivated by this discussion, we make a definition.

Definition 2.3 A flow f at Nash equilibrium in (G^{τ}, ℓ, α) is *canonical* if:

(a) For any path $P \in \mathcal{P}$, the agents assigned to P by f form a (possibly empty or degenerate) subinterval of $[0,1]$.

(b) If $a_1 \leq a_2$, then $\ell_{f(a_1)}(f) \leq \ell_{f(a_2)}(f)$.

(c) If $a_1 \leq a_2$, then $\tau_{f(a_1)}(f) \geq \tau_{f(a_2)}(f)$.

Thus a canonical Nash flow f splits $[0,1]$ into a finite number of subintervals (at most one per s-t path), inducing an ordering on the paths to which f assigns traffic that is nondecreasing in latency and nonincreasing in taxes.

We have the following strengthening of Proposition 2.2.

Proposition 2.4 *Every instance (G^{τ}, ℓ, α) admits a canonical flow at Nash equilibrium.*

The proof of Proposition 2.4 is essentially just a "rearrangement" argument and shows the stronger statement that an arbitrary Nash flow can be reorganized into a canonical one without changing its cost or the induced flow on paths. The details are somewhat tedious and are omitted from this abstract.

The uniqueness of Nash flows is not so universal. However, we have the following reasonably strong uniqueness result when α takes on only finitely many values (i.e., when α is a step function).

Proposition 2.5 *If α takes on only finitely many values and f^1, f^2 are flows at Nash equilibrium for (G^{τ}, ℓ, α), then $\ell_e(f^1_e) = \ell_e(f^2_e)$ for all edges e. If in addition all latency functions are strictly increasing, then f^1 and f^2 induce identical flows on edges.*

We prove Proposition 2.5 using techniques from convex programming. The approach is similar to that of previous works [12, 25] and is therefore omitted from this abstract. Proposition 2.5 is also known to hold for general distribution functions α when G is a set of parallel links [18, Prop 3.3]; we suspect that it holds for general α and general networks G, but have been unable to verify this.

Example 2.6 Consider a two-node network with two links, 1 and 2. If the links have latency functions $\ell_1(x) = x$ and $\ell_2(x) = 1$, then in the absence of taxes a flow at Nash equilibrium assigns all but a measure-zero set of agents to the first link; in a canonical Nash flow, all agents, except possibly for agent 1, are assigned to the first link. By contrast, the flow on edges induced by any optimal flow $\hat{f}$ is $\hat{f}_1 = \hat{f}_2 = \frac{1}{2}$. By placing a tax on the first link, we can discourage selfish agents from using it and alter the set of flows at Nash equilibrium. Ideally, we seek a tax, that will be a function of the distribution α of agent preferences, that deters precisely half of the traffic (the half with greatest sensitivity to taxes) and therefore induces a minimum-latency routing. In the next two sections, we will study how to find such taxes.

3 Existence of Optimal Taxes

In this section we show that carefully chosen edge prices can eradicate all inefficiency due to selfish routing, even in the presence of agent heterogeneity. More precisely, for an instance (G, ℓ, α), call a set τ of taxes *optimal* if there is a minimum-latency flow $\hat{f}$ and a flow f^τ at Nash equilibrium for (G^τ, ℓ, α) so that $\hat{f}$ and f^τ induce identical flows on edges; it follows that f^τ is also a minimum-latency flow. The classical guarantee of marginal cost pricing states that, when α is everywhere equal to 1 and all latency functions are differentiable, the taxes $\tau_e = \hat{f}_e \cdot \ell'_e(\hat{f}_e)$ for all edges e are optimal (where ℓ'_e denotes the derivative of ℓ_e) [4]. Here, we show that every instance (G, ℓ, α) admits an optimal tax, assuming only that all traffic has nonzero sensitivity to edge taxes (i.e., that $\alpha(a) > 0$ for all agents a). Before proving this general result, we consider a seemingly very special type of instance.

3.1 Well-Behaved Instances

In this subsection, we will consider an instance (G, ℓ, α) that satisfies four assumptions, and will call such an instance *well behaved*.

(A1) G is a directed acyclic graph admitting a minimum-latency flow $\hat{f}$ with $\hat{f}_e > 0$ for all edges e.

(A2) Latency functions are "uniformly strictly increasing" in the sense that for some $\delta > 0$, $\ell_e(y) - \ell_e(x) \geq \delta(y - x)$ for any edge e and any $y \geq x \geq 0$.

(A3) The distribution function α is a step function, taking on only finitely many values.

(A4) The distribution function α satisfies $\alpha(0) > 0$; thus all agents have some sensitivity to taxes.

We will remove the first three assumptions in the next subsection. A little reflection shows that if a distribution function fails to satisfy the assumption that $\alpha(a) > 0$ whenever $a > 0$, an assumption only slightly weaker than (A4), then there are instances (even restricting to two-node, two-link networks) in which optimal taxes do not exist.

We next prove that well-behaved instances admit optimal taxes. Our proof will proceed in two parts. The first part constructs a continuous "update map", which takes sets of taxes to intuitively "better" sets of taxes. We will apply Brouwer's fixed point theorem to show that this update map has at least one fixed point. This type of argument is quite standard in economics and game theory; in particular, Nash's theorem on the existence of Nash equilibria in non-zero sum finite matrix games can be proved in this way [20]. To finish the proof, we must show that fixed points of the update map are the desired sets of optimal taxes. Unlike many fixed-point arguments in game theory (including that in the proof of Nash's theorem [20]), this fact is not at all obvious in our application; to prove it, we will make detailed use of the combinatorial structure of the underlying network.

Let (G, ℓ, α) be a well-behaved instance, and $\hat{f}$ a minimum latency flow with $\hat{f}_e > 0$ for all edges e (assumption (A1)). For motivation, consider a tax τ that need not be optimal. Since the instance is well behaved, Proposition 2.5 implies that all flows at Nash equilibrium for (G^τ, ℓ, α) induce the same flow on edges, $\{f_e^\tau\}_{e \in E}$. If $f_e^\tau = \hat{f}_e$ for all edges e, then τ is an optimal tax; if not, we wish to modify τ in a way that steers the Nash flow toward $\hat{f}$. A natural heuristic is to increase the tax on edges with $f_e^\tau > \hat{f}_e$ and decrease the tax on edges with $f_e^\tau < \hat{f}_e$. To apply a fixed-point theorem, we must also ensure that our candidate tax vectors remain bounded throughout an arbitrary number of iterations of our update map. Based on this discussion, we define our update map Γ as follows, where T is a large upper bound on the maximum-allowable tax, to be chosen later:

$$\Gamma(\tau)_e = \min\left\{T, \max\left\{0, \left(\tau_e + \frac{f_e^\tau}{\hat{f}_e} - 1\right)\right\}\right\}.$$

The map Γ is well defined because (G, ℓ, α) is well behaved; in particular, f_e^τ is uniquely defined by τ for each edge e, and $\hat{f}_e > 0$ for all edges e.

Proposition 3.1 *The map $\Gamma : [0,T]^E \to [0,T]^E$ is continuous.*

Proof Sketch. It suffices to prove that the map $\tau \mapsto \{f_e^\tau\}_{e \in E}$ is continuous. Under assumptions (A2) and (A3), this follows from a theorem of Dafermos and Nagurney [11, Thm 3.1] after applying a straightforward reduction of Dafermos [9, §3]. ■

By Brouwer's fixed-point theorem [6, §6], we have the following corollary.

Corollary 3.2 *The map $\Gamma : [0,T]^E \to [0,T]^E$ has at least one fixed point.*

It remains to show that a fixed point of Γ is an optimal tax for the well-behaved instance (G, ℓ, α). Toward this end, let τ be a fixed point of Γ. By the definition of Γ, we can classify the edges of G into one of three types with respect to τ, as follows: *good*, if $f_e^\tau = \hat{f}_e$; *oversaturated*, if $f_e^\tau > \hat{f}_e$ and $\tau_e = T$; or *undersaturated*, if $f_e^\tau < \hat{f}_e$ and $\tau_e = 0$. We will call τ a *good* fixed point if all edges are good w.r.t. τ, and a *bad* fixed point otherwise. Writing ℓ_{max} for $\max_{e \in E} \ell_e(1)$ and n for the number of vertices of G, our key lemma is the following.

Lemma 3.3 *If $T \geq 3n^3 \ell_{max}/\alpha(0)$, then Γ has no bad fixed points.*

The main theorem of this subsection follows immediately from Lemma 3.3 and the definitions.

Theorem 3.4 *A well-behaved instance admits an optimal set of taxes.*

We now discuss Lemma 3.3 and its proof, the details of which can be found in Appendix A. The lemma should be intuitively plausible; in any bad fixed point, some traffic inexplicably uses heavily taxed edges instead of tax-free, undersaturated alternatives. On the other hand, these oversaturated and undersaturated edges can be scattered throughout the network in an arbitrary way; for this reason, making this intuitive argument precise takes some work.

The proof of Lemma 3.3 hinges on defining an appropriate ordering on the vertices of G; this idea was also used, for different purposes, in [23]. To describe the properties we desire of this ordering, let f^τ be a Nash flow for the well-behaved instance (G^τ, ℓ, α) and, for a vertex v, let $d(v)$ denote the length of a shortest s-v path, using τ as edge lengths. The proof approach of Lemma 3.3 (see Appendix A) motivates an ordering satisfying:

(P1) if $f_e^\tau > 0$ with $e = (v, w)$, then w follows v in the ordering

(P2) d-values of vertices are nondecreasing in the ordering.

Since G and hence f^τ are acyclic, these two properties would be jointly attainable if all agents chose s-t paths with minimum-possible tax (see [23]); because of the heterogeneity of agent objective functions, however, (P1) and (P2) can be mutually exclusive goals. Three lemmas of Appendix A show that, on the other hand, it is possible to achieve (P1) while *approximately* satisfying (P2). This weaker statement turns out to be sufficient to prove Lemma 3.3. Intuitively, this weaker result is possible since for sufficiently large taxes all agents are, to first order, choosing minimum-tax paths; edge latencies are only a second-order effect. With all agent objective functions approximately equal, we can define an ordering satisfying (P1) and a relaxed version of (P2).

3.2 General Instances

In this subsection, we show that any instance (G, ℓ, α) with $\alpha(0) > 0$ admits a set of optimal taxes. To do this, we remove assumptions (A1)–(A3) of the previous subsection one by one.

The assumption that G is directed acyclic and that (G, ℓ, α) admits an optimal flow $\hat{f}$ with $\hat{f}_e > 0$ on every edge is easy to remove; without loss of generality, (G, ℓ, α) admits an acyclic optimal flow and edges with $\hat{f}_e = 0$ can be effectively "deleted" with a sufficiently large tax. Details are given in Appendix A.

Lemma 3.5 *An instance satisfying assumptions (A2)–(A4) admits an optimal set of taxes.*

We next remove assumption (A2), that latency functions are uniformly strictly increasing. The proof approach is to approximate more general latency functions with uniformly strictly increasing ones, and conclude with a limiting argument. Again, details are provided in Appendix A.

Lemma 3.6 *An instance satisfying assumptions (A3)–(A4) admits an optimal set of taxes.*

For our most general result, we approximate arbitrary distribution functions with step functions, and apply a limiting argument. The proof is similar to that of Lemma 3.6, and we omit it from this abstract.

Theorem 3.7 *If instance (G, ℓ, α) satisfies $\alpha(0) > 0$, then it admits an optimal set of taxes.*

Remark 3.8 A tax τ is by definition optimal for (G, ℓ, α) if *some* flow at Nash equilibrium for (G^τ, ℓ, α) induces the same flow on edges as an optimal flow and is therefore minimum-latency. The uniqueness assertion of Proposition 2.5 implies that, so long as the distribution function α takes on only finitely many values, *all* Nash flows induced by an optimal tax are minimum-latency. We know of no analogous uniqueness result for more general distribution functions (although we suspect there is one to be discovered); it is therefore conceivable that in an instance with a general distribution function, some Nash flows induced by an optimal tax fail to be minimum-latency.

4 Computing Optimal Taxes

The existence results of the previous section are highly nonconstructive. Theorem 3.4, asserting the existence of optimal taxes for well-behaved instances, relies on Brouwer's fixed point theorem, and no efficient algorithms for finding Brouwer fixed points are known (see e.g. [15]). The extensions to Theorem 3.4 given in Subsection 3.2 require further nonconstructive arguments (e.g., extracting a convergent subsequence from a sequence in a compact set). In this section, we use the methodology of Bergendorff et al. [5] (who only considered agents with identical objective functions) and show how to efficiently compute optimal taxes for an instance with finitely many different types of agents and convex edge latency functions.[2] In fact, we will

[2]Throughout this section, we assume some reasonable encoding of latency and distribution functions; see [1, §14.1] or [25, §5.2] for typical approaches. We assume that a distribution function that is a step function with k steps requires $\Omega(k)$ input bits to describe.

show this in a very strong way: we will prove that the set of optimal taxes can be explicitly described by a polynomial-size set of linear inequalities. Thus an optimal tax can not only be efficiently found, but in fact the optimal tax that optimizes some secondary linear objective function, such as minimizing the taxes paid by network users, can also be computed efficiently. These constructive results complement and do not subsume the existence theorems of Section 3, even in the special case of finitely many distinct agent types; on the contrary, these existence results provide the sole assurance that our linear description of the set of optimal taxes describes a non-empty set!

It can be shown that the existence of an efficient algorithm to compute an optimal tax or an optimal flow in a network with arbitrary continuous, nondecreasing latency functions implies P=NP (proof deferred to the full version). To circumvent this difficulty, in this section we assume that all network latency functions are convex. Under this assumption, the total latency $L(\cdot)$ is a convex function and an optimal flow can be computed efficiently using convex programming.[3] This assumption is satisfied by most (but not all) latency functions of interest, including polynomials with nonnegative coefficients and popular queueing delay functions; see [24] for some specific examples.

We now proceed to our algorithmic results. For an instance (G, ℓ, α) in which α takes on a finite number of values $\alpha_1 < \cdots < \alpha_k$, we will call the agents a with $\alpha(a) = \alpha_i$ the *ith commodity*. If f is a flow for such an instance, we will write $f_P^{(i)}$ for the number (i.e., Lebesgue measure) of agents of commodity i assigned to the path P, and similarly $f_e^{(i)}$ for the number of agents of commodity i that use edge e in their assigned routes. We begin with a reformulation of the definition of a Nash flow, which is an easy consequence of the fact that all agents in a flow at Nash equilibrium select shortest paths according to their personal objective functions.

Lemma 4.1 *Let (G, ℓ, α) be an instance where α takes on a finite number of values $\alpha_1 < \cdots < \alpha_k$. Then:*

(a) a flow on paths $\{\tilde{f}_P\}$ is induced by some flow at Nash equilibrium for (G^τ, ℓ, α) if and only if it minimizes $\sum_{i=1}^k \sum_{P \in \mathcal{P}} [\ell_P(\tilde{f}) + \alpha_i \tau_P] f_P^{(i)}$ over all possible flows on paths $\{f_P\}$;

(b) a flow on edges $\{\tilde{f}_e\}$ is induced by some flow at Nash equilibrium for (G^τ, ℓ, α) if and only if it minimizes $\sum_{i=1}^k \sum_{e \in E} [\ell_e(\tilde{f}_e) + \alpha_i \tau_e] f_e^{(i)}$ over all possible flows on edges $\{f_e\}$.

Lemma 4.1(a) corresponds more directly to the "all agents select shortest paths" intuition, but Lemma 4.1(b), which follows easily from the first assertion, will be more useful for our purposes.

Lemma 4.1 provides an extremely handy criterion for proving that a set of taxes is optimal. In particular, let $\hat{f}$ be an optimal flow for an instance (G, ℓ, α) in which α is a step function taking on values $\alpha_1 < \cdots < \alpha_k$. Define the *cost* $C(\hat{f})$ of $\hat{f}$ by $\sum_{i=1}^k \sum_{e \in E} [\ell_e(\hat{f}_e) + \alpha_i \tau_e] \hat{f}_e^{(i)}$. Lemma 4.1 implies that the flow on edges $\{\hat{f}_e\}_{e \in E}$ is induced by some flow at Nash equilibrium for (G^τ, ℓ, α) if and only if the optimal value of the mathematical program

$$
(P) \qquad
\begin{array}{lll}
\min & \displaystyle\sum_{i=1}^{k} \sum_{e \in E} [\ell_e(\hat{f}_e) + \alpha_i \tau_e] f_e^{(i)} & \\
\text{subject to:} & & \\
& \displaystyle\sum_{e=(v,w)} f_e^{(i)} - \sum_{e=(w,v)} f_e^{(i)} = b_{i,w} & \forall i \in \{1, \ldots, k\}, w \in V \\
& f_e^{(i)} \geq 0 & \forall e \in E, i = 1, 2, \ldots, k
\end{array}
$$

[3]More exactly, under this assumption an optimal flow can be computed to arbitrary precision in polynomial time. This is the best that could be hoped for, since an exact description of an optimal flow may require irrational numbers. We will not discuss this minor detail further in this abstract.

is $C(\hat{f})$, where $b_{i,w}$ is the number r_i of agents of commodity i if $w = t$, is $-r_i$ if $w = s$, and is 0 otherwise. Since $\hat{f}$ and τ are fixed, this is a linear program. By an argument similar to that proving Proposition 2.5, the assumption that all edge latency functions are convex implies that all optimal flows induce the same edge latencies; hence, this linear program is independent of the chosen optimal flow $\hat{f}$. It follows from this observation that a tax τ is optimal if and only if the optimal value of (P) is $C(\hat{f})$.

We now apply strong LP duality to (P), thereby finding that a set τ of taxes is optimal for (G, ℓ, α) if and only if the optimal value of the linear program

$$
(D) \qquad
\begin{aligned}
&\max \quad \sum_{i=1}^{k} r_i z_{t_i}^{(i)} \\
&\text{subject to:} \\
&\qquad z_{s_i}^{(i)} = 0 && \forall i \in \{1, \ldots, k\} \\
&\qquad z_w^{(i)} - z_v^{(i)} \leq \ell_e(\hat{f}_e) + \alpha_i \tau_e && \forall e = (v, w) \in E, i = 1, 2, \ldots, k
\end{aligned}
$$

equals $C(\hat{f})$. The optimal value of both (P) and (D) is at most $C(\hat{f})$ (to see this, take $f_e^{(i)} = \hat{f}_e^{(i)}$ for all i and e in (P)). Thus a set τ of taxes is optimal if and only if the following set of equations and inequalities has a solution in z:

$$
\begin{aligned}
& z_{s_i}^{(i)} = 0 && \forall i \in \{1, \ldots, k\} \\
& z_w^{(i)} - z_v^{(i)} \leq \ell_e(\hat{f}_e) + \alpha_i \tau_e && \forall e = (v, w) \in E, i = 1, 2, \ldots, k \\
& \sum_{i=1}^{k} r_i z_{t_i}^{(i)} = \sum_{i=1}^{k} \sum_{e \in E} [\ell_e(\hat{f}_e) + \alpha_i \tau_e] \hat{f}_e^{(i)}.
\end{aligned}
$$

The number of constraints is polynomial in k and in the size of G and, since an optimal flow on edges can be efficiently computed when latency functions are convex, the system can be constructed in polynomial time. The final key observation is this: this system is not only linear in z for fixed τ, but is linear in (z, τ) *even when τ is allowed to vary*. An optimal tax can thus be found by linear programming; the existence of such a tax is assured by Theorem 3.7.

We summarize the results of this section in a theorem.

Theorem 4.2 *Let (G, ℓ, α) be an instance with convex latency functions in which α takes on only finitely many distinct values. Then a linear description of the taxes optimal for (G, ℓ, α) can be computed in polynomial time. In particular, a set of optimal taxes can be computed in polynomial time.*

5 The Cost of Optimal Taxes

We have shown that optimal taxes exist under quite general conditions (Theorem 3.7). While this answers a very basic question about the power of edge pricing, Theorem 3.7 makes no guarantee about the disutility caused to agents due to such (possibly very large) taxes. In this section, we address this issue.

To account for the disutility caused to agents by taxes, we extend the total latency objective function of Section 2. For a flow f for an instance (G^τ, ℓ, α), we define the *cost* $C(f, \tau)$ of the flow f by

$$
C(f, \tau) = \int_0^1 c_{f(a)}^a(f, \tau) da
$$

where $c_P^a(f, \tau) = \ell_P(f) + \alpha(a)\tau_P$ (see also Section 2). This definition agrees with that of Section 4 in the special case when α takes on only finitely many different values. The definitions of total latency $L(\cdot)$ and cost $C(\cdot,\tau)$ agree if and only if $\tau = 0$. We also note that two different flows inducing the same flow on paths can have different costs (cf., the simpler total latency measure).

As we have spent the previous two sections studying taxes that minimize the total latency of Nash flows, a natural next goal would be to study taxes that minimize the cost of Nash flows. Unfortunately, this optimization problem is provably intractable (assuming $P \neq NP$) [8]. Instead, we will confine our attention to the cost of optimal taxes (taxes that induce a minimum-latency flow). Put differently, we investigate the question: how costly is a tax-induced minimum-latency routing?

Our contribution is a complete characterization of the distribution functions α for which the disutility due to optimal taxes is always at most a constant factor times the disutility due to latency—i.e., times the latency of an optimal flow. We emphasize that this is an extremely strong guarantee, and that there is no reason a priori to believe that *any* distribution function has this property. To better appreciate this, we discuss a modification of Example 2.6. Suppose we replace the latency function of link 1 in that example by the highly nonlinear function $\ell(x) = x^p$ for large p. A simple calculation (see e.g. [26]) shows that, in the absence of taxes, the total latency (or equivalently, the cost) of a Nash flow is arbitrarily larger than that of a minimum-latency flow $\hat{f}$ as $p \to \infty$. In this example, the guarantee above has the following consequence: even though with no taxes the cost of a Nash flow is arbitrarily larger than $L(\hat{f})$, by adding taxes—by making the cost function *more severe*—we can attain a Nash flow with cost within a constant factor of $L(\hat{f})$.[4] Thus, for some distributions α, taxes can improve the cost of a Nash flow by an arbitrarily large factor.

To begin our analysis, we formalize the guarantee we desire of a distribution function.

Definition 5.1 A distribution function α is *ρ-cheap* with parameter $\rho \geq 1$ if the following property holds: for every instance (G,ℓ,α) with $\alpha(0) > 0$, there is a set τ of optimal taxes and a minimum-latency flow f^τ at Nash equilibrium for (G^τ,ℓ,α) such that

$$C(f^\tau,\tau) \leq \rho \cdot L(f^\tau).$$

A distribution function is *cheap* if it is ρ-cheap for some finite $\rho \geq 1$.

It is easy to show that not all distribution functions are cheap. In fact, by considering only two-node two-link networks in which one link has the constant latency function $\ell(x) = 1$, we can derive the following necessary condition on ρ-cheap distributions (see Appendix B for a proof). In our statement of the condition, we employ the notation $\alpha(z^-)$ to mean the left limit $\lim_{z_n \uparrow z} \alpha(z_n)$ of a distribution function α at a point z. This limit always exists because distribution functions are assumed nondecreasing, and is equal to $\alpha(z)$ when α is continuous at z.

Lemma 5.2 *If α is a ρ-cheap distribution function, then*

$$\int_0^z \alpha(a)\,da \leq (\rho - 1)\cdot \alpha(z^-)[1-z] \tag{2}$$

for all $z \in (0,1)$.

Condition (2) is in essence a "growth condition", asserting that $\alpha(a)$ grows extremely quickly with a. The simplest distribution functions satisfying (2) for some value of ρ are the functions $\alpha(a) = (1-a)^{-k}$ for $k > 1$.

Remark 5.3 One particular consequence of condition (2) is that a cheap distribution function α is unbounded (for bounded α, the right-hand side of (2) vanishes as $z \uparrow 1$). Thus, step functions with finitely many values are not cheap.

[4]This arguably counterintuitive effect of taxation is reminiscent of the famous Braess's Paradox [7, 23], but can also occur in networks in which Braess's Paradox cannot (such as in networks of parallel links).

Far more remarkable than the necessary condition of Lemma 5.2—which remains necessary even when restricting attention to two-node, two-link networks—is the fact that condition (2) is also *sufficient* for a distribution to be ρ-cheap. We state this result next; its proof is postponed to Appendix B.

Theorem 5.4 *A distribution function α with $\alpha(0) > 0$ is ρ-cheap if and only if*

$$\int_0^z \alpha(a)\, da \leq (\rho - 1) \cdot \alpha(z^-)[1 - z]$$

for all $z \in (0, 1)$.

Remark 5.5 The condition of Theorem 5.4 is quite strong and is not satisfied by most distribution functions, and thus optimal taxes are in general quite costly. Nonetheless, we find it surprising that *any* natural distribution function is cheap, and satisfying that cheap distribution functions admit such a crisp mathematical characterization.

Acknowledgements

We thank Aaron Archer, David Eppstein, Rob Freund, Eric Maskin, Yuval Rabani, Andreas Schulz, Ilya Segal, Éva Tardos, Adrian Vetta, and Neal Young for helpful discussions.

References

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, 1993.

[2] A. Archer and É. Tardos. Truthful mechanisms for one-parameter agents. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 482–491, 2001.

[3] A. Archer and É. Tardos. Frugal path mechanisms. In *Proceedings of the 13th Annual Symposium on Discrete Algorithms*, pages 991–999, 2002.

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

[5] P. Bergendorff, D. W. Hearn, and M. V. Ramana. Congestion toll pricing of traffic networks. In P. M. Pardalos, D. W. Hearn, and W. W. Hager, editors, *Network Optimization*, pages 51–71. Springer-Verlag, 1997.

[6] K. C. Border. *Fixed Point Theorems with Applications to Economics and Game Theory*. Cambridge, 1985.

[7] D. Braess. Uber ein paradoxon der verkehrsplanung. *Unternehmensforschung*, 12:258–268, 1968. Available from `http://homepage.ruhr-uni-bochum.de/Dietrich.Braess/`.

[8] R. Cole, Y. Dodis, and T. Roughgarden. The cost of taxes for selfish routing. Submitted, 2002.

[9] S. C. Dafermos. The traffic assignment problem for multiclass-user transportation networks. *Transportation Science*, 6:73–87, 1972.

[10] S. C. Dafermos. Toll patterns for multiclass-user transportation networks. *Transportation Science*, 7:211–223, 1973.

[11] S. C. Dafermos and A. Nagurney. Sensitivity analysis for the asymmetric network equilibrium problem. *Mathematical Programming*, 28:174–184, 1984.

[12] S. C. Dafermos and F. T. Sparrow. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards, Series B*, 73B(2):91–118, 1969.

[13] D. W. Hearn and M. V. Ramana. Solving congestion toll pricing models. In P. Marcotte and S. Nguyen, editors, *Equilibrium and Advanced Transportation Modeling*, pages 109–124. Kluwer Academic Publishers, 1998.

[14] D. W. Hearn and M. B. Yildirim. A toll pricing framework for traffic assignment problems with elastic demand. In *Current Trends in Transportation and Network Analysis—Papers in Honor of Michael Florian*. 2002.

[15] M. Hirsch, C. H. Papadimitriou, and S. Vavasis. Exponential lower bounds for finding Brouwer fixpoints. *Journal of Complexity*, 5:379–416, 1989. Preliminary version in *FOCS '87*.

[16] T. Larsson and M. Patriksson. Side constrained traffic equilibrium models—traffic management through link tolls. In P. Marcotte and S. Nguyen, editors, *Equilibrium and Advanced Transportation Modeling*. Kluwer Academic Publishers, 1998.

[17] L. W. McKnight and J. P. Bailey. *Internet Economics*. MIT Press, 1998.

[18] I. Milchtaich. Generic uniqueness of equilibrium in large crowding games. *Mathematics of Operations Research*, 25(3):349–364, 2000.

[19] J. A. Mirrlees. The theory of optimal taxation. In K. J. Arrow and M. D. Intriligator, editors, *Handbook of Mathematical Economics, Volume III*, chapter 24, pages 1197–1249. North-Holland, 1986.

[20] J. F. Nash, Jr. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951.

[21] A. L. Peressini, F. E. Sullivan, and J. J. Uhl. *The Mathematics of Nonlinear Programming*. Springer-Verlag, 1988.

[22] A. C. Pigou. *The economics of welfare*. Macmillan, 1920.

[23] T. Roughgarden. Designing networks for selfish users is hard. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 472–481, 2001.

[24] T. Roughgarden. The price of anarchy is independent of the network topology. In *Proceedings of the 34th Annual ACM Symposium on the Theory of Computing*, pages 428–437, 2002.

[25] T. Roughgarden. *Selfish Routing*. PhD thesis, Cornell University, 2002.

[26] T. Roughgarden and É. Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002. Preliminary version in *FOCS '00*.

[27] D. Schmeidler. Equilibrium points of nonatomic games. *Journal of Statistical Physics*, 7(4):295–300, 1973.

[28] M. J. Smith. The marginal cost taxation of a transportation network. *Transportation Research, Series B*, 13:237–242, 1979.

[29] R. E. Tarjan. *Data Structures and Network Algorithms*. SIAM, 1983.

[30] J. G. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil Engineers, Pt. II*, volume 1, pages 325–378, 1952.

A Proofs from Section 3

A.1 Proofs from Subsection 3.1

We now implement the program described at the end of Subsection 3.1.

Lemma A.1 *Let (G, ℓ, α) be a well-behaved instance, with n vertices and $\ell_{max} = \max_{e \in E} \ell_e(1)$. Let taxes τ induce Nash flow f^τ, and suppose $T \geq 3n^3 \ell_{max}/\alpha(0)$. For a vertex v, let $d(v)$ denote the length of a shortest s-v path, using τ as edge lengths. Suppose $f^\tau_e > 0$, where $e = (v, w)$. Then*

$$d(w) - d(v) \geq \tau_e - \frac{T}{3n^2}.$$

Proof. Since $f^\tau_e > 0$ there is an agent $a \in [0, 1]$ that uses edge e in its s-t path. Let P_1 be the path that agent a uses to get from s to w. Let P_2 be an s-w path with minimum-possible tax, $d(w)$. Since f^τ is a Nash flow, agent a uses a c^a-minimum path; thus, $\ell_{P_1}(f^\tau) + \alpha(a)\tau_{P_1} \leq \ell_{P_2}(f^\tau) + \alpha(a)\tau_{P_2}$. Since $\tau_{P_1} \geq d(v) + \tau_e$ and $\tau_{P_2} = d(w)$, we can derive

$$d(w) - d(v) \geq \tau_e + \frac{1}{\alpha(a)}[\ell_{P_1}(f^\tau) - \ell_{P_2}(f^\tau)] \geq \tau_e - \frac{1}{\alpha(a)}\ell_{P_2}(f^\tau) \geq \tau_e - \frac{n\ell_{max}}{\alpha(0)}.$$

Since $T \geq 3n^3 \ell_{max}/\alpha(0)$, the lemma follows. ■

Lemma A.2 *Under the assumptions and notation of Lemma A.1, suppose there is a path P from v to w with $f^\tau_e > 0$ for all $e \in P$. Then*

$$d(w) - d(v) \geq \tau_P - \frac{T}{3n}.$$

Proof. Sum the previous lemma up over all edges in P. ■

Given a well-behaved instance (G^τ, ℓ, α) and a Nash flow f^τ, an ordering of the vertices of G is *good* if it satisfies property (P1) of Subsection 3.1 and if w following v with $d(w) < d(v)$ implies a path P from v to w with $f^\tau_e > 0$ on every edge of P. A good ordering can always be constructed, for example by topologically sorting the vertices of G w.r.t. the edges e with $f^\tau_e > 0$ and breaking ties among incomparable vertices by putting the vertex with smaller d-value first. The next lemma proves that a good ordering satisfies a relaxed version of property (P2) of Subsection 3.1.

Lemma A.3 *Suppose, after adopting the assumptions and notation of Lemma A.1 and placing a good ordering on the vertices of G, vertex w follows vertex v in the ordering. Then,*

$$d(w) \geq d(v) - \frac{T}{3n}.$$

Proof. Let w follow v in the good ordering. If $d(w) \geq d(v)$ the lemma holds. If $d(w) < d(v)$, then there must be a path P from v to w with $f^\tau_e > 0$ for all $e \in P$. The lemma now follows from Lemma A.2 and the nonnegativity of τ. ■

We can now prove Lemma 3.3.

Proof of Lemma 3.3. Assume for contradiction that Γ has a bad fixed point τ for the well-behaved instance (G, ℓ, α) inducing a Nash flow f^τ. Since τ is bad, $f^\tau_e \neq \hat{f}_e$ for some edge e. As both $\hat{f}$ and f^τ induce acyclic s-t flows on edges carrying one unit of flow, it is straightforward to show that there is at least one oversaturated edge, say $e = (v, w)$.

We seek undersaturated edges to contrast against the oversaturated edge e. We will discover such edges by looking at a particular type of s-t cut. Toward this end, place a good ordering on the vertices of G. We will write $x \preceq y$ if vertex x precedes (or is equal to) y in this ordering. All statements in the sequel that assume some ordering on the vertices refer to the ordering $\preceq$. By the *ith consecutive cut of* G, we mean the partition of G's vertex set into two classes where one class is the first i vertices of G.

Let v and w be the ith and jth vertices according to $\preceq$. Since $f^\tau_e > 0$, property (P1) of good orderings implies that $s \preceq v \prec w \preceq t$. It follows that $j > i$ and that consecutive cuts $i, i+1, \ldots, j-1$ are in fact s-t cuts. Moreover, property (P1) of good orderings and the fact that the net f^τ-flow across any s-t cut is 1 (see e.g. Tarjan [29]) implies that the total f^τ-flow escaping any of these consecutive cuts is precisely 1. By contrast, the total $\hat{f}$-flow escaping each of these cuts is at least 1 (it may be more, since our good ordering need not be a topological one w.r.t. the edges e with $\hat{f}_e > 0$). That consecutive cuts i through $j-1$ each contain an oversaturated edge (namely, e) implies, by counting, that each such cut also contains an undersaturated edge. Since τ is a bad fixed point of Γ, the tax on each of these undersaturated edges is 0.

We next identify a collection of undersaturated edges that "covers" the consecutive cuts i through $j-1$. Precisely, let $e_1 = (v_1, w_1)$ denote an undersaturated edge crossing the ith consecutive cut. We necessarily have $v_1 \preceq v$; if in addition $w_1 \succeq w$, the process halts. Otherwise, we let e_2 be an undersaturated edge crossing the consecutive cut whose last source-side vertex is w_1. Repeating this process, we obtain a sequence $e_1, \ldots, e_p$ of undersaturated edges with $e_i = (v_i, w_i)$ and $w_p \succeq w$. This process must halt with $p \leq n$ since the heads $\{w_1, \ldots, w_p\}$ of the undersaturated edges are strictly increasing.

We next claim that $d(w_p) \leq d(v_1) + \frac{T}{3}$. This follows directly from a chain of inequalities of two different types. First, $d(w_i) \leq d(v_i)$ for $i = 1, 2, \ldots, p$ since undersaturated edges are untaxed (as τ is a bad fixed point). The second type of inequality asserts that $d(v_i) \leq d(w_{i-1}) + \frac{T}{3n}$ for $i = 2, 3, \ldots, p$ and is an immediate consequence of Lemma A.3. Interleaving these two types of inequalities proves the claim.

We are now prepared to derive a contradiction. Since $v_1 \preceq v \prec w \preceq w_1$, two further applications of Lemma A.3 imply that

$$d(w) - d(v) \leq \frac{T}{3}\left[1 + \frac{2}{n}\right] \leq \frac{2}{3}T,$$

since we can assume that $n \geq 2$. On the other hand, since $e = (v, w)$ is an oversaturated edge, $f^\tau_e > 0$ and $\tau_e = T$; by Lemma A.1,

$$d(w) - d(v) \geq \left(1 - \frac{1}{3n^2}\right) T \geq \frac{11}{12}T.$$

These two incompatible inequalities provide a contradiction, showing that the bad fixed point τ cannot exist. ■

A.2 Proofs from Subsection 3.2

Proof of Lemma 3.5. Let (G, ℓ, α) denote an instance satisfying assumptions (A2)–(A4) of Subsection 3.1. This instance (G, ℓ, α) must admit an optimal flow $\hat{f}$ for which the edges e with $\hat{f}_e > 0$ form a directed acyclic graph, since removing flow cycles from a flow on edges only decreases its total latency.

Let $\tilde{G}$ denote the subgraph of G of edges with $\hat{f}_e > 0$. The instance $(\tilde{G}, \ell, \alpha)$ is then well behaved, so by Theorem 3.4 there is a set τ of taxes inducing a Nash flow f^τ for $(\tilde{G}^\tau, \ell, \alpha)$ with $\hat{f}_e = f^\tau_e$ for all $e \in \tilde{G}$. Let $T = \max_{e \in \tilde{G}} \tau_e$, and extend τ to all the edges of G by setting $\tau_e > n(T + \ell_{max}/\alpha(0))$ whenever $\hat{f}_e = 0$, where $\ell_{max} = \max_{e \in \tilde{G}} \ell_e(1)$.

We claim that f^τ is also a Nash flow for (G^τ, ℓ, α). To prove it, it suffices to show that no agent wants to use an s-t path of G^τ containing an edge e with $\hat{f}_e = 0$. This holds since agent $a \in [0, 1]$ experiences personal cost at most $n(\ell_{max} + \alpha(a)T)$ on the path $f^\tau(a)$, and would incur cost greater than $n(\frac{\alpha(a)}{\alpha(0)}\ell_{max} + \alpha(a)T) \geq n(\ell_{max} + \alpha(a)T)$ on any path of G^τ including an edge e with $\hat{f}_e = 0$. ■

We will need a statement slightly stronger than Lemma 3.5 in the sequel. The proofs of Theorem 3.4 and Lemma 3.5 do not use any properties of the optimum flow $\hat{f}$ of an instance other than that it is acyclic; we thus have the following extension of Lemma 3.5 for inducing acyclic flows that need not be minimum-latency.

Corollary A.4 *Let (G, ℓ, α) denote an instance satisfying assumptions (A2)-(A4) of Subsection 3.1 and let $\hat{f}$ be an acyclic flow. Then there is a set τ of taxes and a Nash flow f^τ for (G^τ, ℓ, α) such that $f^\tau_e = \hat{f}_e$ for all $e \in G$.*

Proof of Lemma 3.6. Let (G, ℓ, α) denote an instance satisfying assumptions (A3)–(A4) of Subsection 3.1, admitting acyclic optimal flow $\hat{f}$. For a positive integer k and an edge e of G, define $\ell^k_e(x) = \ell_e(x) + \frac{x}{k}$. We note that $\ell^k \to \ell$ uniformly on $[0,1]^E$. For each k, the instance (G, ℓ^k, α) satisfies assumptions (A2)–(A4) of Subsection 3.1; though $\hat{f}$ need not be optimal for (G, ℓ^k, α), Corollary A.4 applies and there is a set τ^k of taxes and a Nash flow f^k for $(G^{\tau^k}, \ell^k, \alpha)$ such that $f^k_e = \hat{f}_e$ for all $e \in G$. By Proposition 2.4, there is no loss of generality in assuming f^k to be canonical (see Definition 2.3).

We next show how the sequences $\{\tau^k\}$ and $\{f^k\}$ can be chosen to lie in compact subspaces of finite-dimensional Euclidean space. This is straightforward for taxes; inspection of the proofs of Theorem 3.4 and Lemma 3.5 shows that the maximum-allowable tax can be restricted by some function that depends only on $n = |V(G)|$, $\ell_{max} = \max_{e \in E} \ell_e(1)$, and $\alpha(0)$. These three parameters remain uniformly bounded over the sequence $\{(G, \ell^k, \alpha)\}$ of instances, so the sequence $\{\tau^k\}$ of taxes can be chosen to lie in the compact subset $[0, T]^E$ of $\mathcal{R}^E$, for sufficiently large T.

Arguing about the sequence of flows $\{f^k\}$ is more delicate, since each flow is actually a function defined on the unit interval. However, since each flow f^k is canonical, it suffices to deal with more elementary objects. Toward this end, to each canonical flow f^k we associate a *summary vector* v^k in $[-1, 1]^{\mathcal{P} \times \{1,2\}}$. The two components of v^k corresponding to path $P \in \mathcal{P}$ describe the left and right endpoints of the interval of agents assigned to P in f^k, or by convention are both -1 if no such agents exist. Since $\mathcal{P}$ is a finite set, summary vectors are finite-dimensional. The canonical flow f^k can be recovered from v^k, up to the measure-zero set of subinterval endpoints.

The sequences $\{\tau^k\}$ and $\{v^k\}$ of taxes and summary vectors are sequences in the compact spaces $[0, T]^E$ and $[-1, 1]^{\mathcal{P} \times \{1,2\}}$. We can thus select a convergent subsequence of $\{(\tau_k, v_k)\}$. Retaining the original index scheme for simplicity, this implies the existence of convergent sequences $\tau^k \to \tau$ and $v^k \to v$, with a canonical flow f^k corresponding to v^k at Nash equilibrium for (G, ℓ^k, α) for each k. The sum of subinterval lengths in each vector v^k is one; since these vectors lie in a common finite-dimensional space, this property is preserved in the limit. It follows that the vector v defines a canonical flow f in a natural way. As before, the route assignments of f are completely determined up to the measure-zero set of agents that correspond to endpoints of the subintervals of v. We will call these agents *endpoints*. Since each flow f^k induces the same flow on edges as $\hat{f}$, so does the flow f. To finish the proof, we need only show that f is at Nash equilibrium for (G^τ, ℓ, α). We proceed by contradiction.

If f fails to be at Nash equilibrium for (G^τ, ℓ, α) but only endpoints are unsatisfied, then such endpoints can be harmlessly reassigned to their preferred routes to produce a Nash flow for (G^τ, ℓ, α) inducing the same flow on edges as $\hat{f}$. Otherwise, an agent $a \in [0, 1]$ that is not an endpoint is not on a c^a-minimum path in f; put differently, for some path $P \in \mathcal{P}$,

$$\ell_P(f) + \alpha(a)\tau_P < \ell_{f(a)}(f) + \alpha(a)\tau_{f(a)}.$$

Since $(\tau^k, v^k) \to (\tau, v)$, $\ell^k \to \ell$ uniformly on $[0, 1]$, a is not an endpoint of v, and $\alpha(a) < \infty$, for sufficiently large k we have $f^k(a) = f(a)$ and

$$\ell^k_P(f^k) + \alpha(a)\tau^k_P < \ell^k_{f^k(a)}(f^k) + \alpha(a)\tau^k_{f^k(a)}.$$

This contradicts that f^k is at Nash equilibrium for (G, ℓ^k, α), completing the proof. ■

B Proofs from Section 5

Proof of Lemma 5.2. Suppose α is a ρ-cheap distribution function. Let G be a two-node two-link network with links 1 and 2, and with $\ell_2(x) = 1$. For any real numbers $q_1, q_2 \in (0,1)$ there is a choice for the latency function ℓ_1 such that any optimal flow $\hat{f}$ satisfies $\hat{f}_1 = q_1$ and $\ell_1(\hat{f}_1) = q_2$. The function ℓ_1 can be chosen to be a polynomial (with possibly very large degree), if desired.

Choose $q_1, q_2 \in (0,1)$ and ℓ_1 as above; these choices can be arbitrary so long as α is continuous at q_1. This imposes little restriction on q_1 since α, as a nondecreasing function, is discontinuous at only countably many points. Next, let τ be an optimal set of taxes for (G, ℓ, α) and f^τ a minimum-latency flow at Nash equilibrium for (G^τ, ℓ, α) with the properties promised by Definition 5.1. There is no loss of generality in assuming that f^τ is canonical (see Definition 2.3). Since $\ell_1(\hat{f}_1) = q_2 < 1 = \ell_2(\hat{f}_2)$ and the Nash flow f^τ must assign agents to both links (since $\hat{f}_1 \in (0,1)$), $\tau_1 > \tau_2$ must hold. Subtracting a common tax from the two links leaves f^τ at Nash equilibrium and only decreases $C(f^\tau, \tau)$, so there is no loss of generality in assuming that $\tau_2 = 0$. Since f^τ is canonical and τ is an optimal tax, agents $[0, \hat{f}_1)$ are assigned to link 1 and agents $(\hat{f}_1, 1]$ are assigned to link 2; since α is continuous at $\hat{f}_1$, agent $\hat{f}_1$ is indifferent between its two options. This implies that

$$\ell_1(\hat{f}_1) + \alpha(\hat{f}_1)\tau_1 = 1$$

and hence $\tau_1 = (1 - \ell_1(\hat{f}_1))/\alpha(\hat{f}_1)$.

Since α is ρ-cheap,

$$L(f^\tau) + \int_0^{\hat{f}_1} \alpha(a)\tau_1 \, da = C(f^\tau, \tau) \leq \rho \cdot L(f^\tau).$$

Substituting for τ_1 and rearranging, we can derive

$$\int_0^{\hat{f}_1} \alpha(a)\, da \leq (\rho - 1)\alpha(\hat{f}_1) \left[\frac{1}{1 - \ell_1(\hat{f}_1)} - \hat{f}_1 \right].$$

Since $\ell_1(\hat{f}_1)$ can be chosen arbitrarily close to zero for any choice of $\hat{f}_1$, we find that

$$\int_0^z \alpha(a)\, da \leq (\rho - 1)\alpha(z)[1 - z] \tag{3}$$

at all points $z \in (0,1)$ at which α is continuous.

Taking left limits in (3) along points at which α is continuous, we find that

$$\int_0^z \alpha(a)\, da \leq (\rho - 1)\alpha(z^-)[1 - z],$$

for all $z \in (0,1)$. Since α was an arbitrary ρ-cheap distribution, we have derived the promised necessary condition. ■

Proof of Theorem 5.4. Let α denote a distribution function satisfying $\alpha(0) > 0$ and

$$\int_0^z \alpha(a)\, da \leq (\rho - 1) \cdot \alpha(z^-)[1 - z] \tag{4}$$

for all $z \in (0,1)$. We will assume for simplicity that α is continuous; this assumption can be removed with minor modifications to the following proof.

Fix an instance (G, ℓ, α); from Theorem 3.7 we know there is some optimal tax. We next claim something stronger.

Claim: *The instance (G, ℓ, α) admits an optimal tax τ with the property that some s-t path of G receives zero tax.*

Proof of Claim. Let τ be an optimal tax for (G, ℓ, α), inducing a flow f^τ at Nash equilibrium that induces the same flow on edges as an acyclic minimum-latency flow $\hat{f}$. Topologically sort the vertices of G so that all flow-carrying edges go forward in the vertex ordering. Beginning with the vertex preceding t and proceeding backwards in the vertex ordering, perform the following operation for each vertex $v \neq s$: if $\tau_v \geq 0$ is the minimum tax on any flow-carrying edge with tail v, subtract τ_v from the tax on each flow-carrying edge with tail v and add τ_v to the tax on each edge (flow-carrying or not) with head v. This operation does not affect the total tax on any s-t path used by f^τ and can only increase the tax on other s-t paths; hence the flow f^τ remains at Nash equilibrium after these modifications. When the source s is reached, subtract τ_s from the tax on all flow-carrying edges with tail s. Again, f^τ remains at Nash equilibrium. At the conclusion of these modifications, every vertex with an outgoing flow-carrying edge has at least one such edge with zero tax. This implies that some s-t path—indeed, an s-t path of flow-carrying edges—possesses zero tax. ■

Let τ be an optimal tax for (G, ℓ, α) that assigns zero tax to some s-t path, and let f^τ be a minimum-latency flow at Nash equilibrium for (G^τ, ℓ, α). We can take f^τ to be a canonical flow. Let $P_1, P_2, \ldots, P_p$ denote the paths to which f^τ assigns traffic. In what follows, we will abbreviate $f^\tau_{P_i}$ by f_i, $\ell_{P_i}(f^\tau)$ by ℓ_i, and τ_{P_i} by τ_i.

Since f^τ is canonical, we can assume without loss of generality that $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_p$, that $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_p$, and that the subinterval of agents assigned to path P_i precedes (in $[0, 1]$) that of agents assigned to path P_j if and only if $i < j$. Thus $[0, 1]$ is split into p subintervals, with the ith subinterval the agents assigned by f^τ to path P_i. With our assumptions and notation in place, we can now write the cost of f^τ in a manageable form:

$$\begin{aligned} C(f^\tau) &= \sum_{i=1}^{p} \int_{F_{i-1}}^{F_i} [\ell_i + \alpha(a)\tau_i] \; da \\ &= L(f^\tau) + \sum_{i=1}^{p} \tau_i \int_{F_{i-1}}^{F_i} \alpha(a) \; da \end{aligned} \tag{5}$$

where $F_i = \sum_{j=1}^{i} f_j$ denotes the amount of flow f^τ assigns to the paths with index at most i. Defining $\Delta_i = \tau_i - \tau_{i+1} \geq 0$ for $i \in \{1, 2, \ldots, p-1\}$ and $\Delta_p = \tau_p$, equation (5) is equivalent to

$$C(f^\tau) = L(f^\tau) + \sum_{i=1}^{p} \Delta_i \int_0^{F_i} \alpha(a) \; da.$$

Since condition (4) holds, $\lim_{z \uparrow 1} \alpha(z) = +\infty$ and there are agents with arbitrarily large α-value. Since some s-t path is assigned zero tax by τ and f^τ is at Nash equilibrium, it follows that $\tau_p = 0$. Hence,

$$C(f^\tau) = L(f^\tau) + \sum_{i=1}^{p-1} \Delta_i \int_0^{F_i} \alpha(a) \; da.$$

By hypothesis (4),

$$C(f^\tau) \leq L(f^\tau) + (\rho - 1) \sum_{i=1}^{p-1} \Delta_i \alpha(F_i)(1 - F_i). \tag{6}$$

We next make further use of the assumption that f^τ is at Nash equilibrium. Since α is continuous and f^τ is a Nash flow, for each $i < p$ the agent F_i is indifferent between the paths P_i and P_{i+1}. Thus,

$$\ell_i + \alpha(F_i)\tau_i = \ell_{i+1} + \alpha(F_i)\tau_{i+1}$$

implying that

$$\Delta_i \alpha(F_i) = \ell_{i+1} - \ell_i.$$

Plugging this equality into (6) we find that

$$\begin{aligned}
C(f^\tau) &\leq L(f^\tau) + (\rho - 1)\sum_{i=1}^{p-1}(\ell_{i+1} - \ell_i)(1 - F_i) \\
&= L(f^\tau) + (\rho - 1)\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} f_j(\ell_{i+1} - \ell_i) \\
&= L(f^\tau) + (\rho - 1)\sum_{i=2}^{p} f_i(\ell_i - \ell_1) \\
&\leq L(f^\tau) + (\rho - 1)\sum_{i=1}^{p} \ell_i f_i \\
&= \rho \cdot L(f^\tau),
\end{aligned}$$

which completes the proof. ∎