Computing Optimal Tolls with Arc Restrictions and Heterogeneous Players

Tomas Jelinek[1], Marcus Klaas[2], and Guido Schäfer[3]

1 **VU University Amsterdam, The Netherlands**
`t.jelinek@student.vu.nl`
2 **University of Amsterdam, The Netherlands**
`mail@marcusklaas.nl`
3 **CWI and VU University Amsterdam, The Netherlands**
`g.schaefer@cwi.nl`

Abstract

The problem of computing optimal network tolls that induce a Nash equilibrium of minimum total cost has been studied intensively in the literature, but mostly under the assumption that these tolls are unrestricted. Here we consider this problem under the more realistic assumption that the tolls have to respect some given upper bound restrictions on the arcs. The problem of taxing subnetworks optimally constitutes an important special case of this problem. We study the *restricted network toll problem* for both non-atomic and atomic (unweighted and weighted) players; our studies are the first that also incorporate *heterogeneous* players, i.e., players with different sensitivities to tolls.

For non-atomic and heterogeneous players, we prove that the problem is NP-hard even for single-commodity networks and affine latency functions. We therefore focus on parallel-arc networks and give an algorithm for optimally taxing subnetworks with affine latency functions. For weighted atomic players, the problem is NP-hard already for parallel-arc networks and linear latency functions, even if players are homogeneous. In contrast, for unweighted atomic and homogeneous players, we develop an algorithm to compute optimal restricted tolls for parallel-arc networks and arbitrary (standard) latency functions. Similarly, for unweighted atomic and heterogeneous players, we derive an algorithm for optimally taxing subnetworks for parallel-arc networks and arbitrary (standard) latency functions.

The key to most of our results is to derive (combinatorial) characterizations of flows that are inducible by restricted tolls. These characterizations might be of independent interest.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems; G.1.6 Optimization; G.2.1 Combinatorics; G.2.2 Graph Theory

Keywords and phrases Network routing games, coordination mechanisms, network tolls, taxing subnetworks, heterogeneous players

1 Introduction

Motivation and Background. It is a well-known fact that selfish route choices in network routing applications result in outcomes that are undesirable for the society as a whole. In urban road traffic, for example, selfish route choices lead to unnecessary traffic jams, thereby causing environmental pollution, waste of natural resources, time and money. The Texas A&M Transportation Institute states in its *2012 Urban Mobility Report* [13, page 1]: "The 2011 data are consistent with one past trend, congestion will not go away by itself — action is needed! [...] The problem is very large. In 2011, congestion caused urban Americans to travel 5.5 billion hours more and to purchase an extra 2.9 billion gallons of fuel for a congestion cost of $121 billion."

Conference title on which this volume is based on.
Editors: Billy Editor, Bill Editors; pp. 1–12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Road pricing is recognized to be one of the most effective means to reduce congestion in networks. The idea is to let the users pay for the usage of certain segments of the network by imposing tolls. Typically, such tolls incite the users to change their commuting behavior, e.g., by opting for alternative, possibly slightly longer routes, avoiding certain parts of the network at peak hours, etc. As a result, the network becomes less congested because the traffic is better distributed through the network. Currently, road pricing systems are implemented successfully in several large cities across the world (like in Singapore, Kopenhagen, Tel Aviv, London, Dubai, etc.). A fundamental problem in this context is to determine tolls such that the overall congestion of the underlying network is reduced.

In this paper, we study the problem of computing optimal tolls for the arcs of a given network that induce a Nash equilibrium of minimum total cost. This problem has been studied intensively in the literature for decades. One of the earliest articles addressing this problem is due to Beckman, McGuire and Winsten [1], where they show that *marginal cost tolls* induce an optimal flow as Nash equilibrium in non-atomic network routing games. However, most previous studies were conducted under the assumption that the tolls are *unrestricted.* This assumption is too simplistic in many situations and significantly limits the applicability of such tolls in practice. For example, by using marginal cost tolls we lose the ability to control which arcs of the network are tolled and by how much. Clearly, this is undesirable in real-world applications where one can impose tolls only on certain arcs of the network and wants to ensure that they do not exceed predefined amounts.

Only recently, researchers have started to investigate more refined network toll problems like the *taxing subnetworks problem* [8, 10], in which only a subset of the arcs can be tolled, or the *restricted network toll problem* [2], in which tolls have to respect some upper bound restrictions on the arcs. All three studies [2, 8, 10] focus on the case of *non-atomic* players that are *homogeneous*, i.e., all players are assumed to have equal sensitivities to tolls. Here we further advance these investigations.

Our Contributions. We study the *restricted network toll problem* [2] both for non-atomic and atomic (unweighted and weighted) players. In our studies we consider for the first time also the case of *heterogenous* players, i.e., players may have different sensitivities to tolls. Capturing heterogenous players is particularly important if it comes to applications where users experience different disutilities of travel time and monetary cost (due to the tolls). As it turns out, the heterogenous player case gives rise to several new challenges in devising algorithms for the computation of optimal restricted tolls.

The main contributions presented in this paper are as follows:

- In Section 3 we consider the case of non-atomic, heterogeneous players. We prove that the problem of computing optimal restricted tolls is NP-hard even for single-commodity networks with affine latency functions. In light of this negative result, we then focus on parallel-arc networks and derive a combinatorial characterization of flows that are inducible by restricted tolls. Exploiting this characterization, we derive an optimal algorithm for taxing subnetworks with affine latency functions.
- In Section 4 we consider atomic players. We first observe that for weighted players the problem of computing optimal restricted tolls is NP-hard already for parallel-arc networks with linear latency functions (even if the players are homogeneous). We therefore focus on unweighted players and parallel-arc networks with standard latency functions and derive an optimal algorithm to compute restricted tolls for homogeneous players. Further, we obtain an optimal algorithm for taxing subnetworks for heterogeneous players.

As in previous works [2, 8, 10], most of our exact algorithms work only for parallel-arc networks. However, our studies also reveal that this is basically unavoidable, unless one is willing to resort to approximation algorithms (assuming that P $\neq$ NP). Moreover, from a practical point of view the restricted network toll problem for parallel-arc networks is still very well motivated: for example, it captures the problem of pricing fast-lanes (or priority-lanes) of highways that can be used to bypass heavy traffic (like in Tel-Aviv).

Our Techniques. The main difficulty that we face here in designing algorithms to compute optimal restricted tolls is that the underlying problem is a bi-level optimization problem: the feasible tolls constitute a compact set over which we wish to optimize the cost of the corresponding Nash equilibria (which in turn are determined by the tolls). Typically, such bi-level optimization problems are hard to tackle.

The key to most of our algorithmic results is to derive characterizations of flows that are inducible by restricted tolls. For the unrestricted case, several such characterizations can be found in the literature (see, e.g., [4, 6, 9, 15]). Moreover, some of them can be adapted to also incorporate upper bound restrictions on tolls. For example, Fleischer et al. [6] characterize the inducibility of a flow for non-atomic, heterogeneous players by the existence of an optimal solution satisfying certain minimality conditions for a cleverly chosen linear program. The upper bound restrictions can easily be added to this LP formulation such that the same characterization continues to hold. However, the crux is that we cannot simply use this characterization here because it reveals very little about the *structure* of the flows that are inducible by these restricted tolls. In contrast, we derive characterizations that reveal some structural properties of the inducible flows which we then exploit to design our algorithms.

Related Work. Beckman, McGuire and Winsten [1] proved that for non-atomic, homogeneous players marginal cost tolls induce an optimal flow as a Nash equilibrium. The existence of such tolls for non-atomic, heterogeneous players has first been established for single-commodity networks by Cole, Dodis and Roughgarden [4] and then extended to multi-commodity networks by Yang and Huang [15] (see also the independent works by Fleischer, Jain and Mahdian [6] and Karakostas and Kolliopoulos [9]). Fleischer [5] shows that for single-commodity networks linear tolls (in terms of the maximum latency of the optimal flow) are sufficient to enforce an optimal flow as Nash equilibrium.

In the literature one distinguishes between tolls that are *weakly-optimal*, i.e., at least one induced Nash equilibrium is an optimal flow, and *strongly-optimal*, i.e., every induced Nash equilibrium is an optimal flow. Swamy [14] proved the existence of weakly-optimal tolls for atomic, heterogeneous players and splittable flow. For atomic, homogeneous players and unsplittable flow, Caragiannis, Kaklamanis and Kanellopoulos [3] show that for linear latency functions strongly-optimal tolls do not exist for multi-commodity networks or if the players are weighted. They also show that strongly-optimal tolls exist for parallel-arc networks with linear latency functions and unweighted players. Subsequently, Fotakis and Spirakis [7] proved that weakly-optimal tolls can be computed efficiently for single-commodity networks and that these tolls are strongly-optimal for series-parallel networks.

In this paper, we focus on the computation of weakly optimal tolls. Most related to our work are the recent articles [2, 8, 10]. As already mentioned, these studies concentrate on the case of non-atomic players that are homogeneous. Hoefer, Olbrich and Skopalik [8] study the problem of optimally taxing subnetworks. They show that this problem is NP-hard for two-commodity networks and affine latency functions by a non-trivial reduction from partition. We borrow several insights of their proof to establish NP-hardness for single-commodity networks and affine latency functions in the case of heterogeneous, non-atomic players here. They also derive an algorithm to compute optimal tolls for parallel-arc networks and affine latency functions. Recently, Kleinert et al. [10] extended the algorithm in [8] for optimally taxing subnetworks for parallel-arc networks to more general latency functions. The algorithm guarantees polynomial running time for instances satisfying the *inverse concavity property* (see [10] for details). The restricted network toll problem considered here was introduced by Bonifaci, Salek and Schäfer [2]. The authors show that optimal restricted tolls can be computed efficiently for parallel-arc networks and affine latency functions and also derive bounds on the efficiency of restricted tolls for multi-commodity networks and polynomial latency functions.

2 Preliminaries

We provide formal definitions of the concepts introduced in the Introduction. Suppose we are given an instance $\mathcal{I} = (G = (V, A), (\ell_a)_{a\in A}, (s_i, t_i)_{i\in[k]}, (r_i)_{i\in[k]})$ of the non-atomic network routing game, where G is a directed graph with latency functions $(\ell_a)_{a\in A}$ and k commodities $(s_i, t_i)_{i\in[k]}$ of demand $(r_i)_{i\in[k]}$. The goal of every player is to send his flow along a shortest latency path from its source s_i to its destination t_i. Let $\mathcal{P}_i$ denote the set of all simple directed s_i, t_i-paths in G and define $\mathcal{P} := \cup_{i\in[k]}\mathcal{P}_i$. An outcome of the game is a flow $f : [k] \times \mathcal{P} \to \mathbb{R}_+$ that is feasible, i.e., $\sum_{P\in\mathcal{P}_i} f^i_P = r_i$ for every $i \in [k]$. Given a flow f, the total flow on arc $a \in A$ is defined as $f_a := \sum_{i\in[k]} \sum_{P\in\mathcal{P}:a\in P} f^i_P$. The set of arcs used by commodity i is denoted by A^+_i and we define $A^+ = \cup_{i\in[k]}A^+_i$. We define the latency of a path $P \in \mathcal{P}$ with respect to f as $\ell_P(f) := \sum_{a\in P} \ell_a(f_a)$. The total cost $C(f)$ of f is given by its average latency, i.e., $C(f) := \sum_{P\in\mathcal{P}} f_P\ell_P(f)$. A flow that minimizes $C(\cdot)$ is called *optimal* and denoted by f^*. A feasible flow f is called a *Nash flow* (or *Wardrop flow*) with respect to $\ell := (\ell_a)_{a\in A}$ if

$$\forall i \in [k],\ \forall P \in \mathcal{P}_i,\ f^i_P > 0 : \qquad \ell_P(f) \le \ell_{P'}(f) \quad \forall P' \in \mathcal{P}_i. \tag{1}$$

Atomic network routing games are very similar. The only difference to the non-atomic setting is that the flow for each commodity has to be routed along a single path, i.e., $\forall i \in [k],\ \exists! P \in \mathcal{P}_i$ such that $f^i_P = r_i$. What is refered to as commodity $i \in [k]$ in the non-atomic setting, is considered a player of weight r_i in the atomic setting. (Note that we do not assume that the commodities are distinct. Thus, different players might have the same source and destination.) A flow can be regarded as a mapping from $[k]$ into $\mathcal{P}$. On parallel arc networks, $a(i)$ may be used to denote the arc player i uses in a given flow. Players are said to be unweighted if $r_i = 1$ for all $i \in [k]$. In the atomic setting, a feasible flow is a Nash flow when

$$\forall i \in [k],\ \forall P \in \mathcal{P}_i,\ f_P > 0 : \qquad \sum_{a\in P} \ell_a(f_a) \le \sum_{a\in P'} \ell_a(f_a + r_i) \quad \forall P' \in \mathcal{P}_i. \tag{2}$$

Throughout this paper, we assume that the latency functions are non-negative, non-decreasing, differentiable and semi-convex, i.e., $x \cdot \ell_a(x)$ is convex for every arc $a \in A$; such latency functions are also called *standard* [11]. In the non-atomic setting, the cost of a Nash flow is unique if the latency functions are standard; this property is not guaranteed to hold for atomic players.

We study the *restricted network toll problem* as introduced in [2]: We are given an instance $\mathcal{I}$ of the network routing game and threshold values $\theta := (\theta_a)_{a\in A}$ on the arcs. The goal is to determine non-negative tolls $\tau := (\tau_a)_{a\in A}$ for the arcs of the network that obey the bounds defined by the threshold functions $(\theta_a)_{a\in A}$. More formally, a toll vector $\tau = (\tau_a)_{a\in A}$ is called *θ-restricted* if for every arc $a \in A$, $0 \le \tau_a \le \theta_a$. Additionally, we are given a non-negative vector of player *sensitivities* (also called *types*) $\alpha := (\alpha_i)_{i\in[k]}$. α_i represents the fraction of how a player of type i values the cost of one unit of time (latency) compared to one unit of money (toll).[1] Without loss of generality, we assume that $0 < \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_k$ throughout this paper. When $\alpha_i = \alpha_j$ for all $i, j \in [k]$, we say the players are *homogeneous* with respect to their sensitivity to toll; otherwise, we call them *heterogeneous*. In the former case, we can assume without loss of generality that α is normalized to 1.

Given a feasible flow f, we define the *combined cost* that a player of type i experiences by traversing arc $a \in A$ as $\phi_a^i(f_a) = \ell_a(f_a) + \alpha_i \tau_a(f_a)$. The goal of a player of type i is to choose a path P that minimizes the combined cost $\ell_P(f) + \alpha_i \tau_P$, where $\tau_P := \sum_{a \in P} \tau_a$. For θ-restricted tolls τ, let f^τ denote a Nash flow that is induced by τ, i.e., f^τ is a Nash flow with respect to the combined costs $(\phi^i)_{i\in[k]}$. Given the restrictions $\theta = (\theta_a)_{a\in A}$ on the arcs, a θ-restricted toll vector τ is *optimal* if there exists a Nash flow f^τ that is inducible by τ whose cost satisfies $C(f^\tau) \leq C(f^{\bar{\tau}})$ for all Nash flows $f^{\bar{\tau}}$ that are inducible by θ-restricted tolls $\bar{\tau}$. The optimization problem that we are considering in this paper is to compute θ-restricted tolls that are optimal.

An important special case of the θ-restricted network toll problem is the so-called *taxing subnetworks problem* [8]. Here we are given a set $T \subseteq A$ of arcs that are taxable arbitrarily while the arcs in $N := A \setminus T$ are non-taxable. This problem is equivalent to setting $\theta_a = \infty$ for every $a \in T$ and $\theta_a = 0$ for every $a \in N$.

3 Non-Atomic Players

We focus on the non-atomic, heterogeneous player case in this section. We first prove that the problem of computing optimal θ-restricted tolls is NP-hard for single-commodity networks.

▶ **Theorem 1.** *The problem of deciding whether there exist θ-restricted tolls that induce a flow of social cost at most K is NP-complete, even for single-commodity networks and affine latency functions with non-atomic and heterogeneous players.*

Our proof is an adaptation of the NP-hardness result for taxing subnetworks for two-commodity instance, affine latency functions and homogeneous players presented in [8]. The idea is to 'mimic' the behavior of their two-commodity instance by a single-commodity instance. To this aim, we have to overcome several difficulties. The proof is involved and due to lack of space deferred to the full version of the paper.

In light of Theorem 1, we subsequently restrict our attention to parallel-arc networks. We first establish a combinatorial characterization of inducible flows which we then use to derive an optimal algorithm for taxing subnetworks.

3.1 Characterization for Parallel-Arc Networks

We present a characterization of flows that are inducible by θ-restricted tolls in parallel-arc networks. Our approach is algorithmic: We first derive an algorithm for computing tolls that induce a given flow *without* any restrictions. We then show that the computed tolls are component-wise minimal and use this insight to derive our final characterization. Our characterization holds for arbitrary latency functions.

We assume that we have k different player types. Let $f = (f^i)_{i\in[k]}$ be a given flow, where f^i is the flow of player type i. Note that if f is fixed then all latencies of the arcs become

[1] Note that players having the same source and destination might still have different sensitivities in our setting because we do not assume that the commodities are distinct.

Algorithm 1: Computation of flow-inducing tolls

```
Input: flow $f = \sum_{i\in[k]} f^i$
Output: tolls $\tau = (\tau_a)_{a\in A}$ inducing $f$
1 $\tau \leftarrow \mathbf{0}$
2 for $a = 2 \to m^+$ do
3     $\Delta_a \leftarrow \frac{\ell_a - \ell_{a-1}}{\hat{\alpha}_{\min}(a)}$
4     for $\bar{a} = 1 \to (a-1)$ do $\tau_{\bar{a}} \leftarrow \tau_{\bar{a}} + \Delta_a$
5 end
6 return $\tau$
```

constants. For notational convenience, we will therefore omit the reference to f and write ℓ_a to refer to $\ell_a(f_a)$. Throughout this section we assume that $A = [m]$ and that the arcs are ordered such that $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_m$. This is without loss of generality because we can always relabel the arcs accordingly.

Let $m^+ = \max\{a \in A^+\}$ be the arc of largest latency that is used under f. For every arc $a \in A$ we define $L(a) := \{a' \in A \mid a' \geq a\}$ as the set of arcs that succeed a in the order above (including a). Note that the latency of every arc in $L(a)$ is at least ℓ_a. We use $\hat{\alpha}_{\min}(a)$ to refer to the minimal sensitivity of a player that uses an arc in $L(a)$, i.e., $\hat{\alpha}_{\min}(a) = \min\{\alpha_i \mid A_i^+ \cap L(a) \neq \emptyset\}$; we adopt the convention that $\hat{\alpha}_{\min}(a) = \infty$ if no such player exists.

Algorithm 1 describes the computation of a toll vector τ inducing f. We first establish some properties of this algorithm. It will turn to be convenient to view the algorithm as operating in *phases*, where a phase corresponds to the execution of the outer for-loop (Lines 2–5) for a fixed $a \in \{2, \ldots, m^+\}$.

▶ **Lemma 2.** *Let τ be the toll vector computed by Algorithm 1. Then for every two arcs $\hat{a}, \check{a} \in A$ with $\hat{a} \geq \check{a}$ it holds $\tau_{\check{a}} - \tau_{\hat{a}} = \sum_{a=\check{a}+1}^{\hat{a}} \frac{\ell_a - \ell_{a-1}}{\hat{\alpha}_{\min}(a)}$.*

Proof. Note that $\tau_{\check{a}}$ and $\tau_{\hat{a}}$ remain zero during phases $a = 2, \ldots, \check{a}$ and are increased by the same amount Δ_a in phases $a = \hat{a}+1, \ldots, m^+$. In phase $a \in \{\check{a}+1, \ldots, \hat{a}\}$, $\tau_{\check{a}}$ is increased by Δ_a, while $\tau_{\hat{a}}$ remains zero. The claim follows from the definition of $\Delta_a = \frac{\ell_a - \ell_{a-1}}{\hat{\alpha}_{\min}(a)}$. ◀

The following theorem gives a characterization of flows that are inducible by *unrestricted* tolls.

▶ **Theorem 3.** *A flow $f = (f^i)_{i\in[k]}$ of non-atomic, heterogeneous players with sensitivities $0 < \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_m$ is inducible by unrestricted tolls if and only if*

$$\forall i, j \in [k] \text{ with } \alpha_i > \alpha_j,\ \forall a \in A_i^+,\ \forall \bar{a} \in A_j^+ :\ \ell_a \geq \ell_{\bar{a}}. \tag{3}$$

Intuitively, the above condition states that a flow is inducible if more sensitive players are routed on arcs with larger latencies.

Proof. Assume for the sake of contradiction that f is inducible and (3) does not hold, i.e., there exist $i, j \in [k]$ with $\alpha_i > \alpha_j$ and arcs $a \in A_i^+$, $\bar{a} \in A_j^+$ such that $\ell_a < \ell_{\bar{a}}$. Note that f is a Nash flow with respect to the combined costs because it is inducible. Since $a \in A_i^+$ and $\bar{a} \in A_j^+$, we have

$$\ell_a + \alpha_i \tau_a \leq \ell_{\bar{a}} + \alpha_i \tau_{\bar{a}} \quad \text{or, equivalently,} \quad \ell_{\bar{a}} - \ell_a \geq \alpha_i (\tau_a - \tau_{\bar{a}}). \tag{4}$$

$$\ell_{\bar{a}} + \alpha_j \tau_{\bar{a}} \leq \ell_a + \alpha_j \tau_a \quad \text{or, equivalently,} \quad \ell_{\bar{a}} - \ell_a \leq \alpha_j (\tau_a - \tau_{\bar{a}}). \tag{5}$$

Note that $\ell_{\bar{a}} - \ell_a > 0$ by assumption and $\alpha_j > 0$. Inequality (5) therefore implies that $\tau_a - \tau_{\bar{a}} > 0$. Combining (4) and (5), we obtain $\alpha_j(\tau_a - \tau_{\bar{a}}) \geq \alpha_i(\tau_a - \tau_{\bar{a}})$. Dividing both sides by $\tau_a - \tau_{\bar{a}} > 0$ leads to a contradiction because $\alpha_j < \alpha_i$ by assumption.

Now suppose that (3) holds. We show that f is inducible by the tolls τ computed by Algorithm 1. Assume for the sake of contradiction that f is not a Nash flow with respect to the combined cost. Then for some player i there exist arcs $a \in A_i^+$ and $\bar{a} \in A$ satisfying

$$\ell_a + \alpha_i\tau_a > \ell_{\bar{a}} + \alpha_i\tau_{\bar{a}} \quad \text{or, equivalently,} \quad \ell_a - \ell_{\bar{a}} > \alpha_i(\tau_{\bar{a}} - \tau_a). \tag{6}$$

Let $\check{a} = \min\{a, \bar{a}\}$ and $\hat{a} = \max\{a, \bar{a}\}$. By Lemma 2, the difference in toll is

$$\tau_{\check{a}} - \tau_{\hat{a}} = \sum_{a'=\check{a}+1}^{\hat{a}} \frac{\ell_{a'} - \ell_{a'-1}}{\hat{\alpha}_{\min}(a')}. \tag{7}$$

Recall that $\hat{\alpha}_{\min}(a') = \min\{\alpha_j \mid A_j^+ \cap L(a') \neq \emptyset\}$. We distinguish two cases:

Case 1: $\bar{a} \leq a$. We have $\hat{a} = a \in A_i^+$ and thus $a \in L(a')$ for every $a' \in \{\check{a}+1, \dots, \hat{a}\}$. As a consequence, $\hat{\alpha}_{\min}(a') \leq \alpha_i$. Now (7) implies that $\alpha_i(\tau_{\bar{a}} - \tau_a) \geq \ell_a - \ell_{\bar{a}}$, which is a contradiction to (6).

Case 2: $\bar{a} > a$. We have $\check{a} = a \in A_i^+$. Note that by assumption all arcs in $A_j^+ \cap L(a')$ satisfy $\alpha_j \geq \alpha_i$ for every $a' \in \{\check{a}+1, \dots, \hat{a}\}$. Thus, $\hat{\alpha}_{\min}(a') \geq \alpha_i$ and (7) implies that $\alpha_i(\tau_a - \tau_{\bar{a}}) \leq \ell_{\bar{a}} - \ell_a$, which is a contradiction to (6). ◀

The following lemma is crucial in order to obtain our characterization of flows that are inducible by θ-restricted tolls.

▶ **Lemma 4.** *Let f be an inducible flow. Then the tolls τ computed by Algorithm 1 are component-wise minimal tolls that induce f.*

Proof. Assume for the sake of contradiction that τ is not component-wise minimal, i.e., there exists a toll vector τ' which induces f and $\tau'_a < \tau_a$ for some arc $a \in A$. Choose $\bar{a} \in A$ as the arc with largest latency such that $\tau'_{\bar{a}} < \tau_{\bar{a}}$. Note that the toll that Algorithm 1 imposes on arc $m^+ \in A^+$ is $\tau_{m^+} = 0$. Because $\tau_{\bar{a}} > 0$ there must exist at least one arc in A^+ whose latency is strictly larger than $\ell_{\bar{a}}$. Let $\hat{a} \in A^+$ be an arc which has minimal latency among such arcs, i.e.,

$$\hat{a} = \arg\min\{\ell_a \mid a \in A^+,\ \ell_a > \ell_{\bar{a}}\}.$$

Let i be the least toll sensitive player that uses arc $\hat{a}$. By Lemma 2, $\tau_{\bar{a}} - \tau_{\hat{a}} = \frac{\ell_{\hat{a}} - \ell_{\bar{a}}}{\alpha_i}$. Note that for player i the combined costs of $\hat{a}$ and $\bar{a}$ are equal because

$$\phi^i_{\hat{a}}(\tau) = \ell_{\hat{a}} + \alpha_i\tau_{\hat{a}} = \ell_{\bar{a}} + \alpha_i\tau_{\hat{a}} + \alpha_i\left(\frac{\ell_{\hat{a}} - \ell_{\bar{a}}}{\alpha_i}\right) = \ell_{\bar{a}} + \alpha_i\tau_{\bar{a}} = \phi^i_{\bar{a}}(\tau).$$

Further, $\phi^i_{\bar{a}}(\tau') < \phi^i_{\bar{a}}(\tau)$ because $\tau'_{\bar{a}} < \tau_{\bar{a}}$. Because f is a Nash flow with respect to τ' it follows that $\phi^i_{\hat{a}}(\tau') \leq \phi^i_{\bar{a}}(\tau')$. Therefore, $\phi^i_{\hat{a}}(\tau') \leq \phi^i_{\bar{a}}(\tau') < \phi^i_{\bar{a}}(\tau) = \phi^i_{\hat{a}}(\tau)$. This implies that $\tau'_{\hat{a}} < \tau_{\hat{a}}$ which is a contradiction to the choice of $\bar{a}$. ◀

▶ **Theorem 5.** *A flow $f = (f^i)_{i\in[k]}$ for non-atomic, heterogeneous players with sensitivities $0 < \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m$ is inducible by θ-restricted tolls if and only if*
1. *$\forall i, j \in [k]$ with $\alpha_i > \alpha_j$, $\forall a \in A_i^+$, $\forall \bar{a} \in A_j^+$: $\ell_a \geq \ell_{\bar{a}}$.*
2. *$\forall a \in A$: $\theta_a \geq \tau_a$, where τ is the toll vector computed by Algorithm 1 for f.*

Proof. Let f be inducible by θ-restricted tolls τ'. Clearly, f is also inducible by unrestricted tolls and the first condition therefore follows by Theorem 3. Let τ be the toll vector computed by Algorithm 1. Then f is also inducible by τ. By Lemma 4, we have $\tau_a \leq \tau'_a$ for every arc $a \in A$ because τ is component-wise minimal. Since τ' is θ-restricted we conclude that $\tau_a \leq \tau'_a \leq \theta_a$ for every arc $a \in A$, which proves the second condition.

Next, suppose that Conditions 1 and 2 hold. Then f is inducible by unrestricted tolls by Theorem 3. In particular, the proof of Theorem 3 shows that the tolls τ computed by Algorithm 1 induce f. Condition 2 now establishes that τ is θ-restricted, which concludes the proof. ◀

3.2 Taxing Subnetworks

Bonifaci, Salek and Schäfer [2] derive a polynomial-time algorithm for computing optimal θ-restricted tolls for homogeneous, non-atomic players on parallel-arc networks with affine latency functions. We show that this algorithm can be used to determine optimal tolls for taxing subnetworks in the presence of *heterogeneous* players.

Suppose we are given an instance of the taxing subnetworks problem with affine latency functions and k player classes with sensitivities $(\alpha_i)_{i\in[k]}$ and demands $(r_i)_{i\in[k]}$. Let $T \subseteq A$ and $N := A \setminus T$ be the sets taxable and non-taxable arcs, respectively. Set $r := \sum_{i\in[k]} r_i$ and $L = \max_{a\in A} \ell_a(r) - \min_{a\in A} \ell_a(0)$. Note that for homogeneous players and parallel-arc networks the maximum toll needed to induce a given flow is at most L. We can therefore define an instance of the θ-restricted toll problem with $\theta_a = L$ for every $a \in T$ and $\theta_a = 0$ for every $a \in N$. Now, compute an optimal θ-restricted toll vector τ for demand r by running the algorithm in [2]. Let $f^\tau = (f^\tau_a)_{a\in A}$ be the resulting Nash flow. The idea now is to turn the arc flow $f^\tau = (f^\tau_a)_{a\in A}$ into a player flow $f = (f^i)_{i\in[k]}$ such that the properties of Theorem 5 are satisfied. To this aim, we decompose f^τ into k player flows $f = (f^i)_{i\in[k]}$ in such a way that more sensitive players are assigned flow from higher latency arcs. We call this the *canonical decomposition* of f^τ.

▶ **Theorem 6.** *Let $f = (f^i)_{i\in[k]}$ be the canonical decomposition of the flow f^τ as described above. Then f is an optimal θ-restricted flow.*

Proof. Because f^τ is inducible by θ-restricted tolls for homogeneous players, it holds that for every $a \in N$ and $\bar{a} \in A^+$, $\ell_a(f_a) \geq \ell_{\bar{a}}(f_{\bar{a}})$. In particular, all the arcs in $N^+ := N \cap A^+$ have equal latencies and the latencies of arcs in $N \setminus N^+$ are at least as large. Now, the canonical decomposition guarantees that the resulting flow f satisfies (3) of Theorem 3. We can thus use Algorithm 1 to generate tolls which induce f for heterogenous players (as in the proof of Theorem 3). Further, these tolls will not impose any tolls on N because there is no flow-carrying arc with a latency larger than the one in N^+. Since $f^\tau_a = \sum_{i\in[k]} f^i_a$ for every $a \in A$, the total cost is not altered by this decomposition.

We next prove optimality. Let $f = (f_i)_{i\in[k]}$ be an optimal flow inducible by θ-restricted tolls for the game with heterogenous players. Consider the arc flow defined as $f_a := \sum_{i\in[k]} f^i_a$. As before, we know that for every $a \in N$ and $\bar{a} \in A^+$, $\ell_a(f_a) \geq \ell_{\bar{a}}(f_{\bar{a}})$. If $N \cap A^+ \neq \emptyset$ then the maximum latency arc that is used must be in N and all arcs in N have latencies at least as large as this arc. Otherwise, $N \cap A^+ = \emptyset$. In either case, we do not need to impose any toll on the arcs in N to induce f. We conclude that f is also inducible by θ-restricted tolls for homogeneous players. Thus, the total cost of an optimal flow inducible by θ-restricted tolls for homogenous players is at most the cost of one for heterogeneous players. ◀

Algorithm 2: Algorithm for optimal completion of inducible basis f^0

Input: network $(\{s,t\}, A, \ell)$, flow f^0, demand d and latency cap L
Output: optimal flow $f \geq f^0$ such that $\ell_a(f_a) \leq L$, if it exists

```
f ← f^0
while d > 0 do
    Ā ← {a ∈ A | ℓ_a(f_a + 1) ≤ L}
    if Ā = ∅ then return failure
    ā ← arg min_{a∈Ā} (f_a + 1)ℓ_a(f_a + 1) − f_a ℓ_a(f_a)
    f_ā ← f_ā + 1
    d ← d − 1
end
```

4 Atomic Players

We turn to the problem of computing optimal θ-restricted tolls for parallel-arc networks with atomic players. Roughgarden [12] proved that it is NP-hard to compute an optimal flow for weighted atomic players in parallel-arc networks with linear latency functions. As a consequence, computing optimal tolls is NP-hard in this setting, even for homogeneous players and without restrictions on the tolls. We therefore assume that the players are homogeneous and unweighted.

4.1 Characterization of flows inducible by θ-restricted tolls

We first derive a characterization of inducible flows for unweighted homogeneous players on parallel arc networks.

▶ **Lemma 7.** *A flow f of unweighted homomogeneous players on a parallel arc network $(\{s,t\}, A)$ is inducible by θ-restricted tolls if and only if*

$$\ell_a(f_a + 1) + \theta_a \geq \ell_{\hat{a}}(f_{\hat{a}}) \quad \forall a \in A, \tag{8}$$

where $\hat{a} := \arg\max_{a \in A^+} \ell_a(f_a)$.

Proof. Suppose that the restriction in Equation (8) holds. Then the tolls τ defined by $\tau_a = \max\{0, \ell_{\hat{a}}(f_{\hat{a}}) - \ell_a(f_a + 1)\}$ are clearly non-negative and θ-restricted. Furthermore, for any $a \in A^+$ and $\bar{a} \in A$,

$$\begin{aligned} \ell_a(f_a) + \tau_a &\leq \max\{\ell_a(f_a), \ell_a(f_a) + \ell_{\hat{a}}(f_{\hat{a}}) - \ell_a(f_a + 1)\} \\ &\leq \max\{\ell_a(f_a), \ell_{\hat{a}}(f_{\hat{a}})\} = \ell_{\hat{a}}(f_{\hat{a}}) \leq \ell_{\bar{a}}(f_{\bar{a}} + 1) + \tau_{\bar{a}}. \end{aligned}$$

This means no player has an incentive to change its choice. Suppose Equation (8) is not satisfied. Then there is an arc $a \in A$ such that $\ell_a(f_a + 1) + \theta_a < \ell_{\hat{a}}(f_{\hat{a}})$. Then for any θ-restricted toll vector τ, $\ell_a(f_a + 1) + \tau_a < \ell_{\hat{a}}(f_{\hat{a}}) + \tau_{\hat{a}}$. A player on arc $\hat{a}$ will therefore want to switch to arc a, which means that the flow is not inducible by θ-restricted tolls. ◀

4.2 Optimal θ-Restricted Tolls on Parallel-Arc Networks

We exploit the above characterization to obtain an optimal algorithm to compute θ-restricted tolls for unweighted homogeneous players. The idea of the algorithm is to first guess the arc $\hat{a} \in A$ which is the maximum latency flow-carrying arc in a optimal solution and then the amount $i \in [k]$ of flow on it. We then compute for every other arc $a \in A \setminus \{\hat{a}\}$ the minimum

Algorithm 3: Algorithm for finding an optimal flow inducible by θ-restricted tolls

Input: network $(\{s,t\}, A, \ell, \theta)$ and player count k
Output: optimal flow f^* inducible by θ-restricted tolls

$f^* \leftarrow (k, \dots, k)$
for $\hat{a} \in A$ **do**
  for $i \in [k]$ **do**
    $f^0_{\hat{a}} \leftarrow i$
    $d \leftarrow k - i$
    for $a \in A \setminus \{\hat{a}\}$ **do**
      $f^0_a \leftarrow \min\{n \in [k] \mid \ell_a(n+1) + \theta_a \geq \ell_{\hat{a}}(i)\}$
      $d \leftarrow d - f^0_a$
    end
    if $d \geq 0$ *and* $\max_{a \in A^+} \ell_a(f^0_a) \leq \ell_{\hat{a}}(i)$ **then**
      $L \leftarrow \ell_{\hat{a}}(i)$
      $f \leftarrow$ optimal completion of f^0 (using Algorithm 2)
      if *flow completion succeeded and* $C(f) < C(f^*)$ **then** $f^* \leftarrow f$
    end
  end
end

required flow such that Equation (8) is satisfied; call this flow the *inducible basis* f^0. Every flow that contains the inducible basis f^0 is guaranteed to be inducible by θ-restricted tolls. We can thus disregard the restrictions and complete f^0 in the optimal way (as described in Algorithm 2).

▶ **Lemma 8.** *Algorithm 2 returns a flow which is cost minimal among all flows which contain the inducible basis f^0, ship d extra units and for which the latencies of used arcs do not exceed L, if such a flow exists.*

Proof. Suppose such a flow exists. We show that throughout the algorithm there exists an optimal solution f^* which has at least as much flow on each arc as the current flow f. Consider the ith iteration of the algorithm. In this iteration the algorithm increases the flow on $\bar{a}$ by one. (Note that because an optimal solution exists, such an arc always exists.) By our definition of $\bar{A}$, the resulting latency is at most L. Now suppose there is no optimal solution f^* such that $f^*_{\bar{a}} \geq f_{\bar{a}}$. Then after iteration $i-1$, there exists an optimal solution f^* which 'agreed' with the algorithm's flow and an arc a such that $f^*_a > f_a$. Then in this flow f^*, if we move one unit of flow from a to $\bar{a}$, the change in social cost would be

$$(f^*_{\bar{a}}+1)\ell_{\bar{a}}(f^*_{\bar{a}}+1) - f^*_{\bar{a}}\ell_{\bar{a}}(f^*_{\bar{a}}) + (f^*_a - 1)\ell_a(f^*_a - 1) - f^*_a\ell_a(f^*_a). \tag{9}$$

By the choice of $\bar{a}$ and the convexity of $x \cdot \ell_a(x)$,

$$\begin{aligned}(f^*_{\bar{a}}+1)\ell_{\bar{a}}(f^*_{\bar{a}}+1) - f^*_{\bar{a}}\ell_{\bar{a}}(f^*_{\bar{a}}) &\leq (f^*_{\bar{a}}+2)\ell_{\bar{a}}(f^*_{\bar{a}}+2) - (f^*_{\bar{a}}+1)\ell_{\bar{a}}(f^*_{\bar{a}}+1)\\ &= (f_{\bar{a}}+1)f_{\bar{a}}(f_{\bar{a}}+1) - f_{\bar{a}}\ell_{\bar{a}}(f_{\bar{a}})\\ &\leq (f_a+1)f_a(f_a+1) - f_a\ell_a(f_a)\\ &\leq f^*_a\ell_a(f^*_a) - (f^*_a-1)\ell_a(f^*_a-1).\end{aligned}$$

So the change in the social cost (9) is non-positive. Therefore the optimal solution f^* can be altered without increasing the social cost so that it 'agrees' again with f, a contradiction. ◀

We summarize the algorithm for finding an optimal flow inducible by θ-restricted tolls in Algorithm 3.

▶ **Theorem 9.** *Algorithm 3 finds in polynomial time a flow of minimum total cost which is inducible by θ-restricted tolls.*

Proof. Atomic network routing games with unweighted atomic players admit at least one Nash equilibrium. Since the number of feasible flows is finite, there exists an optimal solution f^*. Let $\hat{a}$ be its maximum latency arc and i the flow on this arc. Consider the iteration of the algorithm with the same choice of $\hat{a}$ and i. The algorithm then puts as much flow on every arc to ensure that it can stand its toll. Because f^* is inducible under θ-restrictions, it must ship at least as much flow on every arc. After this, the algorithm finds a cost-minimal flow of d units on the arcs $A \setminus \{\hat{a}\}$ with the added restriction that $\hat{a}$ remains the maximum-latency arc. Because of this restriction, the resulting flow remains inducible under θ-restrictions. It does this by increasing flow on the arc in such a way that the increase in social cost is minimized. This produces an optimal solution by Lemma 8. It is not hard to see that the algorithm never outputs a flow which is not inducible by θ-restricted tolls. It puts as much flow on every arc a such that it can stand its toll $\ell_{\hat{a}}(f_{\hat{a}}) - \ell_a(f_a + 1)$. If this requires more flow than there is demand, the flow is discarded and will never be returned. ◀

▶ Remark. Note that once an optimal flow inducible by θ-restricted tolls is found we can extract the respective tolls as described in the proof of Lemma 7.

4.3 Optimally Taxing Subnetworks with Heterogeneous Players

With the help of Algorithm 3, we can also compute an optimal solution to the taxing subnetworks problem on parallel-arc networks with heterogeneous players. We compute this optimal flow in polynomial time by the following steps: Run Algorithm 3 on the given network with k players and θ-restrictions as given. This returns a flow $(f_a)_{a\in A}$. We decompose this arc flow into a player flow $(f^i)_{i\in[k]}$ by assigning the most sensitive players to the arcs in N arbitrarily and the remaining players to the arcs in T using the canonical decomposition described in Section 3.2.

▶ **Theorem 10.** *The process described above generates an optimal θ-restricted flow.*

Proof. First we show that the flow is inducible by tolls on T. Run Algorithm 3 on the network $(\{s,t\}, T)$ to define tolls on arcs T. Since the canonical decomposition decomposes the flow on T, these tolls discourage players on T to change to a different arc in T. Furthermore, the maximum latency arc $\hat{a}$ in T^+ has a zero toll. Now note that for the flow f to be inducible in the homogeneous case, for every $a \in N, \bar{a} \in A^+$ it must hold that $\ell_a(f_a + 1) \geq \ell_{\bar{a}}(f_{\bar{a}})$. Since players on T have no incentive to switch to $\hat{a}$, they surely do not have any incentive to switch to an arc in N. Now consider the players on N. They would not change to another arc in N, or f would not be in equilibrium. Let a' be the maximum latency arc in A^+. Suppose $a' \in N$. We impose an additional toll to all arcs in T of $\max\{\ell_{a'}(f_{a'}) - \ell_{\hat{a}}(f_{\hat{a}})/\bar{\alpha}, 0\}$, where $\bar{\alpha} := \min_{i:a(i)\in N} \alpha_i$ denotes the minimum sensitivity amongst players on N. Then the cost that a player on N sees on $\hat{a}$ is at least

$$\ell_{\hat{a}}(f_{\hat{a}}) + \bar{\alpha} \cdot \frac{\ell_{a'}(f_{a'}) - \ell_{\hat{a}}(f_{\hat{a}})}{\bar{\alpha}} = \ell_{a'}(f_{a'}).$$

Since the players on $\hat{a}$ have no incentive to change to other arcs in T, neither do players in N, as their sensitivity to toll is at least as high. Because this extra toll is added to all arcs in T, the players on T are still in equilibrium. Consider an arbitrary user i that uses an arc in T. Then on arc $\hat{a}$ it sees cost of at most

$$\alpha_i \cdot \frac{\ell_{a'}(f_{a'}) - \ell_{\hat{a}}(f_{\hat{a}})}{\bar{\alpha}} + \ell_{\hat{a}}(f_{\hat{a}}) \leq \ell_{a'}(f_{a'}) - \ell_{\hat{a}}(f_{\hat{a}}) + \ell_{\hat{a}}(f_{\hat{a}}) = \ell_{a'}(f_{a'}),$$

so no player on T has an incentive to switch to N. Now let the maximum latency arc $a' \in T$. Then $\hat{a} = a'$. Because the players on T cannot gain by deviating to a different arc in T, for some player $i \in [k]$ such that $a(i) = \hat{a}$,

$$\alpha_i \cdot \tau_a + \ell_a(f_a) \geq \ell_{\hat{a}}(f_{\hat{a}}) \qquad \forall a \in T.$$

Since $\bar{\alpha} > \alpha_i$, for every user on some $\bar{a} \in N$

$$\bar{\alpha} \cdot \tau_a + \ell_a(f_a) \geq \alpha_i \cdot \tau_a + \ell_a(f_a) \geq \ell_{\hat{a}}(f_{\hat{a}}) \geq \ell_{\bar{a}}(f_{\bar{a}}) \qquad \forall a \in T.$$

So the decomposed flow is inducible by tolls.

It remains to prove optimality. Take an optimal solution to the heterogeneous variant of the problem. Again we know that for every $a \in N, \bar{a} \in A^+$ it must hold that $\ell_a(f_a+1) \geq \ell_{\bar{a}}(f_{\bar{a}})$, or it can never be induced by tolls which are zero on N. For arc flow f' defined by $f'_a := \sum_{i=1}^k f_a^i$, the component-wise minimal tolls that induce these are equal to $\max\{0, \ell_{\hat{a}}(f_{\hat{a}}) - \ell_a(f_a+1)\}$ for $a \in A$. Then for $a \in N$, $\tau_a = 0$. So the minimum cost solution to the heterogeneous problem is at most that of the homogeneous variant, which proves optimality. ◀

References

1. M. Beckmann, B. McGuire, and C. Winsten. *Studies in the Economics of Transportation*. Yale University Press, New Haven, 1956.
2. V. Bonifaci, M. Salek, and G. Schäfer. Efficiency of restricted tolls in non-atomic network routing games. In *Proc. 4th Symp. on Algorithmic Game Theory*, pages 302–313, 2011.
3. I. Caragiannis, C. Kaklamanis, and P. Kanellopoulos. Taxes for linear atomic congestion games. *ACM Transactions on Algorithms*, 7(1):1–31, 2010.
4. R. Cole, Y. Dodis, and T. Roughgarden. Pricing network edges for heterogeneous selfish users. In *Proc. 35th Symp. on Theory of Computing*, pages 521–530, 2003.
5. L. Fleischer. Linear tolls suffice: New bounds and algorithms for tolls in single source networks. *Theoretical Computer Science*, 348(2-3):217–225, 2005.
6. L. Fleischer, K. Jain, and M. Mahdian. Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games. In *Proc. 45th Symp. on Foundations of Computer Science*, pages 277–285, 2004.
7. D. Fotakis and P. Spirakis. Cost-balancing tolls for atomic network congestion games. In *Proc. 3rd Workshop on Internet and Network Economics*, pages 179–190, 2007.
8. M. Hoefer, L. Olbrich, and A. Skopalik. Taxing subnetworks. In *Proc. 4th Workshop on Internet and Network Economics*, pages 286–294, 2008.
9. G. Karakostas and S. G. Kolliopoulos. Edge pricing of multicommodity networks for heterogeneous selfish users. In *Proc. 45th Symp. on Foundations of Computer Science*, pages 268–276, 2004.
10. I. Kleinert, M. Klimm, T. Harks, and R. H. Möhring. Computing network tolls with support constraints. Networks, to appear.
11. T. Roughgarden. The price of anarchy is independent of the network topology. *Computer and Systems Sciences*, 67(2):341–364, 2003.
12. T. Roughgarden. On the severity of Braess's paradox: Designing networks for selfish users is hard. *Computer and Systems Sciences*, 72(5):922–953, 2006.
13. D. Schrank, B. Eisele, and T. Lomax. TTI's 2012 urban mobility report, 2012. http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-report-2012.pdf.
14. C. Swamy. The effectiveness of stackelberg strategies and tolls for network congestion games. *ACM Transactions on Algorithms*, 8(4):1–19, 2012.
15. H. Yang and H.-J. Huang. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. *Transportation Research B*, 38(1):1 – 15, 2004.