Algorithmica (2009) 53: 132–153
DOI 10.1007/s00453-007-9018-5

Stackelberg Strategies for Selfish Routing in General Multicommodity Networks

George Karakostas · Stavros G. Kolliopoulos

Received: 18 September 2006 / Accepted: 26 July 2007 / Published online: 13 September 2007

Abstract A natural generalization of the selfish routing setting arises when some of the users obey a central coordinating authority, while the rest act selfishly. Such behavior can be modeled by dividing the users into an α fraction that are routed according to the central coordinator's routing strategy (*Stackelberg strategy*), and the remaining $1-\alpha$ that determine their strategy selfishly, given the routing of the coordinated users. One seeks to quantify the resulting price of anarchy, i.e., the ratio of the cost of the worst traffic equilibrium to the system optimum, as a function of α. It is well-known that for $\alpha = 0$ and linear latency functions the price of anarchy is at most $4/3$ (J. ACM 49, 236–259, 2002). If α tends to 1, the price of anarchy should also tend to 1 for any reasonable algorithm used by the coordinator.

We analyze two such algorithms for Stackelberg routing, LLF and SCALE. For general topology networks, multicommodity users, and linear latency functions, we show a price of anarchy bound for SCALE which decreases from $4/3$ to 1 as α increases from 0 to 1, and *depends only on* α. Up to this work, such a tradeoff was known only for the case of two nodes connected with parallel links (SIAM J. Comput. 33, 332–350, 2004), while for general networks it was not clear whether such a result could be achieved, even in the single-commodity case. We show a weaker bound for LLF and also some extensions to general latency functions.

Research of G. Karakostas supported by an NSERC Discovery Grant and MITACS.
Research of S. Kolliopoulos partially supported by the University of Athens under the project Kapodistrias.

G. Karakostas (✉)
Department of Computing and Software, McMaster University, 1280 Main st. W., Hamilton, ON L8S 4K1, Canada
e-mail: karakos@mcmaster.ca

S.G. Kolliopoulos
Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimiopolis Ilissia, Athens 157 84, Greece

The existence of a central coordinator is a rather strong requirement for a network. We show that we can do away with such a coordinator, as long as we are allowed to impose taxes (tolls) on the edges in order to steer the selfish users towards an improved system cost. As long as there is at least a fraction α of users that pay their taxes, we show the existence of taxes that lead to the simulation of SCALE by the tax-payers. The extension of the results mentioned above quantifies the improvement on the system cost as the number of tax-evaders decreases.

Keywords Selfish routing · Stackelberg strategies · Price of anarchy

1 Introduction

We study the performance of a network shared by noncooperative nonatomic users. Every selfish user needs to route an infinitesimal amount of flow from a specified origin to a specified destination node. Let f be a flow vector defined on the network paths, which describes a given traffic pattern according to the standard multicommodity flow conventions. Every path P has a latency function $l_P(f)$ which expresses the disutility (delay) experienced by all users on the path due to the aggregated flow of all users using some edge of P. Each selfish user wants to choose a minimum-latency path from her origin to her destination node. The user interaction is modelled by studying the system in the steady state captured by the classic *traffic equilibrium* concept [24]. The traffic equilibrium is characterized by the *Wardrop principle*: for every origin-destination pair (s_i, t_i) the disutility on every used $s_i - t_i$ path is equal and less than or equal to the disutility on any unused $s_i - t_i$ path. Hence, in equilibrium no user has an incentive to unilaterally switch paths. There is a large literature on traffic equilibria in transportation science, see [18].

Selfish behavior induces inefficiency from the system perspective. Motivated by decentralized data networks, Koutsoupias and Papadimitriou [13] were the first to propose as a measure of this inefficiency the worst-possible ratio between the system cost of an equilibrium and of an optimal routing designed by a central coordinator. This ratio is called the *the price of anarchy*. In the context of selfish routing, we define the *system* (*social*) *cost* as the total latency of the users. The price of anarchy for selfish routing was studied in the seminal paper by Roughgarden and Tardos [20]. They showed that for linear latency functions the price of anarchy is at most 4/3, and this is tight. For arbitrary continuous latency functions the price of anarchy is unbounded [20]. Several other results have pinpointed the price of anarchy $\rho(\mathcal{L})$ for various families $\mathcal{L}$ of latency functions [6, 17, 18]. See the recent survey by Roughgarden [19] for a comprehensive overview. Parameterizing the price of anarchy solely by the latency type is legitimate: under mild assumptions the price of anarchy is independent of the network topology [17].

These results refer to one extreme of selfish routing, namely to the case where all users are selfish. The other extreme is the system optimum where all users are coordinated and follow the predetermined optimal routing. The natural question that arises then is: what happens when only a *fraction* of the users are selfish, while the rest follow a predetermined policy? Are there such policies that can *always* improve the price of anarchy, given that a non-zero fraction of users can be coordinated? If so, how does the improvement depend on this fraction? For example, if the improvement is insignificant even when almost all users are coordinated, then such a policy is obviously of little value. Issues like these are important for real networks [12], since in general, it is quite possible that they do not fall in one of the two extremes, but that their users are a mixture of selfish and coordinated ones. As we shall see, surprisingly little is known for such networks. In fact, the only case that has been thoroughly studied, even for the case of linear latency functions, is the case of a network with two nodes connected by a number of parallel links [15]. To our knowledge, the current work is the first that deals with the issues above for *general topology* networks, and with *multiple* origin-destination pairs for the users (*multicommodity case*).

Stackelberg Routing Our main results deal with *Stackelberg routing*, a notion first proposed by Korilis, Lazar and Orda [12]. An α fraction of the users are *coordinated* and the rest are *selfish.* The coordinated users are controlled by a coordinator who assigns them to routes computed by an algorithm of choice. This algorithm is the *Stackelberg policy.* Let $\bar{f}$ be the corresponding flow vector output by the algorithm. The remaining $1-\alpha$ fraction of the users choose paths selfishly by taking into account the specified routes of the coordinated users: if the selfish users reach a traffic pattern x, they experience latency $l_P(x+\bar{f})$ on a path P. The concept is inspired by Stackelberg games (see, e.g., [2]) where players are asymmetric and are divided into *leaders* and *followers.* The followers react rationally (in our terms selfishly) to the strategies imposed on them by the leaders. An important difference between Stackelberg games and Stackelberg routing is that in the former setting, each leader is selfishly interested in her own individual utility. In Stackelberg routing, coordinated users aim to improve the social cost.

A given Stackelberg policy σ induces an associated equilibrium in which the Wardrop principle holds for every selfish user. This is a *Stackelberg equilibrium.* Given a Stackelberg policy σ, the worst-possible ratio between the cost of a Stackelberg equilibrium, and the minimum total latency is an expression that should depend on α, and we call it *the price of anarchy curve of* σ. For convenience, we treat the price of anarchy curve interchangeably as a scalar (if one thinks of α as fixed) and as a function (if one thinks of α as variable). Let $\mathcal{L}$ be the family of latency functions at hand. Clearly (i) the curve of any policy σ passes through the points $(0, \rho(\mathcal{L}))$ and $(1,1)$. Conceivably, for any 'reasonable' Stackelberg policy (ii) the curve also has to be a continuous nonincreasing function of α. We call a curve fulfilling Conditions (i) and (ii) *normal.*

Previous Results on Stackelberg Routing As mentioned above, rather little is known for Stackelberg routing. Roughgarden [15] defined two natural Stackelberg policies SCALE and Largest Latency First (LLF). SCALE simply sets the flow on every path equal to α times the optimal flow f^{opt}. LLF in the context of parallel links orders the links in terms of their latency in the optimal solution and saturates them one-by-one, from largest to smallest, until there are no centrally controlled users remaining. Roughgarden [15] not only obtained normal curves for LLF on parallel links but he also proved the optimality of LLF for such networks with linear latency functions. More specifically he obtained a $4/(3+\alpha)$ price of anarchy for linear latency functions and a $1/\alpha$ bound for general latency functions. Both bounds are tight [15].

For non-linear latency functions, there are examples of four-node networks where it is impossible to achieve the $1/\alpha$ bound (Proposition B.3.1 in [16]). On multicommodity networks the performance can be arbitrarily bad for latency functions with negative coefficients [18], but we do not consider such functions here. Obtaining a $4/3-\epsilon$ guarantee for linear latencies or a bound weaker than $1/\alpha$ for general latency functions are mentioned in [15] and [19], respectively as open problems, even for the single-commodity case.

Our Results This paper establishes for the first time the existence of a normal curve for Stackelberg routing with linear latency functions on general multicommodity networks. This should be contrasted with the earlier results that applied only to the single-commodity parallel links network [15].

More specifically, we analyze SCALE and a version of LLF which we call strong LLF (cf. Sect. 2) for linear latency functions. For SCALE, our price of anarchy curve is $\frac{4}{3}-\frac{X}{3}$ where $X=\frac{(1-\sqrt{1-\alpha})(3\sqrt{1-\alpha}+1)}{2\sqrt{1-\alpha}+1}$. See Fig. 2 for a plot. Hence we show that a very simple policy to implement (SCALE) achieves a very significant improvement of the price of anarchy for the linear latency case. In view of the simplicity of the policy (SCALE) that achieves such an improvement, it is rather surprising that virtually no progress has been made since [15] appeared. One possible explanation is the fact that our analysis examines the network as a *whole*, and avoids the edge-by-edge bounding that has been the staple of classic results on the price of anarchy, e.g., [6, 17, 20]. The technical machinery that makes our approach possible is the analysis of selfish routing by Perakis [14]. We use the special structure of SCALE in order to relax one of the "hard" constraints that the bound of [14] needs to satisfy. Moreover, we demonstrate that our upper bound analysis for SCALE is nearly tight for every α, by giving a set of linear latency functions on the Braess graph for which SCALE performs very close to our upper bound. More details appear in Sect. 3. Our analyses of SCALE and strong LLF can be extended to general latency functions using the concept of Jacobian similarity [14], a notion adapted from Hessian similarity in interior point methods [14, 22]. The latter approach, which is outlined in Sect. 5, has the potential of yielding bounds that are specific to individual families of latency functions.

For parallel links Roughgarden [15] gives an example where LLF outperforms SCALE. On the other hand on the instance of the Braess paradox (cf. Sect. 4), SCALE outperforms LLF. Finally on our hard example for SCALE (cf. Sect. 3) which shares the same underlying graph with the Braess paradox, LLF outperforms SCALE. Hence the two policies are incomparable, in the sense that no policy dominates the other on all possible inputs. The three types of instances we described suggest that in order to achieve the best possible curve, both the network topology and the latency functions matter. Although this appears to be in stark contrast with the independence of the price of anarchy for selfish routing from the network topology [17], it should not come as a surprise: Stackelberg routing has an algorithmic component which is lacking from "traditional" selfish routing.

Selfish Routing with Tax Evasion The existence of a central coordinator is a rather strong requirement for a network. A well-studied alternative for mitigating the effects of selfishness goes back to the origins of traffic equilibria (see [3]): impose monetary taxes (tolls) per-unit-of-flow on the edges. Selfish users are conscious both of the travel latency and the monetary cost on a path. It is by now known that such taxes exist even when users are *heterogeneous*, i.e., they are divided into classes where each class has a different sensitivity level towards paying taxes [8, 10, 25]. See also the work in [4].

In the same way Stackelberg routing establishes partial control over the users by centrally coordinating only an α fraction of them, we examine whether similar effects can be achieved when only an α fraction of the users pay taxes. Equivalently, one can think of the remaining $1-\alpha$ fraction of the users as tax-evaders having a zero sensitivity to taxes. In Sect. 6 we show that there is a set of edge taxes so that the price of anarchy obtained is equal to the price of anarchy of the SCALE policy. As the fraction of law-abiding citizens increases from 0 to 1, the system cost is improved accordingly.

This work was first published as a McMaster University technical report [11]. Independently of our work, Correa and Stier-Moses [5] and Swamy [23] have obtained results for general latency functions. Other recent work on Stackelberg routing includes [9, 21].

2 Preliminaries

A directed network $G=(V,E)$, with parallel edges allowed, is given on which a set of users each want to route an infinitesimal amount of flow (traffic) from a specified origin to a destination node in G. Users are divided into k classes (commodities). The demand of class $i=1,\dots,k$, is $d_i>0$ and the corresponding origin–destination pair is (s_i,t_i). A *feasible* vector x is a valid flow vector (defined on the path or edge space as appropriate) that satisfies the standard multicommodity flow conventions and routes demands d_i for every commodity i. We use feasible flow vectors throughout the paper to characterize traffic patterns. We use K to denote the (convex) set of all the feasible vectors. As in [15] we assume separable costs: each edge e is assigned a nonnegative, nondecreasing latency function $l_e(f_e)$ that gives the delay experienced by any user on e due to congestion caused by the total flow f_e that passes through e. We also assume the standard additive model: for a path P, $l_P(f)=\sum_{e\in P}l_e(f_e)$.

Stackelberg policies can be classified as weak or strong [18]. A *weak Stackelberg policy* controls demand αd_i from each commodity for a parameter $\alpha\in(0,1)$. A *strong Stackelberg policy* gives more power to the coordinator: he can control as much demand from each commodity as he sees fit under the condition that the total demand controlled equals $\alpha\sum_{i=1}^k d_i$. In the single-commodity case, strong and weak policies coincide.

Let f^* be the flow vector of the selfish users and $\bar{f}$ the strategic flow of the coordinated users. The additive cost model makes it easy to view our flows sometimes as path flows and sometimes as edge flows. The system cost of feasible flow x is defined as $\sum_P x_P l_P(x)$. Let $C_{\mathrm{eq}} := \sum_P (f_P^* + \bar{f}_P) l_P(f^* + \bar{f})$ denote the cost at equilibrium. We denote by f^{opt} a flow that optimizes the system cost and the optimum itself as C_{opt}, i.e., $C_{\mathrm{opt}} := \sum_P f_P^{\mathrm{opt}} l_P(f^{\mathrm{opt}})$. The *SCALE policy* is a weak one defined by setting $\bar{f}_e := \alpha f_e^{\mathrm{opt}}$ for every e. Note that this is equivalent to setting $\bar{f}_P := \alpha f_P^{\mathrm{opt}}$ for all paths P. The *strong LLF policy* imposes a total order on the paths used by all commodities based on nondecreasing $l_P(f^{\text{opt}})$ values and breaking ties arbitrarily. It then saturates paths one by one from the largest latency to the smallest until the total demand of the controlled users equals $\alpha\sum_{i=1}^k d_i$. By saturating a path P, we mean that $\bar{f}_P := f_P^{\text{opt}}$. Note that the last path P_* in the ordering that is assigned positive flow may carry less than f_{P*}^{opt}. P_* is assigned as much flow as possible without exceeding a total of $\alpha\sum_{i=1}^k d_i$.

In this work we will use the concept of β-function defined by Correa et al. [6]. Let $\mathcal{L}$ be a family of continuous and non-decreasing latency functions. For every function $l \in \mathcal{L}$ and every value $v \geq 0$, let us define:

$$\beta(v,l) := \frac{1}{vl(v)} \max_{x\geq 0}\{x(l(v)-l(x))\}.$$

In addition, let us define

$$\beta(l) := \sup_{v\geq 0} \beta(v,l).$$

Then $\beta(\mathcal{L})$ is defined as follows:

$$\beta(\mathcal{L}) := \sup_{l\in\mathcal{L}} \beta(l).$$

3 Linear Latency Functions

In this section, we examine the case of linear (or affine) latency functions. That is, for all e, $l_e(f_e) = a_e f_e + b_e$, with $a_e, b_e \geq 0$.

A First Attempt Existing upper bounds on the price of anarchy depend to a large extent on the behavior of the latency function on individual edges. This is what we call the "edge-by-edge" approach. The definitions of the anarchy value $\alpha(\mathcal{L})$ by Roughgarden [17] and the $\beta(\mathcal{L})$ parameter by Correa et al. [6], where $\mathcal{L}$ is a class of latency functions, are particularly revealing in this context. In order to gain intuition into the problem we initially try an analysis of Stackelberg routing using similar arguments. We assume that the coordinator uses the SCALE policy. Let $\beta = \beta(\mathcal{L})$. The definition of β implies that for any edge e

$$f_e^{\text{opt}} l_e(f_e^* + \bar{f}_e) \leq \beta(f_e^* + \bar{f}_e) l_e(f_e^* + \bar{f}_e) + f_e^{\text{opt}} l_e(f_e^{\text{opt}}). \tag{1}$$

We can get a better upper bound when edge e is underutilized by the selfish users. Define an edge e to be *light* if $f_e^* \leq c\bar{f}_e$ for a suitable $c > 0$. An edge which is not light is called *heavy*. Define $\delta \in [0,1]$ such that $\sum_{e \text{ light}} f_e^{\text{opt}} l(f_e^{\text{opt}}) = (1-\delta)C_{\text{opt}}$ and $\sum_{e \text{ heavy}} f_e^{\text{opt}} l(f_e^{\text{opt}}) = \delta C_{\text{opt}}$.

Lemma 1 *Let c, δ be defined as above. Then for a general network with linear latency functions and a fraction α of coordinated users, SCALE achieves a price of anarchy* $\frac{C_{\text{eq}}}{C_{\text{opt}}} \leq \frac{4}{3}[1 - \frac{\alpha^2}{4}(1-\delta)]$.

Proof Since the $l_e(\cdot)$ functions are nondecreasing, we have that for the light edges

$$\sum_{e \text{ light}} f_e^{\text{opt}} l_e(f_e^* + \bar{f}_e) \le \sum_{e \text{ light}} f_e^{\text{opt}} l_e(\alpha(1+c) f_e^{\text{opt}}) \le \sum_{e \text{ light}} f_e^{\text{opt}} l_e(f_e^{\text{opt}}) \tag{2}$$

under the assumption that $\alpha(1+c) \le 1$. For heavy edges, (1) yields that

$$\sum_{e \text{ heavy}} f_e^{\text{opt}} l_e(f_e^* + \bar{f}_e) \le \beta \sum_{e \text{ heavy}} (f_e^* + \bar{f}_e) l_e(f_e^* + \bar{f}_e) + \sum_{e \text{ heavy}} f_e^{\text{opt}} l_e(f_e^{\text{opt}}). \tag{3}$$

For linear latency functions it is well-known that $\beta \le 1/4$ [6], hence later we will use the value $\beta = 1/4$. The analysis is affected by the amount of cost that f^{opt} pays on the light and heavy edges, respectively. From now on, we make use of the assumption that the edge latency functions are linear. By summing (2), (3) over all the edges we obtain that

$$\begin{aligned}
\sum_{e} f_e^{\text{opt}} l_e(f_e^* + \bar{f}_e) &\le \beta C_{\text{eq}} - \beta \sum_{e \text{ light}} (f_e^* + \bar{f}_e) l_e(f_e^* + \bar{f}_e) + C_{\text{opt}} \\
&\le \beta C_{\text{eq}} - \beta \alpha^2 \sum_{e \text{ light}} f_e^{\text{opt}} l_e(f_e^{\text{opt}}) + C_{\text{opt}} \\
&\le \beta C_{\text{eq}} + [1 - \beta \alpha^2 (1-\delta)] C_{\text{opt}}, \qquad (4)
\end{aligned}$$

where the second inequality is due to the fact that the l_e's are linear and $\alpha \le 1$.

Let $\hat{f} := f^{\text{opt}} - \bar{f}$ be the flow that remains if we remove flow $\bar{f}$ from the optimal flow f^{opt}. Note that $\hat{f}$ is a flow that satisfies demands $\hat{d}_i \le d_i$, for all commodities $i = 1, \ldots, k$. In the special case, where $\bar{f} = \alpha f_{\text{opt}}$, $\hat{d}_i = (1-\alpha) d_i$, for $i = 1, \ldots, k$. Then from the variational inequality

$$\sum_{e} l_e(f_e^* + \bar{f}_e)(x_e - f_e^*) \ge 0, \quad \forall x = \text{flow that satisfies demands } \hat{d}_i, i = 1, \ldots, k \tag{5}$$

that f^* satisfies as a traffic equilibrium [7], we get the following for $x := \hat{f}$:

$$C_{\text{eq}} = \sum_{e} (f_e^* + \bar{f}_e) l_e(f_e^* + \bar{f}_e) \le \sum_{e} (\hat{f}_e + \bar{f}_e) l_e(f_e^* + \bar{f}_e) = \sum_{e} f_e^{\text{opt}} l_e(f_e^* + \bar{f}_e). \tag{6}$$

By using (6) in (4), and replacing β by $1/4$ we get the lemma. □

If $\delta < 1$, Lemma 1 yields a normal curve. Hence we would like to have δ as small as possible. The parameter c must satisfy $\alpha(c+1) \le 1$, and, by definition, the bigger c is the smaller δ potentially is. Therefore we should pick $c := \frac{1-\alpha}{\alpha}$.

Note, though, that, even with this choice of c, it may still be the case that $\delta = 1$. In this case, the bound we have calculated is not better than the classic $4/3$ that holds when no Stackelberg policy is used. The "edge-by-edge" approach led us to believe that, at least for SCALE, the easy case is when f^{opt} pays a substantial fraction of its cost on edges that are underutilized by the selfish users. After completing our upper and lower bound derivations we will have demonstrated instead that a small δ is the difficult case.

An Improved Upper Bound for SCALE In order to prove our main result for SCALE and linear latency functions we will have to depart from the approach used above and examine the network as a whole. If for every edge e, the latency function is of the form $l_e(f_e) = a_e f_e$, it is well known that the price of anarchy is 1. The inefficiency of selfish routing for general affine functions could be attributed, in a sense, to the existence of load-independent latency terms $b_e > 0$ in some $l_e()$ functions. Our analysis exploits the fact that the SCALE and LLF policies decrease the total influence of these terms on the social cost at equilibrium.

Theorem 1 *For general multicommodity networks with linear latency functions and a fraction α of users coordinated by the SCALE policy, the price of anarchy is bounded as follows*

$$\frac{C_{\mathrm{eq}}}{C_{\mathrm{opt}}} \le \frac{4}{3} - \frac{X}{3}, \quad \textit{where } X = \frac{(1-\sqrt{1-\alpha})(3\sqrt{1-\alpha}+1)}{2\sqrt{1-\alpha}+1}.$$

Proof A lemma of Perakis [14] provides us with an important tool for our analysis. It was originally derived to deal with asymmetric and non-separable cost functions. Consider the latency function as a vector-valued function $L : \mathbb{R}_+^m \to \mathbb{R}_+^m$, with $L(f) = Gf + b$ and $m = |E|$. In our case, G is a diagonal matrix containing the a_e's and $b^T = [b_e]_{e\in E}$. In this notation $C_{\mathrm{eq}} = [L(f^* + \bar{f})]^T(f^* + \bar{f})$. From the proof of Theorem 3 in [14] we can abstract away the following fact, which isolates the contribution of the flow-dependent part of the latency to the total cost:

Lemma 2 ([14]) *Given the notation above, let $f \in K$ be a vector that satisfies $[L(f)]^T(f_{\mathrm{opt}} - f) \ge 0$. For any scalars $a_1, a_2 \ge 0$ such that $\begin{bmatrix} a_1 G^T & \frac{-G^T}{2} \\ \frac{-G}{2} & a_2 G^T \end{bmatrix}$ is positive semi-definite, we have that*

$$f^T G^T f^{\mathrm{opt}} \le a_1 f^T G f + a_2 (f^{\mathrm{opt}})^T G f^{\mathrm{opt}}.$$

In our case, G is symmetric, and $G \succeq 0$ since G is a diagonal matrix with entries $G[e,e] = a_e \ge 0$, for all $e \in E$. In this case, Lemma 2 can be reduced to a more malleable form, which is implicit in [14]:

Lemma 3 ([14]) *If for all edges e, $l_e(f_e) = a_e f_e + b_e$ with $a_e, b_e \ge 0$, then for any $a_1, a_2 \ge 0$ that satisfy $a_1 a_2 \ge 1/4$*

$$C_{\mathrm{eq}} \le a_1 \sum_e a_e (f_e^* + \bar{f}_e)^2 + a_2 \sum_e a_e (f_e^{\mathrm{opt}})^2 + \sum_e b_e f_e^{\mathrm{opt}}.$$

Proof For every $a_1, a_2 \ge 0$, we show that the semidefinite constraint of Lemma 2 is equivalent to the following holding for every $2m$-dimensional vector $X = [X_1\ X_2]^T$, where X_1, X_2 are m-dimensional vectors:

$$[X_1\ X_2]\cdot\begin{bmatrix} a_1 G & \frac{-G}{2} \\ \frac{-G}{2} & a_2 G\end{bmatrix}\cdot\begin{bmatrix} X_1 \\ X_2\end{bmatrix}\geq 0 \Leftrightarrow a_1 X_1^T G X_1 + a_2 X_2^T G X_2 - X_1^T G X_2 \geq 0. \tag{7}$$

If $a_1 a_2 \geq \frac{1}{4}$, then the following holds for any two numbers x_1^e, x_2^e:

$$a_e \cdot (a_1 (x_1^e)^2 + a_2 (x_2^e)^2 - x_1^e x_2^e) \geq a_e \cdot (\sqrt{a_1} x_1^e - \sqrt{a_2} x_2^e)^2 \geq 0. \tag{8}$$

By considering the coordinates x_1^e, x_2^e of X_1, X_2 separately, applying (8) to them, and finally adding over all edges e, we get (7).

We show that for $f = f^* + \bar{f}$ the second hypothesis of Lemma 2 is also satisfied. By (5)

$$[L(f^* + \bar{f})]^T ((f_{\text{opt}} - \bar{f}) - f^*) \geq 0 \Leftrightarrow [L(f^* + \bar{f})]^T (f_{\text{opt}} - (f^* + \bar{f})) \geq 0. \tag{9}$$

Inequality (9) yields that

$$C_{\text{eq}} = [L(f^* + \bar{f})]^T (f^* + \bar{f}) \leq [L(f^* + \bar{f})]^T f_{\text{opt}} = (f^* + \bar{f})^T G^T f^{\text{opt}} + b^T f_{\text{opt}}.$$

By using Lemma 2 to upperbound the right-hand side the proof is complete. □

Note that in order to apply Lemma 3 we are free to pick a_1, a_2 subject to the constraints of the Lemma. This is exactly the point where the SCALE policy helps us to get a better bound for the price of anarchy: while [14] also gets to pick a_1, a_2 subject to these constraints *and* the extra constraint $a_2 \geq 1$, we will not have to obey the latter constraint. The details of the proof follow.

We rewrite the right-hand side of Lemma 3 in terms of paths:

$$\begin{aligned} C_{\text{eq}} &\leq a_1 \sum_P (f_P^* + \bar{f}_P) \sum_{e\in P} a_e (f_e^* + \bar{f}_e) + a_2 \sum_P f_P^{\text{opt}} \sum_{e\in P} a_e f_e^{\text{opt}} + \sum_P f_P^{\text{opt}} \sum_{e\in P} b_e \\ &= a_1 C_{\text{eq}} - a_1 \sum_P (f_P^* + \bar{f}_P) \sum_{e\in P} b_e + a_2 \sum_P f_P^{\text{opt}} \sum_{e\in P} a_e f_e^{\text{opt}} + \sum_P f_P^{\text{opt}} \sum_{e\in P} b_e. \end{aligned} \tag{10}$$

Let

$$\Delta := -a_1 \sum_P (f_P^* + \bar{f}_P) \sum_{e\in P} b_e + \sum_P f_P^{\text{opt}} \sum_{e\in P} b_e.$$

Then (10) can be written as

$$(1 - a_1) C_{\text{eq}} \leq a_2 \sum_P f_P^{\text{opt}} \sum_{e\in P} a_e f_e^{\text{opt}} + \Delta. \tag{11}$$

Since we have assumed that $a_1 \geq 0$ and $b_e \geq 0$, for all $e \in E$, we get that

$$\Delta \leq -a_1 \sum_P \bar{f}_P \sum_{e\in P} b_e + \sum_P f_P^{\text{opt}} \sum_{e\in P} b_e. \tag{12}$$

By the definition of SCALE, we have that on every P, $\bar{f}_P = \alpha f_P^{\mathrm{opt}}$. Therefore

$$\Delta \le (1 - a_1\alpha) \sum_P f_P^{\mathrm{opt}} \sum_{e \in P} b_e. \tag{13}$$

From (11), (13), if we require that $a_1 \le 1$, we have that

$$\frac{C_{\mathrm{eq}}}{C_{\mathrm{opt}}} \le \frac{\max\{a_2, 1 - \alpha a_1\}}{1 - a_1}.$$

To obtain the best possible price of anarchy we solve the program:

$$\begin{aligned} \min \frac{\max\{a_2, 1 - \alpha a_1\}}{1 - a_1} \quad & \text{s.t.} \\ a_1 a_2 \ge \frac{1}{4}, & \\ a_1 \le 1, & \\ a_1, a_2 \ge 0. & \end{aligned}$$

By setting

$$a_1 := \frac{1 - \sqrt{1 - \alpha}}{2\alpha}, \qquad a_2 := \frac{1 + \sqrt{1 - \alpha}}{2}$$

all constraints are satisfied (note that $a_1 \le \frac{1}{2}$), the two expressions in the max of the objective function become equal, and Theorem 1 follows. □

A Nearly Tight Example for SCALE Consider the graph of the Braess paradox (Fig. 1). This is a directed graph with four vertices s, t, u, v and five edges (s, u), (u, t), (u, v), (s, v), (v, t). There is a single commodity to be routed from s to t of total demand 1. We set the latency of edge (u, v) to be identically equal to zero. For the other edges we define a latency function $l(x)$ which is parameterized

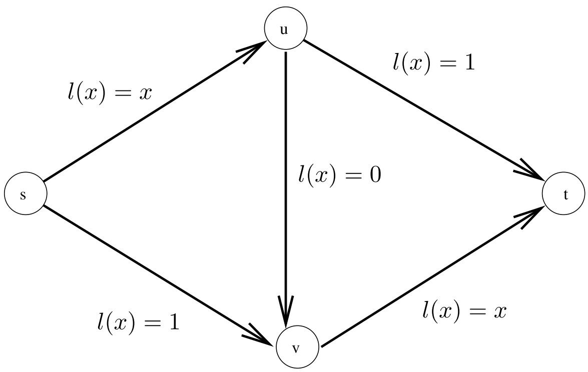

Fig. 1 The Braess paradox instance

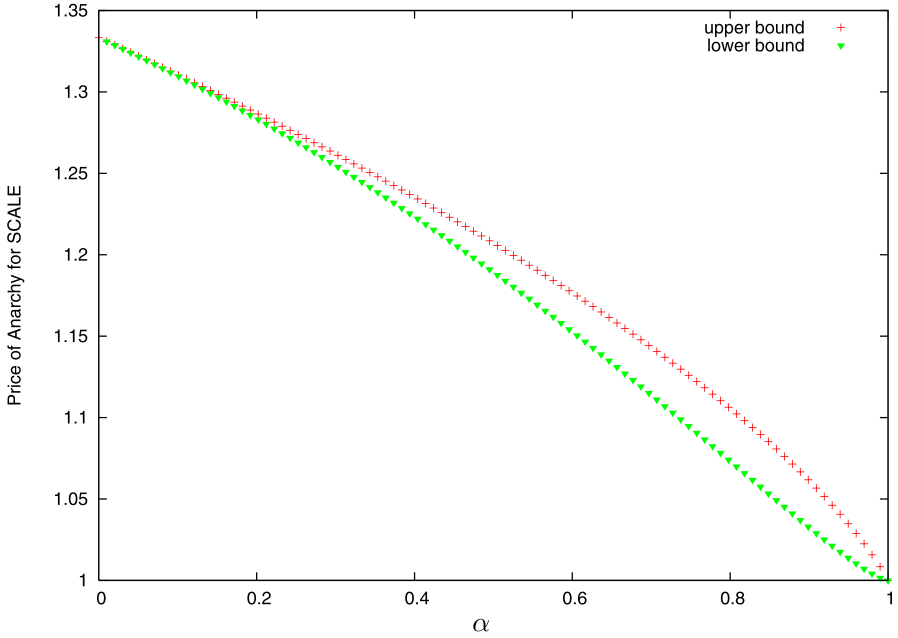

Fig. 2 Our upper and lower bounds for SCALE, as obtained in Sect. 3 plotted as functions of the fraction α of the users that are coordinated

by α. For (s,u), (v,t) the latency is $\frac{\alpha+2\sqrt{1-\alpha}}{2-\alpha-2\sqrt{1-\alpha}}x$, and the remaining two edges have latency $x+\frac{2\sqrt{1-\alpha}}{2-\alpha-2\sqrt{1-\alpha}}$. One can verify that in the optimal solution the upper and lower paths carry flow $1/2$ each, therefore $C_{\mathrm{opt}}=\frac{1}{2}+\frac{\alpha+6\sqrt{1-\alpha}}{2(2-\alpha-2\sqrt{1-\alpha})}$. In the Stackelberg equilibrium the coordinated users push $\alpha/2$ units of flow along each of the paths $s-u-t$ and $s-v-t$. The selfish users push $1-\alpha$ units of flow along the path $s-u-v-t$. The resulting price of anarchy is $\frac{2\alpha-\alpha^2-2\alpha\sqrt{1-\alpha}+4\sqrt{1-\alpha}}{1+2\sqrt{1-\alpha}}$, and this lower bound is at most an additive factor of 0.0323 away from our upper bound (this maximum gap happens for $\alpha=0.81\ldots$). See Fig. 2.

Recall the quantity δ we defined earlier. In the example just produced, one can verify that the fraction of C_{opt} that is paid on the heavy edges (s,u), (v,t) is $\frac{\alpha+2\sqrt{1-\alpha}}{2+4\sqrt{1-\alpha}}$, which for every $\alpha\in[0,1]$ is less than $1/2$. Moreover, in the case where δ is at least some constant fraction, one can modify the proof of Theorem 1 to obtain an improved price of anarchy. In particular, the right-hand side of (11) can be written as a sum of two parts, one for the heavy edges and one for the light ones. The use of the additive model makes easy the transition from a sum over paths to a sum over edges and vice versa.

$$(1-a_1)C_{\mathrm{eq}}\le a_2\sum_{e\ \mathrm{heavy}}a_e(f_e^{\mathrm{opt}})^2+\Delta_h+a_2\sum_{e\ \mathrm{light}}a_e(f_e^{\mathrm{opt}})^2+\Delta_l, \qquad (14)$$

where

$$\Delta_h := -a_1 \sum_{e\ \text{heavy}} (f_e^* + \bar{f}_e) b_e + \sum_{e\ \text{heavy}} f_e^{\text{opt}} b_e$$

and

$$\Delta_l := a_1 \sum_{e\ \text{light}} (f_e^* + \bar{f}_e) b_e + \sum_{e\ \text{light}} f_e^{\text{opt}} b_e.$$

For the contribution of the light edges to the right-hand side of (14) one proceeds as before by using that $f_e^* \geq 0$, while for the contribution of the heavy edges one uses that $f_e^* > c\bar{f}_e$. The improved upper bound is obtained by minimizing $\frac{\max\{a_2, 1-a_1(\alpha c+\alpha)\}\delta + \max\{a_2, 1-\alpha a_1\}(1-\delta)}{1-a_1}$ subject to the constraints $a_1 a_2 \geq \frac{1}{4}$, $a_1 \leq 1, a_1, a_2 \geq 0$. The quantity $c > 0$ is the one from the definition of the light edges.

By the preceding arguments and the plot in Fig. 2 we conclude that SCALE hits its worst-case performance when δ is rather small. This occurs when the optimal solution pays most of its cost on the light edges, i.e., edges that are underutilized by the selfish users. Natural as this insight is, it appears to contradict the bound of Lemma 1 which was obtained by the edge-by-edge approach. The apparent contradiction is resolved when one notices that the bound of Lemma 1 is a weak one: even for $\delta = 0$, it is larger than the bound of Theorem 1 for all $\alpha \leq 0.919$ and becomes only marginally better for larger values of α.

An Upper Bound for Strong LLF Let a path be *good* if it is used by the coordinated users as dictated by strong LLF. Therefore, path P is good iff $\bar{f}_P > 0$. Paths that are not good are called *bad*. There is a $\lambda \in [0, 1]$ such that $\sum_{P\ \text{bad}} f_P^{\text{opt}} \times [\sum_{e\in P}(a_e f_e^{\text{opt}} + b_e)] = (1-\lambda)C_{\text{opt}}$ and $\sum_{P\ \text{good}} f_P^{\text{opt}}[\sum_{e\in P}(a_e f_e^{\text{opt}} + b_e)] = \lambda C_{\text{opt}}$.

Theorem 2 *Let λ be defined as above. Then for general multicommodity networks with linear latency functions and a fraction α of users coordinated by the strong LLF policy, the price of anarchy is bounded as follows*:

$$\frac{C_{\text{eq}}}{C_{\text{opt}}} \leq \begin{cases} \dfrac{4}{3}, & \text{if } \lambda \in \left[0, \dfrac{1}{3}\right), \\ \dfrac{2(1-\lambda)^2}{2-\lambda-\sqrt{4\lambda - 3\lambda^2}}, & \text{if } \lambda \in \left[\dfrac{1}{3}, 1\right]. \end{cases}$$

Proof By decomposing the right hand side of (12) into two parts, one for the good and one for the bad paths, we get

$$\Delta \leq -a_1 \sum_{P\ \text{good}} \bar{f}_P \sum_{e\in P} b_e + \sum_{P\ \text{good}} f_P^{\text{opt}} \sum_{e\in P} b_e - a_1 \sum_{P\ \text{bad}} \bar{f}_P \sum_{e\in P} b_e + \sum_{P\ \text{bad}} f_P^{\text{opt}} \sum_{e\in P} b_e. \tag{15}$$

Under the LLF policy, all good paths P but one are saturated, meaning $f_P^{\text{opt}} = \bar{f}_P$. We can replace the offending path Π (i.e., the one on which $0 < \bar{f}_\Pi < f_\Pi^{\text{opt}}$) by two copies of the same path in the flow decomposition of f^{opt}, both with the same latency.

One copy gets flow $\bar{f}_\Pi$ out of a total of f_Π^{opt} and is included in the set of good paths, and the other copy gets the rest $f_\Pi^{\text{opt}} - \bar{f}_\Pi$ and is included in the bad paths. With this new path set, all good paths are saturated, i.e., $f_P^{\text{opt}} = \bar{f}_P$. All of the above can be seen as just a change of the set of indices used in the $\sum$ notation for the paths P of flow f^{opt}. We use this new set of indices (decomposition) of f^{opt} from now on. Then

$$-a_1 \sum_{P \text{ good}} \bar{f}_P \sum_{e\in P} b_e + \sum_{P \text{ good}} f_P^{\text{opt}} \sum_{e\in P} b_e = (1-a_1) \sum_{P \text{ good}} f_P^{\text{opt}} \sum_{e\in P} b_e,$$

and (15) becomes

$$\Delta \le (1-a_1) \sum_{P \text{ good}} f_P^{\text{opt}} \sum_{e\in P} b_e - a_1 \sum_{P \text{ bad}} \bar{f}_P \sum_{e\in P} b_e + \sum_{P \text{ bad}} f_P^{\text{opt}} \sum_{e\in P} b_e. \tag{16}$$

Recall that on a bad path P, $\bar{f}_P = 0$. If in addition we require that $a_2 \le 1$, (11), (16) yield

$$\begin{aligned}
(1-a_1)C_{\text{eq}} &\le a_2 \sum_{P} f_P^{\text{opt}} \sum_{e\in P} a_e f_e^{\text{opt}} + (1-a_1) \sum_{P \text{ good}} f_P^{\text{opt}} \sum_{e\in P} b_e + \sum_{P \text{ bad}} f_P^{\text{opt}} \sum_{e\in P} b_e \\
&\le \sum_{P \text{ bad}} f_P^{\text{opt}} \left[\sum_{e\in P} (a_e f_e^{\text{opt}} + b_e)\right] \\
&\quad + \max\{a_2, 1-a_1\} \sum_{P \text{ good}} f_P^{\text{opt}} \left[\sum_{e\in P} (a_e f_e^{\text{opt}} + b_e)\right]
\end{aligned}$$

which, in turn, implies that

$$\frac{C_{\text{eq}}}{C_{\text{opt}}} \le \frac{1-\lambda+\max\{a_2, 1-a_1\}\lambda}{1-a_1}. \tag{17}$$

We will pick a_1, a_2, subject to all the constraints on them we have assumed so far, so that we get the smallest possible upper bound on the price of anarchy from (17).

First we assume that $a_2 \ge 1 - a_1$ and therefore (17) implies that $\frac{C_{\text{eq}}}{C_{\text{opt}}} \le \frac{1-(1-a_2)\lambda}{1-a_1}$. Hence we would like to minimize $\frac{1-(1-a_2)\lambda}{1-a_1}$ subject to the constraints $a_2 \ge 1 - a_1, a_1 a_2 \ge \frac{1}{4}, 0 \le a_1, a_2 \le 1$. For the case $\lambda \in [\frac{1}{3}, 1]$ the minimum is achieved by picking $a_1 := \frac{\sqrt{4\lambda-3\lambda^2}-\lambda}{4(1-\lambda)}, a_2 := \frac{1}{4a_1}$, while for $\lambda \in [0, \frac{1}{3})$ the minimum is achieved by picking $a_1 := \frac{1}{4}, a_2 := 1$.

If we assume that $a_2 < 1 - a_1$, then we do not get a better upper bound. Therefore our analysis of LLF yields the upper bounds given in the statement of Theorem 2. □

Since LLF picks the most expensive paths of f^{opt} to saturate, and $\bar{f}$ satisfies a fraction α of the overall demand, we have that $\lambda \ge \alpha$ (note that in the definition of λ above, each flow path in the decomposition pays the latency of the path due to the whole flow through the edges of the path). The upper bound for the price of anarchy computed above is a decreasing function of λ, hence we can replace λ with α in it, and still get valid upper bounds that depend only on α. If $\lambda > \alpha$ our analysis yields a price of anarchy bound that is even better.

4 Plots

In this section, we provide various plots of the curves for the linear latency functions mentioned earlier. All the plots were obtained using Gnuplot. Our hard example of Sect. 3 was a modification of the Braess paradox instance with respect to the latency functions. The exact Braess paradox instance is defined in Fig. 1. One can easily verify that on the latter instance the price of anarchy curve of SCALE is $4/3-(1/3)(2\alpha-\alpha^2)$. For LLF, both paths used by the optimum solution have equal latency. Regardless of tie breaking, the curve of LLF is $4/3-(1/3)(2\alpha-2\alpha^2)$ for $\alpha \leq 1/2$ and $4/3-(4\alpha/3-2\alpha^2/3-1/3)$ for $\alpha > 1/2$.

Figure 2 shows the upper and lower bounds we obtained in Sect. 3. Figure 3 shows our LLF upper bound when $\lambda = 1$ in Theorem 2 and the corresponding lower bound obtained from the Braess paradox. Finally, Fig. 4 compares the existing lower bounds for the two policies on general networks, by drawing their performance on the Braess paradox defined in the previous paragraph. For comparison reasons, in the same plot we also give the performance of SCALE in our nearly tight example from Sect. 3. Note that the latter is a significantly stronger lower bound for SCALE than the lower bound obtained by the performance of the policy on the Braess paradox instance defined above.

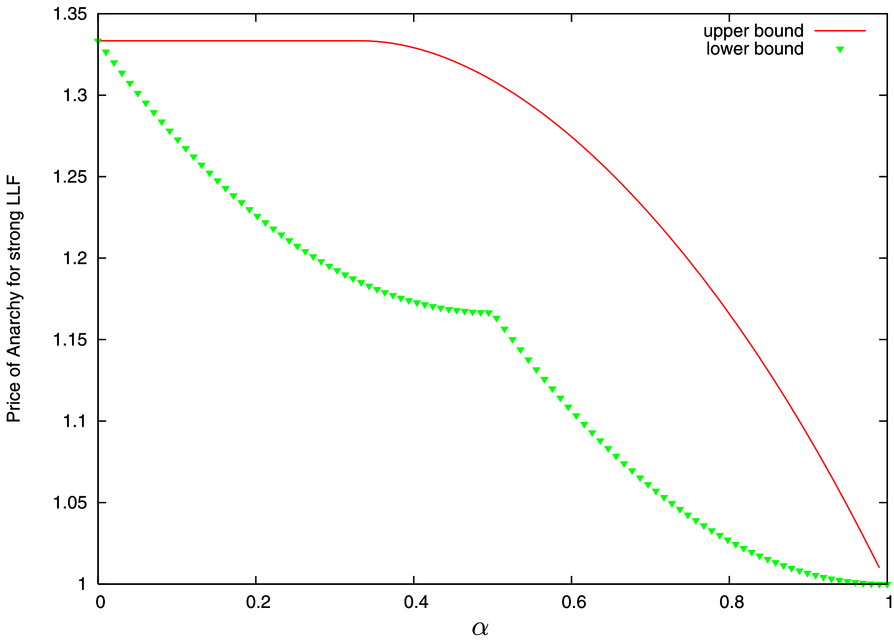

Fig. 3 Our upper bound for strong LLF, as obtained in Sect. 3, under the further assumption that $\lambda = \alpha$. The lower bound is the exact performance of LLF on the instance of the Braess paradox

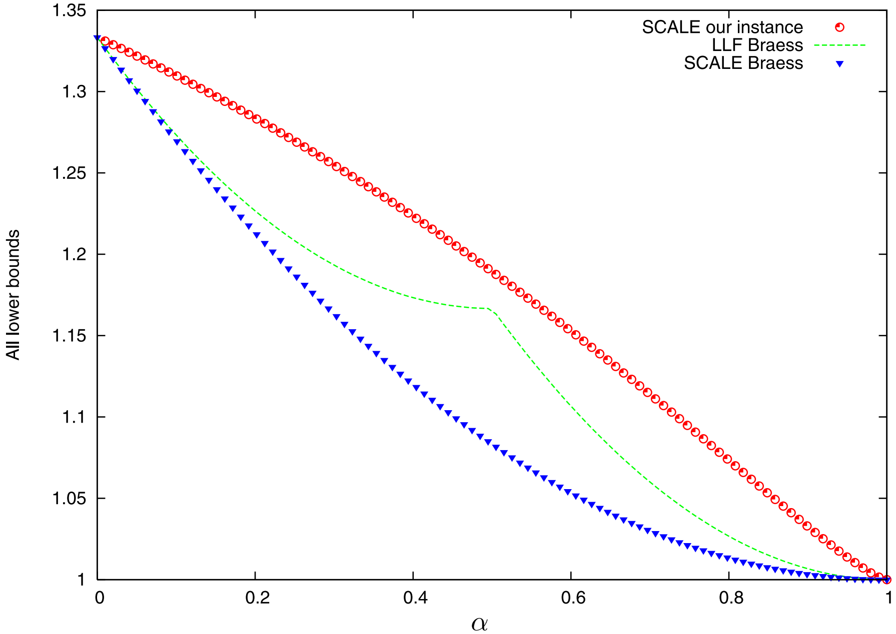

Fig. 4 The performance of the SCALE policy on our hard instance from Sect. 3 vs. the performance of the LLF and SCALE policies on the instance of the Braess paradox. Observe that SCALE outperforms LLF on the latter instance

It is worth remarking that there is a value of α for which our upper bound for SCALE from Sect. 3 is very close to the lower bound for both policies. For $\alpha = 1/2$ our upper bound is within an additive 0.027 factor from $7/6$ which is the performance of LLF on the Braess paradox instance.

5 General Latency Functions

The analysis of the linear case can be extended to general latency functions that satisfy certain properties. Recall the vector-valued function notation $L()$ for the latency function. According to Perakis [14], $L(x)$ satisfies the *Jacobian similarity property* if it has a positive semidefinite Jacobian matrix ($\nabla L(x) \succeq 0$, for every $x \in K$) and there exists constant $A \geq 1$ such that for all $w \in \mathbb{R}^m$, for all $x, \bar{x} \in K$

$$\frac{1}{A} w^T \nabla L(x) w \leq w^T \nabla L(\bar{x}) w \leq A w^T \nabla L(x) w.$$

The concept of Jacobian similarity originates from the Hessian similarity notion in interior point methods (see e.g., [22]). The value A is known to be independent of the matrix dimension, for positive definite $\nabla L(x)$. If the Jacobian matrix is *strongly positive definite*, i.e., it has eigenvalues bounded away from zero, then A is upper-bounded by

$$\frac{\max_{x\in K}\lambda_{\max}(S(x))}{\min_{x\in K}\lambda_{\min}(S(x))},$$

where $S(x)=\frac{\nabla L(x)+\nabla L(x)^T}{2}$. If $L(x)=Gx+b$ with $G\succeq 0$, then $A=1$ [14].

In our case, $\nabla L(x)$ is a diagonal matrix with the diagonal entry corresponding to edge e being equal to $\frac{dl_e(x_e)}{dx_e}$. Such a matrix is positive semidefinite if these derivatives are nonnegative for all $x\in K$. This is the natural and common assumption that the latency functions are increasing. Therefore our results below will assume only that the latency functions are increasing.

Generalizing the earlier remarks on the affine case we can abstract the following from Perakis [14]:

Lemma 4 ([14]) *If* (i) *for all edges e, $l_e(f_e)$ is a continuously differentiable function with $\frac{dl_e(f_e)}{df_e}\geq 0$, and $l_e(0)\geq 0$ for all $f\in K$ and* (ii) *the matrix $\nabla L(x)$ satisfies the Jacobian similarity property for some $A\geq 1$, then*

$$C_{\mathrm{eq}}\leq a_1A\sum_e(f_e^*+\bar{f}_e)[l_e(f_e^*+\bar{f}_e)-l_e(0)]+C_{\mathrm{opt}}$$
$$+(a_2-1)A\sum_e f_e^{\mathrm{opt}}[l_e(f_e^{\mathrm{opt}})-l_e(0)]$$

for any $a_1,a_2\geq 0$ that satisfy $a_1a_2\geq 1/4$.

We can define $Z:=-a_1A\sum_e(f_e^*+\bar{f}_e)l_e(0)+\sum_e f_e^{\mathrm{opt}}l_e(0)$, and the lemma yields that

$$(1-a_1A)C_{\mathrm{eq}}\leq[(a_2-1)A+1]\sum_e f_e^{\mathrm{opt}}[l_e(f_e^{\mathrm{opt}})-l_e(0)]+Z. \tag{18}$$

For the SCALE policy $Z\leq(1-\alpha a_1A)\sum_e f_e^{\mathrm{opt}}l_e(0)$, and therefore we can obtain that

$$(1-Aa_1)C_{\mathrm{eq}}\leq[(a_2-1)A+1]C_{\mathrm{opt}}-A(\alpha a_1+a_2-1)\sum_e f_e^{\mathrm{opt}}l_e(0)$$

under the conditions $a_1a_2\geq 1/4$, $a_1\leq 1/A$, $a_1,a_2\geq 0$. We distinguish two cases:

1. $\alpha a_1+a_2\geq 1$: In this case, we have that

$$(1-Aa_1)C_{\mathrm{eq}}\leq[(a_2-1)A+1]C_{\mathrm{opt}}.$$

Hence we are looking for a_1,a_2 that solve the following minimization problem:

$$\min\frac{Aa_2+1-A}{1-Aa_1}\quad \text{s.t.}$$
$$\alpha a_1+a_2\geq 1,$$

$$a_1 a_2 \geq \frac{1}{4},$$
$$a_1 \leq \frac{1}{A},$$
$$a_1, a_2 \geq 0.$$

If we set $a_2 = 1/4a_1$ then we must have that

$$a_1 \notin \left(\frac{1-\sqrt{1-\alpha}}{2\alpha}, \frac{1+\sqrt{1-\alpha}}{2\alpha}\right).$$

The objective function is increasing for

$$a_1 \in \left[\frac{A-\sqrt{A^2+4(1-A)}}{4(A-1)}, \frac{A+\sqrt{A^2+4(1-A)}}{4(A-1)}\right]$$

and decreasing for the other values of a_1 in $[0, \frac{1}{A})$. If $\frac{A-\sqrt{A^2+4(1-A)}}{4(A-1)} \leq \frac{1-\sqrt{1-\alpha}}{2\alpha}$, then we set $a_1 := \frac{A-\sqrt{A^2+4(1-A)}}{4(A-1)}$ otherwise we set $a_1 := \frac{1-\sqrt{1-\alpha}}{2\alpha}$.

If we set $a_2 = 1 - \alpha a_1$, then we must have

$$a_1 \in \left(\frac{1-\sqrt{1-\alpha}}{2\alpha}, \frac{1+\sqrt{1-\alpha}}{2\alpha}\right).$$

If $\frac{1-\sqrt{1-\alpha}}{2\alpha} \leq \frac{1}{A}$ then we set $a_1 := \frac{1-\sqrt{1-\alpha}}{2\alpha}$, else the problem is infeasible. So this case does not add something new to the previous bound.

2. $\alpha a_1 + a_2 \leq 1$: In this case, since $l_e(0) \leq l_e(f_e^{\text{opt}})$, for all $e \in E$, we have that

$$(1 - Aa_1)C_{\text{eq}} \leq [1 - A\alpha a_1]C_{\text{opt}}.$$

Hence we are looking for a_1, a_2 that solve the following minimization problem:

$$\min \frac{1 - A\alpha a_1}{1 - Aa_1} \quad \text{s.t.}$$
$$\alpha a_1 + a_2 \leq 1,$$
$$a_1 a_2 \geq \frac{1}{4},$$
$$a_1 \leq \frac{1}{A},$$
$$a_1, a_2 \geq 0.$$

This case produces the same bounds as the previous one.

Hence we get the following theorem for general (increasing) latency functions and the SCALE strategy:

Theorem 3 *Let $A \geq 1$ be the Jacobian similarity property parameter for the latency function matrix L, and α the coordinated fraction of flow that follows the SCALE strategy. Then the price of anarchy is upper-bounded as follows*:

$$\frac{C_{\text{eq}}}{C_{\text{opt}}} \leq \frac{A + 4(1-A)a_1}{4a_1(1-Aa_1)},$$

where

1. *If* $\frac{A-\sqrt{A^2+4(1-A)}}{4(A-1)} \leq \frac{1-\sqrt{1-\alpha}}{2\alpha}$ *then* $a_1 = \frac{A-\sqrt{A^2+4(1-A)}}{4(A-1)}$.
2. *If* $\frac{A-\sqrt{A^2+4(1-A)}}{4(A-1)} > \frac{1-\sqrt{1-\alpha}}{2\alpha}$ *then* $a_1 = \frac{1-\sqrt{1-\alpha}}{2\alpha}$.

Note that the bound of Theorem 3 depends only on the family of latency functions (through A) and α. Also observe that for A slightly greater than 1, part 1 of Theorem 3 holds. In that case, the bound depends only on A. For the case $A = 1$, it is easy to see that the analysis coincides with the analysis of the linear latency functions case.

6 The Effect of Tax Evasion on Networks

So far we have assumed that the network is subject to a central coordinating authority that can decide the routing of a fraction α of the overall traffic, while allowing the rest to act selfishly. In this section, we explore whether the same effects can be achieved when *no such central authority exists*, i.e., there is no notion of leader and follower in the Stackelberg sense. Instead we wish to use taxes (tolls) on the edges of the network assuming that all users are selfish but an α fraction of them are still law-abiding tax-paying citizens. The remaining $1-\alpha$ fraction does not believe in paying taxes. In this section, we show that such taxes do exist.

We assume throughout the section that the latency functions $l_e()$ are continuous, increasing and take only nonnegative values. The flow for every origin-destination pair (commodity) $i = 1, \dots, k$ of rate d_i in the network is split into two parts: $\bar{f}^i$ corresponds to the set of tax-payers with rate αd_i and f^i corresponds to the set of tax-evaders with rate $(1-\alpha)d_i$. The tax payers can be heterogeneous: they attach an *importance factor* $a(i) > 0$ to their disutility due to taxation. Let $f_e := \sum_i f_e^i$, $\bar{f}_e := \sum_i \bar{f}_e^i$, for all e. We are looking for the existence of nonnegative edge taxes b_e, for all $e \in E$, such that for every commodity i (i) the tax-paying users $\bar{f}^i$ perceive edge costs $l_e(f_e + \bar{f}_e) + a(i) \cdot b_e$, for all $e \in E$, (ii) the tax-evaders f^i perceive edge costs $l_e(f_e + \bar{f}_e)$, and (iii) the b_e's are such that the tax-payers are forced to implement the SCALE policy. The latter means that at the traffic equilibrium we must have

$$\sum_{i=1}^{k} \bar{f}_e^i = \alpha f_e^{\text{opt}}, \quad \forall e \in E. \tag{19}$$

A key observation is that if, in addition, we assume that all latency functions $l_e(\cdot)$ are strictly increasing, then conditions (19) are equivalent to

$$\sum_{i=1}^{k} \bar{f}_e^i \le \alpha f_e^{\mathrm{opt}}, \quad \forall e \in E. \tag{20}$$

We prove this similarly to Claim 1 in [10]. Assume for the sake of contradiction that some inequalities in (20) are strict. Define the flow $\bar{f}/\alpha$, by routing for each commodity i, flow equal to $\bar{f}^i/\alpha$. Then for every edge e, $\bar{f}_e/\alpha \le f_e^{\mathrm{opt}}$, and $l_e(\bar{f}_e/\alpha) \le l_e(f_e^{\mathrm{opt}})$. By nonnegativity,

$$(\bar{f}_e/\alpha)l_e(\bar{f}_e/\alpha) \le f_e^{\mathrm{opt}} l_e(f_e^{\mathrm{opt}}), \quad \forall e \in E.$$

If for some edge e, $\bar{f}_e/\alpha < f_e^{\mathrm{opt}}$, it follows that $0 \le l_e(\bar{f}_e/\alpha) < l_e(f_e^{\mathrm{opt}})$. Therefore for this particular edge

$$(\bar{f}_e/\alpha)l_e(\bar{f}_e/\alpha) < f_e^{\mathrm{opt}} l_e(f_e^{\mathrm{opt}}).$$

It follows that $\sum_{e\in E}(\bar{f}_e/\alpha)l_e(\bar{f}_e/\alpha) < \sum_{e\in E} f_e^{\mathrm{opt}} l_e(f_e^{\mathrm{opt}})$, which contradicts the optimality of f^{opt}.

We use the framework of [1, 10] to incorporate constraints (20) into a complementarity problem that describes the traffic equilibrium in our case. The complementarity problem (CP) is defined by constraints (21–31):

$$\bar{f}_P^i\left(\sum_{e\in P}\frac{l_e(f_e+\alpha f_e^{\mathrm{opt}})}{a(i)}+\sum_{e\in P} b_e-\bar{u}_i\right)=0 \quad \forall i,\ \forall P\in\mathcal{P}_i, \tag{21}$$

$$f_P^i\left(\sum_{e\in P} l_e(f_e+\alpha f_e^{\mathrm{opt}})-u_i\right)=0 \quad \forall i,\ \forall P\in\mathcal{P}_i, \tag{22}$$

$$\sum_{e\in P}\frac{l_e(f_e+\alpha f_e^{\mathrm{opt}})}{a(i)}+\sum_{e\in P} b_e \ge \bar{u}_i \quad \forall i,\ \forall P\in\mathcal{P}_i, \tag{23}$$

$$\sum_{e\in P} l_e(f_e+\alpha f_e^{\mathrm{opt}}) \ge u_i \quad \forall i,\ \forall P\in\mathcal{P}_i, \tag{24}$$

$$\bar{u}_i\left(\sum_{P\in\mathcal{P}_i}\bar{f}_P^i-\alpha d_i\right)=0 \quad \forall i, \tag{25}$$

$$u_i\left(\sum_{P\in\mathcal{P}_i} f_P^i-(1-\alpha)d_i\right)=0 \quad \forall i, \tag{26}$$

$$\sum_{P\in\mathcal{P}_i}\bar{f}_P^i \ge \alpha d_i \quad \forall i, \tag{27}$$

$$\sum_{P\in\mathcal{P}_i} f_P^i \ge (1-\alpha)d_i \quad \forall i, \tag{28}$$

$$b_e\left(\sum_i \bar{f}_e^i - \alpha f_e^{\mathrm{opt}}\right) = 0 \quad \forall e \in E, \tag{29}$$

$$\sum_i \bar{f}_e^i \le \alpha f_e^{\mathrm{opt}} \quad \forall e \in E, \tag{30}$$

$$f_P^i, \bar{f}_P^i, b_e, u_i, \bar{u}_i \ge 0 \quad \forall P, e, i. \tag{31}$$

To provide intuition we describe briefly the meaning of the constraints of (CP). For details the reader is referred to [1, 10]. (CP) defines two traffic equilibria that must be reached simultaneously. Constraints (21), (23), (25), (27), (29), (30) express the equilibrium problem for the tax-payers and constraints (22), (24), (26), (28) the equilibrium problem for the tax-evaders. Let us take a closer look at the equilibrium problem for the tax-payers. Constraints (21), (23) express the Wardrop principle. The variable $\bar{u}_i$ is the common disutility experienced by all tax-payers who belong to commodity i. Constraint (27) enforces that tax-payers in user class i satisfy rate at least αd_i. If they satisfy a rate strictly greater than αd_i, this comes for free since by Constraint (25) $\bar{u}_i$ must be zero. In [1] it is shown that these constraints form an exact formulation of the traffic equilibrium problem for the tax-payers. Constraints (30) enforce (20). The Lagrange multipliers b_e for (30), appearing in (29), will be the desired taxes. Similar considerations apply to the equilibrium problem for the tax-evaders except of course that for them there are no "capacity" constraints like (30).

To prove the existence of the tax vector b with the properties (i)–(iii) described above, it is enough to show that (CP) has a solution.

By using the fact that $\alpha f_e^{\mathrm{opt}}$ is a known constant for every edge e when f^{opt} is known, it follows from [1] that the complementarity problem (CP′) below with variables f_P^i, u_i has a solution (f^*, u^*):

$$\begin{aligned}
&f_P^i\left(\sum_{e\in P} l_e(f_e + \alpha f_e^{\mathrm{opt}}) - u_i\right) = 0 \quad \forall i,\ \forall P \in \mathcal{P}_i,\\
&\sum_{e\in P} l_e(f_e + \alpha f_e^{\mathrm{opt}}) \ge u_i \quad \forall i,\ \forall P \in \mathcal{P}_i,\\
&u_i\left(\sum_{P\in\mathcal{P}_i} f_P^i - (1-\alpha)d_i\right) = 0 \quad \forall i,\\
&\sum_{P\in\mathcal{P}_i} f_P^i \ge (1-\alpha)d_i \quad \forall i,\\
&f_P^i, u_i \ge 0 \quad \forall P, e, i.
\end{aligned} \tag{CP$'$}$$

In turn, by using the arguments from the proof of Theorem 2 in [10] we can show that the following complementarity problem (CP′) with variables $\bar{f}_P^i, b_e, \bar{u}_i$ is equivalent to pair of primal and dual linear programs and also has a solution $(\bar{f}^*, b^*, \bar{u}^*)$:

$$\bar{f}_P^i\left(\sum_{e\in P}\frac{l_e(f_e^* + \alpha f_e^{\mathrm{opt}})}{a(i)} + \sum_{e\in P} b_e - \bar{u}_i\right) = 0 \quad \forall i,\ \forall P \in \mathcal{P}_i,$$

$$\sum_{e\in P}\frac{l_e(f_e^*+\alpha f_e^{\mathrm{opt}})}{a(i)}+\sum_{e\in P}b_e\ge \bar{u}_i\quad \forall i,\ \forall P\in\mathcal{P}_i, \tag{CP$''$}$$

$$\bar{u}_i\left(\sum_{P\in\mathcal{P}_i}\bar{f}_P^i-\alpha d_i\right)=0\quad\forall i,$$

$$\sum_{P\in\mathcal{P}_i}\bar{f}_P^i\ge\alpha d_i\quad\forall i,$$

$$b_e\left(\sum_i\bar{f}_e^i-\alpha f_e^{\mathrm{opt}}\right)=0\quad\forall e\in E,$$

$$\sum_i\bar{f}_e^i\le\alpha f_e^{\mathrm{opt}}\quad\forall e\in E,$$

$$\bar{f}_P^i,b_e,\bar{u}_i\ge 0\quad\forall P,e,i.$$

Now it is clear that $(f^*,\bar{f}^*,b^*,u^*,\bar{u}^*)$ is a solution of (CP), and we can use taxes b_e^* on each edge e to induce the tax-payers to follow the SCALE policy. Then all our results about the effects of SCALE hold also for this setting. If the latency functions are non-strictly monotone, we have obtained that there exists one tax-induced equilibrium in which the tax-paying users implement the SCALE policy [1, 10]. If the latency functions are strictly monotone, in every tax-induced equilibrium the tax payers implement SCALE [1, 10]. We summarize our findings in the next theorem.

Theorem 4 *Let x_e be the total flow through edge e for some traffic assignment. If for all $e\in E$, the functions $l_e()$ are strictly monotone, and $l_e()\ge 0$, there is a $b\in\mathbb{R}_+^{|E|}$ such that if an α fraction of the users (called the tax-payers) experiences edge disutility*

$$l_e(x_e)+a(i)\cdot b_e,\quad\forall e\in E,$$

while the rest experience disutility $l_e(x_e)$, for all $e\in E$, then the tax payers induce in equilibrium the flow vector αf^{opt}. Here f^{opt} is the flow that minimizes the system cost $\sum_P x_P l_P(x)$.

7 Discussion

Perakis [14] derives the price of anarchy for non-separable asymmetric latency functions. Therefore our results from Sect. 3 are bound to extend to that setting as well.

There are several issues that are left open. Can one get a strictly decreasing curve for LLF? Moreover the difference between the upper and lower bounds for LLF is currently considerable. For SCALE it would be interesting to close the rather small gap that exists between our upper and lower bounds. It would be interesting also if one could determine the instances on which SCALE outperforms LLF and vice versa. Finally, and perhaps more importantly, is there an optimal Stackelberg strategy for general multicommodity networks?

References

1. Aashtiani, H.Z., Magnanti, T.L.: Equilibria on a congested transportation network. SIAM J. Algebr. Discret. Methods **2**, 213–226 (1981)
2. Başar, T., Olsder, G.J.: Dynamic Noncooperative Game Theory. SIAM, Philadelphia (1999)
3. Beckmann, M., McGuire, C.B., Winsten, C.B.: Studies in the Economics of Transportation. Yale University Press (1956)
4. Cole, R., Dodis, Y., Roughgarden, T.: Pricing network edges for heterogeneous selfish users. In: Proceedings of the 35th Annual ACM Symposium on Theory of Computing, pp. 521–530 (2003)
5. Correa, J.R., Stier Moses, N.E.: Personal communication (2006)
6. Correa, J.R., Schulz, A.S., Stier Moses, N.E.: Selfish routing in capacitated networks. Math. Oper. Res. **29**, 961–976 (2004)
7. Dafermos, S.C.: Traffic equilibrium and variational inequalities. Transp. Sci. **14**(1), 42–54 (1980)
8. Fleischer, L., Jain, K., Mahdian, M.: Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games. In: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science, pp. 277–285 (2004)
9. Kaporis, A.C., Spirakis, P.G.: The price of optimum in Stackelberg games on arbitrary single commodity networks and latency functions. In: Proceedings of the 18th Annual ACM Symposium on Parallelism in Algorithms and Architectures, pp. 19–28 (2006)
10. Karakostas, G., Kolliopoulos, S.G.: Edge pricing of multicommodity networks for heterogeneous selfish users. In: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science, pp. 268–276 (2004)
11. Karakostas, G., Kolliopoulos, S.G.: Stackelberg strategies for selfish routing in general multicommodity networks. Technical Report AdvOL2006/08, McMaster University (June 2006)
12. Korilis, Y.A., Lazar, A.A., Orda, A.: Achieving network optima using Stackelberg routing strategies. IEEE/ACM Trans. Netw. **5**, 161–173 (1997)
13. Koutsoupias, E., Papadimitriou, C.: Worst-case equilibria. In: Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science, pp. 404–413 (1999)
14. Perakis, G.: The price of anarchy when costs are non-separable and asymmetric. In: Proceedings of the 10th Conference on Integer Programming and Combinatorial Optimization, pp. 46–58 (2004)
15. Roughgarden, T.: Stackelberg scheduling strategies. SIAM J. Comput. **33**, 332–350 (2004). Conference version in STOC 2001
16. Roughgarden, T.: Selfish routing. Ph.D. thesis, Cornell University (2002)
17. Roughgarden, T.: The price of anarchy is independent of the network topology. J. Comput. Syst. Sci. **67**, 341–364 (2003). Special issue on STOC 2002
18. Roughgarden, T.: Selfish Routing and the Price of Anarchy. MIT Press, Cambridge (2005)
19. Roughgarden, T.: Selfish routing and the price of anarchy. Optima (2006)
20. Roughgarden, T., Tardos, É.: How bad is selfish routing? J. ACM **49**, 236–259 (2002)
21. Sharma, Y., Williamson, D.P.: Stackelberg thresholds in network routing games or the value of altruism. In: Proc. ACM Conference on Electronic Commerce, pp. 93–102 (2007)
22. Sun, J.: A convergence analysis for a convex version of Dikin's algorithm. Ann. Oper. Res. **62**, 357–374 (1996)
23. Swamy, C.: The effectiveness of Stackelberg strategies and tolls for network congestion games. In: Proc. 18th ACM-SIAM Symposium on Discrete Algorithms (2007)
24. Wardrop, J.G.: Some theoretical aspects of road traffic research. Proc. Inst. Civil Eng. Part II **1**, 325–378 (1952)
25. Yang, H., Huang, H.-J.: The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. Transp. Res. B **38**, 1–15 (2004)