MATHEMATICS OF OPERATIONS RESEARCH
Vol. 35, No. 2, May 2010, pp. 330–346
ISSN 0364-765X | EISSN 1526-5471 | 10 | 3502 | 0330

DOI 10.1287/moor.1100.0442

Stackelberg Routing in Arbitrary Networks

Vincenzo Bonifaci
Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany, bonifaci@mpi-inf.mpg.de

Tobias Harks
Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany,
harks@math.tu-berlin.de

Guido Schäfer
Center for Mathematics and Computer Science (CWI), Algorithms, Combinatorics and Optimization,
1098 XG Amsterdam, The Netherlands, g.schaefer@cwi.nl

We investigate the impact of *Stackelberg routing* to reduce the price of anarchy in network routing games. In this setting, an α fraction of the entire demand is first routed centrally according to a predefined *Stackelberg strategy* and the remaining demand is then routed selfishly by (nonatomic) players. Although several advances have been made recently in proving that Stackelberg routing can, in fact, significantly reduce the price of anarchy for certain network topologies, the central question of whether this holds true in general is still open. We answer this question negatively by constructing a family of single-commodity instances such that every Stackelberg strategy induces a price of anarchy that grows linearly with the size of the network. Moreover, we prove upper bounds on the price of anarchy of the largest-latency-first (LLF) strategy that only depend on the size of the network. Besides other implications, this rules out the possibility to construct constant-size networks to prove an unbounded price of anarchy. In light of this negative result, we consider bicriteria bounds. We develop an efficiently computable Stackelberg strategy that induces a flow whose cost is at most the cost of an optimal flow with respect to demands scaled by a factor of $1+\sqrt{1-\alpha}$. Finally, we analyze the effectiveness of an easy-to-implement Stackelberg strategy, called SCALE. We prove bounds for a general class of latency functions that includes polynomial latency functions as a special case. Our analysis is based on an approach that is simple yet powerful enough to obtain (almost) tight bounds for SCALE in general networks.

Key words: network routing games; Stackelberg routing; inefficiency of equilibria
MSC2000 subject classification: Primary: 91A10, 68W01; secondary: 91A13, 68W40
OR/MS subject classification: Primary: games, nonatomic; secondary: analysis of algorithms
History: Received November 27, 2008; revised July 24, 2009. Published online in *Articles in Advance* April 30, 2010.

1. Introduction. Over the past years, the impact of the behavior of selfish, uncoordinated users in congested networks has been investigated intensively in the theoretical computer science and operations research literature. In this context, *network routing games* have proved to be an appropriate means of modeling selfish behavior in networks. The basic idea is to model the interaction between the selfish network users as a *noncooperative game*. We are given a directed graph with latency functions on the arcs and a set of origin-destination pairs, called *commodities*. Every commodity has a *demand* associated with it that specifies the amount of flow that needs to be sent from the respective origin to the destination. We assume that every demand represents a large population of players, each controlling an infinitesimal small amount of flow of the entire demand (such players are also called *nonatomic*). The latency that a player experiences to traverse an arc is given by a (nondecreasing) function of the total flow on that arc. We assume that every player acts selfishly and routes his flow along a minimum-latency path from its origin to the destination; this corresponds to a common solution concept for noncooperative games, that of a *Nash equilibrium* (here *Nash* or *Wardrop flow*; see Wardrop [37]). In a Nash flow, no player can improve his own latency by unilaterally switching to another path.

It is well-known that Nash equilibria can be *inefficient* in the sense that they need not achieve socially desirable objectives (Braess [3], Dubey [10]). In the context of network routing games, a Nash flow in general does not minimize the total cost; said differently, selfish behavior may cause a performance degradation in the network. Koutsoupias and Papadimitriou [22] initiated the investigation of the efficiency loss caused by selfish behavior. They introduced a measure to quantify the inefficiency of Nash equilibria, which they termed the *price of anarchy*. The price of anarchy is defined as the worst-case ratio of the cost of a Nash equilibrium over the cost of a system optimum. In recent years, considerable progress has been made in quantifying the degradation in network performance caused by the selfish behavior of noncooperative network users. In a seminal work, Roughgarden and Tardos [32] showed that the price of anarchy for network routing games with nonatomic players and linear latency functions is 4/3; in particular, this bound holds independently of the underlying network topology. The case of more general families of latency functions has been studied by Roughgarden [27] and Correa et al. [7]. (For an overview of these results, we refer to the book by Roughgarden [30].) Despite these bounds for specific classes of latency functions, it is known that the price of anarchy for general latency functions is unbounded even on simple parallel-arc networks (Roughgarden and Tardos [32]).

Because of this large efficiency loss, researchers have proposed different approaches to reduce the price of anarchy in network routing games. One of the most prominent approaches is the use of *Stackelberg routing* (Korilis et al. [21], Roughgarden [28]). In this setting, it is assumed that a fraction $\alpha \in [0, 1]$ of the entire demand is controlled by a central authority, termed *Stackelberg leader*, while the remaining demand is controlled by the selfish nonatomic players, called the *followers*. In a *Stackelberg game*, the Stackelberg leader first routes the centrally controlled flow according to a predetermined policy, called the *Stackelberg strategy*, and then the remaining demand is routed by the selfish followers. The aim is to devise Stackelberg strategies so as to minimize the price of anarchy of the resulting combined flow with respect to the optimal solution for the entire demand.

Although Roughgarden [28] showed that computing the *best* Stackelberg strategy, i.e., one that minimizes the price of anarchy of the induced flow, is NP-hard even for parallel-arc networks and linear latency functions, several advances have been made recently in proving that Stackelberg routing can indeed significantly reduce the price of anarchy in network routing games. A well-studied Stackelberg strategy is the *largest-latency-first* (*LLF*) strategy. Intuitively, LLF tries to save the part of an optimal flow that is unattractive for the selfish followers by sending flow along paths of large latencies. More precisely, LLF computes an optimal flow for the entire demand and orders the paths that carry a positive amount of flow by nonincreasing latencies. According to this order, it then iteratively sends as much flow as possible along these paths (not exceeding the optimal flow value) until an α fraction of the demand has been routed.

Roughgarden [28] showed that for parallel-arc networks, the LLF strategy reduces the price of anarchy to $1/\alpha$ *independently* of the latency functions. That is, even if the Stackelberg leader controls only a small constant fraction of the overall demand, the price of anarchy reduces to a constant (while it is unbounded in the absence of any centralized control). More recently, Swamy [36] obtained a similar result for single-commodity, series-parallel networks and Fotakis [12] for parallel-arc networks and unsplittable flows. Despite these positive results, a central question regarding the effectiveness of Stackelberg routing was still open: Does every routing game admit a Stackelberg strategy inducing a bounded price of anarchy? More precisely, is there a function $g(\cdot)$ such that, for any Stackelberg routing game, there is a Stackelberg strategy inducing a flow with cost at most $g(\alpha)$ times the cost of the optimal flow? This question has been posed explicitly by Roughgarden [26, Open Problem 4].

Besides these efforts, researchers have also tried to characterize the effectiveness of easy-to-implement Stackelberg strategies for specific classes of latency functions. One of the simplest Stackelberg strategies is SCALE (see also Roughgarden [28]), which simply computes an optimal flow for the entire demand and then scales this flow down by α. The current best-known bound for the price of anarchy induced by SCALE on multicommodity networks and linear latency functions is credited to Karakostas and Kolliopoulos [18]. More recently, Swamy [36] derived the first general bounds for polynomial latency functions.

1.1. Our results. We investigate the impact of Stackelberg routing to reduce the price of anarchy in network routing games with nonatomic players. Our contributions are the following:

(i) We show that there exists a family of single-commodity networks for which every Stackelberg strategy induces a price of anarchy of $\Omega(k)$, where k is a parameter that represents the size of the network. By increasing the size of the network, we can thus show that the price of anarchy is unbounded. The result holds independently of the fraction $\alpha \in (0, 1)$ of the centrally controlled demand. This settles the open question raised by Roughgarden [26].

(ii) We prove that for every fixed α, the price of anarchy for the LLF strategy is bounded by $O(b(n, m, k))$, where $b(n, m, k)$ is some function depending on the number of vertices, arcs, and commodities of the network, both for single-commodity and multicommodity networks. This complements the negative result above and shows that no small (i.e., constant-size) networks exist that enable to prove an unbounded price of anarchy. These are also the first upper bounds for a Stackelberg strategy that hold for both *arbitrary* networks and *arbitrary* latency functions.

(iii) In light of our negative result, we investigate the effectiveness of Stackelberg routing strategies compared to an optimum flow for a larger demand, i.e., we consider bicriteria bounds. We develop an efficiently computable Stackelberg strategy inducing a flow whose cost is at most the cost of an optimal flow with respect to demands increased by a factor of $1 + \sqrt{1 - \alpha}$.

(iv) We give upper bounds on the efficiency of SCALE for a general class of latency functions, which, among others, contains polynomial latency functions with nonnegative coefficients. We also derive the first tight lower bounds for SCALE. Our bound is tight for concave latency functions; for higher-degree polynomials, our bounds are almost tight (though there remains a small gap for small values of α). Our results also imply that for concave

latency functions and general networks, SCALE achieves an approximation guarantee of less than 1.12 with respect to the best Stackelberg strategy (which is NP-hard to compute).

1.2. Significance. Our negative result settles an important open question regarding the applicability of Stackelberg routing in general networks. Though most existing results show that the performance degradation because of the absence of central control is *independent* of the underlying network topology, our results shows that the network topology matters in the context of Stackelberg routing. On one hand, we present a family of instances that show that the price of anarchy of every Stackelberg strategy is unbounded if we are allowed to increase the size of the network arbitrarily. On the other hand, we prove that the price of anarchy for LLF is bounded in terms of the size of the input network. Besides these structural insights, our negative result also has an impact on several other related settings outlined here.

A basic assumption inherent in almost all network routing games that have been studied in the past is that players are entirely selfish. However, experiments in economics show that this assumption is too simplistic in many scenarios (see also Chen and Kempe [4] and the references therein). In Stackelberg routing games, we abandon this assumption (at least partially) because we assume that only a fraction of the players is selfish while the other players may behave *arbitrarily*. Note that the behavior of the nonselfish players can be seen as a potential Stackelberg strategy. As a consequence, our negative result also carries over to these settings.

Most notably in this context is the very recent work by Chen and Kempe [4], who introduce a new network routing game with nonatomic players that is capable to model the players' degree of altruism. Every player i has an *altruism level* β_i and the utility function is a linear combination of a selfish part (player i's latency) and an altruistic part (the average latency of all players). By varying β_i from 1 to 0 to -1, player i's degree of altruism ranges from altruistic to selfish to spiteful, respectively. The authors show, among other results, that if all players have a uniform altruism level of $\beta > 0$, i.e., there are no entirely selfish players, then the price of anarchy is bounded by $1/\beta$ for arbitrary networks and semiconvex latency functions. On the other hand, our negative result implies that if the players who are entirely selfish ($\beta_i = 0$) only control a nonzero fraction of the overall demand, then the price of anarchy is unbounded, even for single-commodity networks and independently of the altruism levels of the nonselfish players ($\beta_i \neq 0$). In fact, based on this negative result, Chen and Kempe [4] restrict their analysis of the price of anarchy for nonuniform altruism levels to parallel-arc networks.

Fotakis [12] and Harks [14] studied Stackelberg routing for atomic congestion games and atomic splittable network games, respectively. Our lower-bound construction can be easily adapted to the unsplittable flow setting as well as to the atomic splittable case. Thus, it follows that even for symmetric congestion games (with or without fractional assignments), there exist no Stackelberg strategies inducing a bounded price of anarchy.

There are numerous applications that can be interpreted as a Stackelberg routing game. Here, we focus on highlighting only one of them: the routing of Internet traffic within the domain of an Internet service provider (see also Sharma and Williamson [33]). Here, the Internet service provider centrally controls a fraction of the overall traffic traversing its domain. In this setting, our second result provides the Internet service provider with an efficient algorithm to route the centrally controlled traffic. The performance of this routing algorithm is characterized by a smooth trade-off curve that scales between the absence of centralized control (doubling the demands is sufficient) and completely centralized control (no scaling is necessary). Additionally, our result has a nice interpretation for the class of (practical relevant) M/M/1-latency functions that model arc capacities: In order to beat the cost of an optimal flow, it is sufficient to scale all arc capacities by $1 + \sqrt{1-\alpha}$. Our bound is a natural generalization of the bicriteria bound by Roughgarden and Tardos [32] for the entirely selfish setting (see Correa et al. [8] for other related results).

1.3. Techniques. In order to prove that the price of anarchy of every Stackelberg strategy is unbounded, we construct a family of network instances. The crucial insight that we exploit in the multicommodity case is that one can devise a graph topology and corresponding latency functions such that for every commodity whose demand is not entirely controlled by the Stackelberg leader, the selfish followers have an incentive to harm some other players by inducing a constant latency on their path (the latency along this path would be zero otherwise). Because no Stackelberg leader can control all the commodities (assuming $\alpha \neq 1$), we can ensure that the total cost induced by the followers grows with the number of commodities. We believe that these ideas might turn out to be useful in order to prove negative results in other settings that involve selfish behavior. Our single-commodity instance simulates the multicommodity instance by introducing a supersource and supersink that are connected to the origins and destinations of the commodities, respectively. In order to control the amount of flow that is routed through every commodity, we tailor the latency functions so as to mimic capacities on these arcs.

We also show that the LLF strategy induces a price of anarchy that is bounded by $O(\alpha^{-1} \cdot b(n, m, k))$, where $b(n, m, k)$ is a function that depends on the number of vertices, arcs, and commodities of the network. In order to prove this, we bound the price of anarchy of LLF in terms of the worst-case ratio between the maximum latency that a selfish follower experiences if the followers are routed according to a Nash flow and the maximum latency that a follower experiences if the followers are routed according to an arbitrary flow. To the best of our knowledge, this relation has not been observed before and might be of independent interest. Our upper bounds then simply follow from existing results, characterizing the ratio of the largest latency in a Nash flow and that of a flow that minimizes the maximum latency (Correa et al. [9], Lin et al. [24], Roughgarden [29]).

We introduce a general approach, which we term λ-*approach*, to prove upper bounds on the price of anarchy of Stackelberg strategies for specific classes of latency functions. This approach is simple yet powerful enough to obtain (almost) tight bounds for SCALE in general networks. For polynomial latency functions, our approach yields upper bounds that significantly improve the currently best bounds by Swamy [36]. For linear latency functions, we derive an upper bound that coincides with a previous bound of Karakostas and Kolliopoulos [18]. Their analysis is based on a (rather involved) machinery presented in Perakis [25]. However, our analysis is much simpler; in particular, we do not rely on the machinery in Perakis [25]. Moreover, we show that this bound also holds for concave latency functions. A number of real-world problems may be formulated as network flow problems involving concave latency functions. Cost functions of this type are useful when dealing with network routing problems in the presence of economy of scale (see Gallo et al. [13]). We present a generalized Braess instance that shows that for the concave case, our bound is tight; a similar instance can be used to show that for higher-degree polynomials with nonnegative coefficients, our bounds are almost tight and leave only a small gap for small values of α. We are confident that our λ-approach will prove useful to derive upper bounds on the price of anarchy in other settings. For instance, the λ-approach can be applied to prove upper bounds when flows are unsplittable. So far, such upper bounds for general networks are only known for linear latency functions (see Fotakis [12]).

1.4. Related work. The idea of using Stackelberg strategies to improve the performance of a system was first proposed by Korilis et al. [21], who identified necessary and sufficient conditions for the existence of Stackelberg strategies that induce a system optimum; their model differs from the one discussed here. Roughgarden [28] first formulated the problem and model considered here. Roughgarden [28] also proposed some natural Stackelberg strategies such as SCALE and LLF. For parallel-arc networks, he showed that the price of anarchy for LLF is bounded by $4/(3 + \alpha)$ and $1/\alpha$ for linear and arbitrary latency functions, respectively. Both bounds are tight. He also showed that for certain types of Stackelberg strategies, which he termed *weak* strategies (see §2 for a definition), the price of anarchy for multicommodity networks can be unbounded (Roughgarden [28]). However, this did not rule out the existence of effective Stackelberg strategies in general. Moreover, Roughgarden [28] also proved that it is NP-hard to compute the best Stackelberg strategy. Kumar and Marathe [23] investigated approximation schemes to compute the best Stackelberg strategy and gave a polynomial-time approximation scheme for the case of parallel-arc networks.

Karakostas and Kolliopoulos [18] proved upper bounds on the price of anarchy for SCALE and LLF. Their bounds hold for arbitrary multicommodity networks and linear latency functions. Their analysis is based on a result obtained by Perakis [25] to bound the price of anarchy for network routing games with asymmetric and nonseparable latency functions. Furthermore, Karakostas and Kolliopoulos [18] showed that their analysis for SCALE is almost tight. More recently, Swamy [36] obtained upper bounds on the price of anarchy for SCALE and LLF for polynomial latency functions. Swamy [36] also proved a bound of $1 + 1/\alpha$ for single-commodity, series-parallel networks with arbitrary latency functions. Fotakis [12] studied LLF and a randomized version of SCALE for the case of unsplittable flows. He proved upper and lower bounds on the price of anarchy for linear latency functions. For parallel-arc networks, Fotakis [12] proved that LLF still achieves an upper bound of $1/\alpha$ for arbitrary latency functions in this case.

Correa and Stier-Moses [6] proved, besides some other results, that the use of *opt-restricted strategies*, i.e., strategies in which the Stackelberg leader sends no more flow on every arc than the system optimum, does not increase the price of anarchy. Sharma and Williamson [33] considered the problem of determining the smallest value of α such that the price of anarchy can be improved. They obtained results for parallel-arc networks and linear latency functions. Kaporis and Spirakis [16] studied a related question of finding the minimum demand that the Stackelberg leader needs to control in order to enforce an optimal flow.

Another prominent way to reduce the price of anarchy in nonatomic network routing games is the use of nonnegative tolls on arcs of the network. In the area of transportation networks, this concept has been called *congestion toll pricing* (see, for example, Knight [19], Beckmann et al. [2], Smith [35], and Hearn and

Ramana [15]). This mechanism assigns tolls to certain arcs of the network, which are charged to those users that decide to take routes through them. If users value latency relative to toll the same, Beckmann et al. [2] showed that charging users the difference between the marginal cost and the real cost in the socially optimal solution (marginal cost pricing) leads to an equilibrium flow that is optimal. Cole et al. [5] considered the case of heterogeneous users, that is, users value latency relative to cost differently. For single-commodity networks, Cole et al. [5] showed the existence of tolls that induce an optimal flow as Nash flow. Finally, Fleischer et al. [11], Karakostas and Kolliopoulos [17], and Yang and Huang [38] proved that there are tolls inducing an optimal flow for heterogenous users even in general networks.

2. Model and notation. In a network routing game, we are given a directed network $G = (V, A)$ and k origin-destination pairs $(s_1, t_1), \ldots, (s_k, t_k)$ called *commodities*. We let n and m refer to the number of vertices and arcs of G, respectively. For every commodity $i = 1, 2, \ldots, k$, a demand $r_i > 0$ is given that specifies the amount of flow with origin s_i and destination t_i. The interpretation here is that r_i corresponds to a large population of nonatomic players, each controlling an infinitesimally small amount of the entire demand that needs to be sent from s_i to t_i. Let $\mathcal{P}_i$ be the set of all paths from s_i to t_i in G and let $\mathcal{P} = \cup_i \mathcal{P}_i$. A *flow* is a function $f\colon \mathcal{P} \to \mathbb{R}_+$. The flow f is *feasible* (with respect to r) if for all i, $\sum_{P\in\mathcal{P}_i} f_P = r_i$. For a given flow f, we define the flow on an arc $a \in A$ as $f_a = \sum_{P\ni a} f_P$.

Moreover, each arc $a \in A$ has an associated variable *latency* $l_a\colon \mathbb{R}_+ \to \mathbb{R}_+$. For each $a \in A$, the latency function l_a is assumed to be nondecreasing and differentiable. If not indicated otherwise, we assume that $xl_a(x)$ is a convex function of x. Such functions are called *standard* (Roughgarden [27]). The latency of a path P with respect to a flow f is defined as the sum of the latencies of the arcs in the path, denoted by $l_P(f) = \sum_{a\in P} l_a(f_a)$. The triple (G, r, l) is called an *instance*.

We assume that every nonatomic player aims at routing his flow along a path that has minimum latency. Informally, a *Nash flow* (or *selfish flow*) is a feasible flow such that no player has an incentive to unilaterally deviate from his path. More formally, a feasible flow f is a Nash flow if for every $i = 1, 2, \ldots, k$ and $P, P' \in \mathcal{P}_i$ with $f_P > 0$, $l_P(f) \le l_{P'}(f)$. That is, all $s_i - t_i$ paths to which f assigns a positive amount of flow are paths of minimum latency; in particular, these paths have equal latency. The *cost* of a flow f is $C(f) = \sum_{P\in\mathcal{P}} f_P l_P(f)$. Equivalently, $C(f) = \sum_{a\in A} f_a l_a(f_a)$. It is well-known that if f and f' are Nash flows for the same instance, then $C(f) = C(f')$ (see, e.g., Roughgarden and Tardos [32]). A feasible flow of minimum cost is called *optimal* and denoted by o.

In a Stackelberg network game, we are given, in addition to G, r, and l, a parameter $\alpha \in (0, 1)$. A (*strong*) *Stackelberg strategy* (Roughgarden [28]) is a flow g feasible with respect to $r' = (\alpha_1 r_1, \ldots, \alpha_k r_k)$ for some $\alpha_1, \ldots, \alpha_k \in [0, 1]$ such that $\sum_{i=1}^k \alpha_i r_i = \alpha \sum_{i=1}^k r_i$. If $\alpha_i = \alpha$ for all i, g is called a *weak Stackelberg strategy* (Roughgarden [30]). Thus, both strong and weak strategies route a fraction α of the overall traffic but a strong strategy can choose how much flow of each commodity is centrally controlled. For single-commodity networks the two definitions coincide. A Stackelberg strategy g is called *opt restricted* if $g_a \le o_a$ for all $a \in A$.

Given a Stackelberg strategy g, let $\tilde{l}_a(x) = l_a(g_a + x)$ for all $a \in A$ and let $\tilde{r} = r - r'$. Then, a flow h is called *a Nash flow induced by* g if it is a Nash flow for the instance $(G, \tilde{r}, \tilde{l})$. Smith [34, Equation (9)] has proved that the Nash flow h can be characterized by the following *variational inequality*: h is a Nash flow induced by g if and only if for all flows x feasible with respect to $\tilde{r}$, $\sum_{a\in A} h_a \tilde{l}_a(h_a) \le \sum_{a\in A} x_a \tilde{l}_a(h_a)$ or, equivalently,

$$\sum_{a\in A} h_a l_a(g_a + h_a) \le \sum_{a\in A} x_a l_a(g_a + h_a). \tag{1}$$

We will mainly be concerned with the cost of the combined induced flow $g + h$, given by $C(g + h) = \sum_{a\in A}(g_a + h_a) l_a(g_a + h_a)$. In particular, we are interested in bounding the ratio $C(g+h)/C(o)$, called the *price of anarchy*.

In the remainder of the paper, we assume that the reader is familiar with the asymptotic notations $O(\cdot)$, $\Omega(\cdot)$, and $\Theta(\cdot)$; their definition can be found in any book on the analysis of algorithms, for example the one by Knuth [20]. We will also use the shorthand $[k] := \{1, 2, \ldots, k\}$.

3. Limits of Stackelberg routing. In this section, we prove that there does not exist a Stackelberg strategy that induces a price of anarchy bounded by a function of α only. More precisely, we show that for any fixed $\alpha \in (0, 1)$, the ratio between the cost of the flow induced by any Stackelberg strategy and the optimum can be arbitrarily large, even in single-commodity networks.

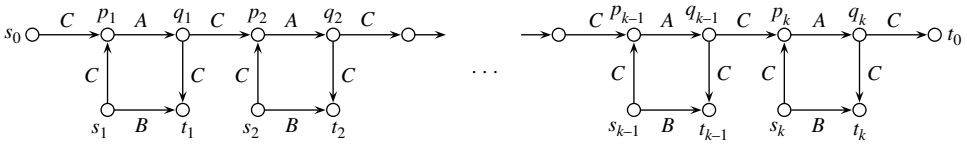

FIGURE 1. The graph G_k used in the proof of Theorem 3.1. Arcs are labeled with their types.

3.1. Multicommodity networks. We first show this claim for multicommodity networks. In this case, such a result was already known to hold for weak Stackelberg strategies (Roughgarden [30]); here, we prove that it also holds for strong Stackelberg strategies.

THEOREM 3.1. *Let $M > 0$ and $\alpha \in (0, 1)$. There is a multicommodity instance $\mathcal{I} = (G, r, l, \alpha)$ such that if g is any strong Stackelberg strategy for $\mathcal{I}$ inducing a Nash flow h and o is an optimal flow for the instance (G, r, l), then $C(g + h) \geq M \cdot C(o)$.*

To prove Theorem 3.1, we will use an instance based on the graph depicted in Figure 1. For a positive integer k, the graph G_k has $4k + 2$ vertices $V_k = \{s_0, t_0, s_1, t_1, p_1, q_1, \ldots, s_k, t_k, p_k, q_k\}$. The arc set A_k is the union of three sets, $\{(p_i, q_i)\colon i \in [k]\}$, $\{(s_i, t_i)\colon i \in [k]\}$, and $\{(s_i, p_i), (q_i, t_i), (q_i, p_{i+1})\colon i \in [k]\} \cup \{(s_0, p_1), (q_k, t_0)\}$. We call the arcs in these sets of type A, B, and C, respectively (see Figure 1). There are $k + 1$ commodities $0, 1, \ldots, k$. Commodity i has origin s_i and destination t_i. The demand is $r_0 := (1 - \alpha)/2$ for commodity 0 and $r_1 := (1 + \alpha)/2k$ for all other commodities. Thus, the total demand is $r_0 + kr_1 = 1$.

The latency of an arc is determined by its type. Type B arcs have constant latency 1 and type C arcs have constant latency 0. Type A arcs have latency $l_\epsilon(x)$, where the function $l_\epsilon(x)$ is defined as follows:

$$l_\epsilon(x) = \begin{cases} 0, & \text{if } x \leq r_0; \\ 1 - \dfrac{r_0 + r_1 - x}{(1 - \epsilon)r_1}, & \text{if } x \geq r_0 + 2\epsilon r_1. \end{cases}$$

Here, ϵ is any positive constant such that $\epsilon < (1 - \alpha)/(1 + \alpha)$. In the interval $(r_0, r_0 + 2\epsilon r_1)$, the function l_ϵ is defined arbitrarily so that, overall, it is a standard and convex function (see also Figure 2). In particular, $l_\epsilon(x) \geq 1 - (r_0 + r_1 - x)/((1 - \epsilon)r_1)$ for all x.

Let us first bound the cost of the optimal flow.

LEMMA 3.1. $C(o) \leq 1$.

PROOF. Consider the flow $\bar{f}$ where each commodity is routed along the shortest path (in terms of number of arcs) from origin to destination. The latency on the $s_0 - t_0$ path is zero because the load on each arc of the path is r_0 and $l_\epsilon(r_0) = 0$. The latency of each other $s_i - t_i$ path is 1. Then, $C(o) \leq C(\bar{f}) = k \cdot r_1 = (1 + \alpha)/2 \leq 1$. □

PROOF OF THEOREM 3.1. For $i = 1, 2, \ldots, k$, let g_i be the amount of flow sent by the Stackelberg strategy over the arc (s_i, t_i). Because the total value of the flow controlled by any Stackelberg strategy is α, we have $\sum_{i=1}^{k} g_i \leq \alpha$.

The crucial point is that without loss of generality, all the selfish flow induced by g on an $s_i - t_i$ path, $i \neq 0$, will be sent along the path (s_i, p_i, q_i, t_i). Indeed, if the arc (s_i, t_i) contained some selfish flow $h_i > 0$, the latency of the path (s_i, p_i, q_i, t_i) would be $l_\epsilon(r_0 + r_1 - g_i - h_i) < 1 = l_{(s_i, t_i)}(g_i + h_i)$. However, this contradicts the definition of Nash flows. Thus, the combined flow on each (p_i, q_i) arc is exactly $r_0 + r_1 - g_i$. Now, let P_0 be the unique $s_0 - t_0$ path. We have

$$l_{P_0}(g + h) \geq \sum_{i=1}^{k} l_\epsilon(r_0 + r_1 - g_i) \geq \sum_{i=1}^{k} \left(1 - \frac{g_i}{(1 - \epsilon)r_1}\right) \geq k - \frac{\alpha}{(1 - \epsilon)r_1} = \frac{1}{1 - \epsilon} \cdot \left(\frac{1 - \alpha}{1 + \alpha} - \epsilon\right) \cdot k.$$

The last inequality follows from $\sum_i g_i \leq \alpha$ and the last equality from $r_1 = (1 + \alpha)/2k$. Because $\epsilon < (1 - \alpha)/(1 + \alpha)$, we conclude that $l_{P_0}(g + h) = \Omega(k)$. Together with Lemma 3.1, we obtain

$$C(g + h) \geq r_0 \cdot l_{P_0}(g + h) = \tfrac{1}{2} \cdot (1 - \alpha) \cdot \Omega(k) = \Omega(k) \cdot C(o).$$

Thus, the ratio of $C(g + h)/C(o)$ can be made arbitrarily large by picking a sufficiently large k. □

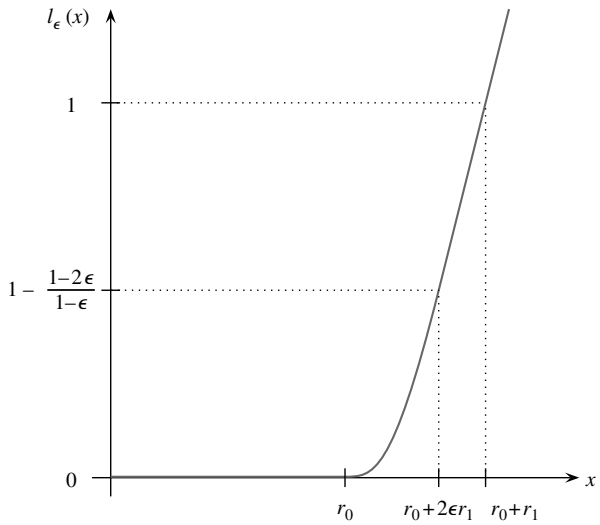

FIGURE 2. The latency function $l_\epsilon(x)$ used in the proof of Theorem 3.1.

REMARK 3.1. We remark that the proof of Theorem 3.1 also works for undirected networks. In these networks, flow can be sent across an edge in both directions and the aggregated flow of an edge is defined as the sum of the flows traversing that edge (in either direction). To see that the lower-bound proof still holds, observe that the selfish flow of commodity $i \in [k]$ is still routed along the (s_i, p_i, q_i, t_i) path. The selfish flow sent from s_0 to t_0 now has potentially more paths available than in the directed case. However, it is easy to see that this flow is sent along the $(s_0, p_1, q_1, \ldots, p_k, q_k, t_0)$ path and, thus, the proof goes through without change. We do not know, however, whether the lower bound for single-commodity networks presented can be extended to undirected networks.

3.2. Single-commodity networks. We use the insights gained in the previous section to prove the following stronger result.

THEOREM 3.2. *Let $M > 0$ and $\alpha \in (0, 1)$. There is a single-commodity instance $\mathcal{I} = (G, r, l, \alpha)$ such that, if g is any strong Stackelberg strategy for $\mathcal{I}$ inducing a Nash flow h and o is an optimal flow for the instance (G, r, l), then $C(g+h) \geq M \cdot C(o)$.*

Theorem 3.2 extends Theorem 3.1 to single-commodity networks. The main idea behind the proof of Theorem 3.2 is to simulate the instance used in Theorem 3.1 by creating a supersource s and a supersink t and connecting them to the sources and sinks of the original network (see also Figure 3). If somehow we were able to enforce the $s-t$ flow to split according to the demand vector of the multicommodity instance, the result would easily follow as in the proof of Theorem 3.1. In order to achieve this, we use latency functions that simulate capacities on the arcs connecting the supersource to the sources and the sinks to the supersink. Although these "capacities" might be exceeded, we will make sure that if the excess flow is too large, the price of anarchy will already be large enough for our purposes.

To prove Theorem 3.2, we use the instance $G_k = (V_k, A_k)$ depicted in Figure 3. For a positive integer k, the graph G_k has $4k+4$ vertices. There is a single commodity (s, t) with unit demand. Define $r_0 := (1-\alpha)/2$ and $r_1 := (1+\alpha)/2k$. Note that the total demand is equal to $r_0 + kr_1$. Every arc is of one of five different *types*

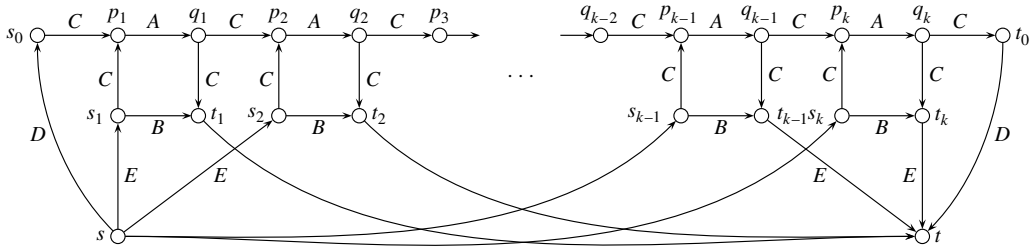

FIGURE 3. The graph G_k used in the proof of Theorem 3.2. Arcs are labeled with their types.

$\{A, B, C, D, E\}$ as indicated in Figure 3. The latency of an arc is determined by its type. Type B arcs have constant latency 1 and type C arcs have constant latency 0. Arcs of type A have the following latency function:

$$l_0(x) = \begin{cases} 0, & \text{if } x \le r_0; \\ 1 - \dfrac{r_0 + r_1 - x}{r_1}, & \text{if } x > r_0. \end{cases}$$

Although $l_0(x)$ is not differentiable at r_0, it can be approximated with arbitrarily small error by standard functions.

For fixed L and τ, let $u_{L,\tau}(x)$ be any standard function satisfying $u_{L,\tau}(L) = 0$ and $u_{L,\tau}(L+\tau) = M/\tau$. Type D arcs have latency $u_{r_0,\,\delta/3k^3}(x)$, and type E arcs have latency $u_{r_1,\,\delta/3k^3}(x)$. We will fix the constant δ later in the proof.

LEMMA 3.2. $C(o) \le 1$.

PROOF. Let P_0 be the path $(s, s_0, p_1, q_1, p_2, \ldots, p_k, q_k, t_0, t)$. For $i \in [k]$, let P_i be the path (s, s_i, t_i, t). Consider the feasible flow f such that $f_{P_0} = r_0$ and $f_{P_i} = r_1$ for $i \in [k]$. The latency induced by f is zero on arcs of type A, C, D, and E, and one on arcs of type B. Thus, $C(o) \le C(f) = k \cdot r_1 = (1+\alpha)/2 \le 1$. □

Lemma 3.3 will allow us to focus on the case where the combined flow on type D and E arcs does not exceed a certain threshold value.

LEMMA 3.3. *For any Stackelberg strategy g inducing a Nash flow h, the following hold*:
(i) *If a is a type D arc and $g_a + h_a \ge r_0 + \delta/3k^3$, then $C(g+h) \ge M \cdot C(o)$.*
(ii) *If a is a type E arc and $g_a + h_a \ge r_1 + \delta/3k^3$, then $C(g+h) \ge M \cdot C(o)$.*

PROOF. We prove statement (i); the proof for (ii) is similar. We have $C(g+h) \ge (g_a + h_a) \cdot l_a(g_a + h_a) = (g_a + h_a) \cdot u_{r_0,\,\delta/3k^3}(g_a + h_a) \ge (r_0 + \delta/3k^3) \cdot M/(\delta/3k^3) \ge M$. The proof follows from Lemma 3.2. □

For the remainder of the proof, we assume that there is no arc satisfying the conditions of Lemma 3.3; otherwise, Theorem 3.2 follows immediately.

LEMMA 3.4. *For any Stackelberg strategy g inducing a Nash flow h, the following hold*:
(i) *For any arc $a = (q_{i-1}, p_i)$, $i \in [k]$, $g_a + h_a \ge r_0 - \delta/k$.*
(ii) *For any arc $a = (s, s_i)$, $i \in [k]$, $g_a + h_a \ge r_1 - \delta/k$.*

PROOF. Regarding (i), we will prove by induction on i the stronger claim

$$g_a + h_a \ge r_0 - (2i+1)\delta/3k^2.$$

For $i = 1$, notice that, by Lemma 3.3, the flow along each of $(s, s_1), \ldots, (s, s_k)$ is at most $r_1 + \delta/3k^3$, so the flow on (s, s_0) must be at least $1 - \sum_{i=1}^{k}(r_1 + \delta/3k^3) = 1 - kr_1 - \delta/3k^2 = r_0 - \delta/3k^2$. However, the flow on (s, s_0) is the same as that on arc $(s_0, p_1) = (q_0, p_1)$. Notice that a similar argument allows us to also conclude that the flow on each (s, s_i) arc ($i \in [k]$) is at least $r_1 - \delta/3k^2$. This implies (ii) for all $i \in [k]$.

To prove (i) for $i > 1$, consider the ith block in the graph (Figure 4) and let $f = g + h$. By flow conservation, $f_{(q_i,\,p_{i+1})} = f_{(q_{i-1},\,p_i)} + f_{(s,\,s_i)} - f_{(t_i,\,t)}$. Using induction and Lemma 3.3,

$$\begin{aligned} f_{(q_i,\,p_{i+1})} &= f_{(q_{i-1},\,p_i)} + f_{(s,\,s_i)} - f_{(t_i,\,t)} \\ &\ge (r_0 - (2i-1)\delta/3k^2) + (r_1 - \delta/3k^2) - (r_1 + \delta/3k^3) = r_0 - (2i+1)\delta/3k^2. \quad \square \end{aligned}$$

We are now ready to conclude the proof of Theorem 3.2.

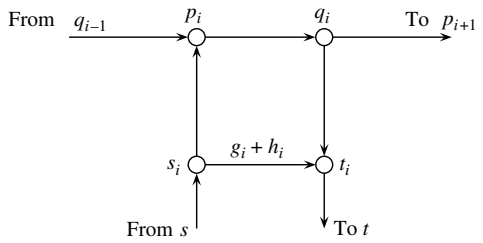

FIGURE 4. The ith block of the graph G_k.

PROOF OF THEOREM 3.2. For any $i \in [k]$, consider the ith block in the graph (Figure 4). Let g_i, h_i be the Stackelberg and selfish flow on the arc (s_i, t_i), respectively. We have two cases:

(i) $h_i = 0$. In this case, using Lemma 3.4, the flow on arc (p_i, q_i) is at least $r_0 - \delta/k + r_1 - \delta/k - g_i$. The latency on that same arc is thus at least $l_0(r_0 + r_1 - 2\delta/k - g_i)$.

(ii) $h_i > 0$. In this case, the Nash flow on path $P_i' = (s, s_i, t_i, t)$ is strictly positive. Consider the path $P_i'' = (s, s_i, p_i, q_i, t_i, t)$. By definition of Nash flow, $l_{P_i''}(g+h) \geq l_{P_i'}(g+h)$. Notice that the two paths P_i', P_i'' share all their nonzero-latency arcs except for (s_i, t_i) (only present in P_i') and (p_i, q_i) (only present in P_i''). Thus, $l_{P_i''}(g+h) \geq l_{P_i'}(g+h)$ implies $l_{(p_i,q_i)}(g+h) \geq l_{(s_i,t_i)}(g+h) = 1$. As a consequence, $l_{(p_i,q_i)}(g+h) \geq 1 = l_0(r_0+r_1) \geq l_0(r_0+r_1-2\delta/k-g_i)$ because g_i and δ/k are nonnegative.

In both cases, $l_{(p_i,q_i)}(g+h) \geq l_0(r_0+r_1-2\delta/k-g_i) \geq 1-(g_i+2\delta/k)/r_1$.

The latency on the path $P_0 = (s, s_0, p_1, q_1, \ldots, p_k, q_k, t_0, t)$ is at least

$$l_{P_0}(g+h) \geq \sum_{i=1}^{k} l_{(p_i,q_i)}(g+h) \geq \sum_{i=1}^{k}\left(1 - \frac{g_i + 2\delta/k}{r_1}\right) \geq k - \frac{\alpha}{r_1} - \frac{2\delta}{r_1} = \left(\frac{1-\alpha-4\delta}{1+\alpha}\right)k.$$

The last inequality is a consequence of the fact that the total Stackelberg flow is α; thus, $\sum_i g_i \leq \alpha$.

Choosing $\delta < (1-\alpha)/4$, we can conclude that $l_{P_0}(g+h) = \Omega(k)$. Together with Lemmas 3.2 and 3.4, this gives

$$C(g+h) \geq (r_0 - \delta/k) \cdot l_{P_0}(g+h) \geq (\tfrac{1}{2} \cdot (1-\alpha) - \delta) \cdot \Omega(k) = \Omega(k) \cdot C(o).$$

Thus, the ratio $C(g+h)/C(o)$ can be made arbitrarily large by picking a sufficiently large k. □

REMARK 3.2. Suppose the Stackelberg leader is solely interested in minimizing the cost of the flow that he controls, i.e., $C_1(g+h) = \sum_{a \in A} g_a l_a(g_a + h_a)$. Our result also implies that even the ratio $C_1(g+h)/C(o)$ can be unbounded, independent of the Stackelberg strategy g.

4. Upper bounds for LLF. The results of the previous section reveal that the price of anarchy of every Stackelberg strategy is unbounded, even in single-commodity networks. Note that in our proofs, we crucially exploit that the size of the network can be made arbitrarily large. More precisely, we constructed a family of graphs G_k with $n = \Theta(k)$ vertices and $m = \Theta(k)$ arcs and showed that the price of anarchy grows as a function of k. A natural question that arises is whether it is necessary to expand the network in order to prove an unbounded price of anarchy. Or, said differently, is it possible to raise the price of anarchy beyond any fixed $M > 0$, even for constant-size networks (for instance, the Braess graph)?

We answer this question negatively by proving (for any fixed α) an upper bound on the price of anarchy of $O(b(n, m, k))$, where $b(n, m, k)$ is some function depending on the number of vertices, arcs, and commodities of the network. The upper bound holds for the particular Stackelberg strategy known as largest-latency-first (LLF); see Roughgarden [28] and Swamy [36]. Besides complementing the negative results of the previous section, these are also the first upper bounds for LLF in general networks that hold for *arbitrary* latency functions.

LLF works as follows for a given instance $\mathcal{I} = (G, r, l, \alpha)$: First, compute an optimal flow o for (G, r, l) and then successively saturate the paths used by o in nonincreasing order of their latencies until we have routed an α fraction of the overall demand. More precisely, we initially set $g_a := 0$ for all arcs $a \in A$ and define the *residual demand* as $\Delta := \alpha\Delta_0 := \alpha \sum_{i=1}^{k} r_i$. While Δ remains positive, we repeatedly find a path P such that $l_P(o) = \max_{P:(o-g)_P>0} l_P(o)$ and set $g_a := g_a + \min\{\Delta, (o-g)_P\}$ for all arcs $a \in P$ and $\Delta := \max\{0, \Delta - (o-g)_P\}$. Because o is an acyclic flow, the flow g can be computed in polynomial time. Clearly, g is opt restricted because $g_a \leq o_a$ for every arc $a \in A$ by construction. Observe that LLF is a strong Stackelberg strategy.

Consider the instance $\tilde{\mathcal{I}} = (G, \tilde{r}, \tilde{l})$ (as defined in §2). Recall that $\tilde{l}_a(x) := l_a(g_a + x)$ for all $a \in A$. The *maximum latency* of a flow f is defined as $L(f) := \max_{P \in \mathcal{P}: f_P > 0} \tilde{l}_P(f)$. Let h be a Nash flow and let $o^{\max}$ denote a flow that minimizes the maximum latency. Then, $\rho_L := L(h)/L(o^{\max})$ denotes the worst-case ratio between the maximum latency of a Nash flow and the maximum latency of an arbitrary flow. To prove the upper bound, we bound the price of anarchy induced by LLF in terms of ρ_L. The upper bound will then follow from the previously known fact that ρ_L can be bounded in terms of the network size only (Roughgarden [29], Lin et al. [24]).

THEOREM 4.1. *Let $\mathcal{I} = (G, r, l, \alpha)$ be a multicommodity instance with m arcs and let g be the LLF strategy. Then, $C(g+h) \leq (m + 1/\alpha)\rho_L C(o)$.*

PROOF. Consider the quantity $L^g := \min_{P \in \mathcal{P}: g_P > 0} l_P(o)$. We claim that

$$L(h) \leq \rho_L L(o^{\max}) \leq \rho_L L(o-g) \leq \rho_L L^g.$$

The first inequality follows from the definition of ρ_L, the second inequality follows because $o-g$ is feasible for $\tilde{\mathcal{I}}$, and the third inequality follows because $L(o-g) \le L^g$ by the definition of LLF.

We further observe that

$$\alpha\Delta_0 L^g \le \sum_{P\in\mathcal{P}} g_P l_P(o) = \sum_{a\in A} g_a l_a(o_a) \le C(o).$$

The first inequality follows from the definition of L^g while the second is trivially valid because g is opt restricted. We are now ready to bound the cost C_2 of the followers:

$$\begin{aligned} C_2(g+h) &:= \sum_{a\in A} h_a l_a(g_a+h_a) = \sum_{P\in\mathcal{P}} h_P \tilde{l}_P(h) \\ &\le L(h) \sum_{P\in\mathcal{P}} h_P = (1-\alpha)\Delta_0 L(h) \\ &\le (1-\alpha)\Delta_0\rho_L L^g \le \frac{1-\alpha}{\alpha}\rho_L C(o). \end{aligned}$$

For bounding the cost C_1 of the Stackelberg leader, we partition the set of arcs into $A_1 := \{a\in A\colon h_a > 0\}$ and $A_2 := \{a\in A\colon h_a = 0\}$. Then

$$\begin{aligned} C_1(g+h) &:= \sum_{a\in A} g_a l_a(g_a+h_a) = \sum_{a\in A_1} g_a l_a(g_a+h_a) + \sum_{a\in A_2} g_a l_a(g_a+h_a) \\ &\le \sum_{a\in A_1} g_a l_a(g_a+h_a) + C(o) \\ &\le |A_1|\alpha\Delta_0 L(h) + C(o) \le m\rho_L C(o) + C(o) = (m\rho_L+1)C(o). \end{aligned}$$

Combining the bounds for C_1 and C_2 yields

$$C(g+h) \le \left(\left(m+\frac{1}{\alpha}-1\right)\rho_L + 1\right)C(o).$$

As $\rho_L \ge 1$, Theorem 4.1 is proved. □

COROLLARY 4.1. *Let $\mathcal{I} = (G, r, l, \alpha)$ be a single-commodity instance with n vertices and m arcs, and let g be the LLF strategy. Then, $C(g+h) \le (n-1)(m+1/\alpha)C(o)$.*

PROOF. Roughgarden [29] proves that $\rho_L \le n-1$ for single-commodity instances with n vertices. □

COROLLARY 4.2. *Let $\mathcal{I} = (G, r, l, \alpha)$ be a multicommodity instance with n vertices, m arcs, and k commodities, and let g be the LLF strategy. Then, $C(g+h) \le b(n,m,k)(m+1/\alpha)C(o)$, where $b(n,m,k) = 2^{O(\min\{kn,\, m\log n\})}$.*

PROOF. Lin et al. [24] prove that $\rho_L = 2^{O(\min\{kn,\, m\log n\})}$ for any multicommodity instance with n vertices, m arcs, and k commodities. □

5. A bicriteria bound for general latency functions. As we have seen in the previous sections, no Stackelberg strategy controlling a constant fraction of the traffic can reduce the price of anarchy to a constant, even if we consider single-commodity networks. In light of this negative result, we therefore compare the cost of a Stackelberg strategy on an instance $\mathcal{I} = (G, r, l, \alpha)$ to the cost of an optimal flow for the instance $\mathcal{I}^\beta = (G, \beta r, l)$ in which the demand vector has been scaled up by a factor $\beta > 1$.

We propose the following simple Stackelberg strategy, which we term *augmented SCALE* (*ASCALE*):

(i) Compute an optimal flow o^β for the instance $\mathcal{I}^\beta$.

(ii) Define the Stackelberg flow by $g := (\alpha/\beta)o^\beta$.

We prove that the resulting flow induced by the Stackelberg strategy ASCALE satisfies $C(g+h) \le C(o^\beta)$ if we choose $\beta = 1+\sqrt{1-\alpha}$. This result can be seen as a generalization of the result by Roughgarden and Tardos [32] that the cost of a Nash flow is always less than or equal to the cost of the optimal flow for an instance in which demands have been doubled. Our bound gives a smooth transition from absence of centralized control (where doubling the demands is sufficient) to completely centralized control (where no augmentation is necessary).

LEMMA 5.1. *If g is the ASCALE strategy, $C(g+h) \le \sum_{a\in A}(1/\beta)o^\beta_a l_a(g_a+h_a)$.*

PROOF. Consider the flow $(1-\alpha)g/\alpha$; it is a flow feasible with respect to $(1-\alpha)r$. Using the variational inequality (1), we get

$$\sum_{a\in A} h_a l_a(g_a+h_a) \le \frac{1-\alpha}{\alpha}\sum_{a\in A} g_a l_a(g_a+h_a).$$

Adding $\sum_a g_a l_a(g_a+h_a)$ to both sides and using $g=(\alpha/\beta)o^\beta$ proves Lemma 5.1. □

THEOREM 5.1. *If g is the ASCALE strategy, $C(g+h) \le 1/(\beta-1)\cdot(1-\alpha/\beta)\cdot C(o^\beta)$. Furthermore, this bound is tight.*

PROOF. We first show that for *every* arc $a\in A$,

$$o_a^\beta l_a(g_a+h_a) \le (g_a+h_a)l_a(g_a+h_a) + \left(1-\frac{\alpha}{\beta}\right)o_a^\beta l_a(o_a^\beta). \tag{2}$$

There are two cases. When $g_a+h_a \ge o_a^\beta$, the inequality holds simply because its left-hand side is upper bounded by the first summand of the right-hand side. Otherwise, if $o_a^\beta > g_a+h_a$, we obtain

$$\begin{aligned} o_a^\beta l_a(g_a+h_a) &\le (g_a+h_a+o_a^\beta-g_a)l_a(g_a+h_a) = (g_a+h_a)l_a(g_a+h_a) + \left(1-\frac{\alpha}{\beta}\right)o_a^\beta l_a(g_a+h_a) \\ &\le (g_a+h_a)l_a(g_a+h_a) + \left(1-\frac{\alpha}{\beta}\right)o_a^\beta l_a(o_a^\beta). \end{aligned}$$

Summing (2) over all $a\in A$, we obtain

$$\sum_{a\in A} o_a^\beta l_a(g_a+h_a) \le C(g+h) + \left(1-\frac{\alpha}{\beta}\right)C(o^\beta).$$

Invoking Lemma 5.1, we get

$$\beta\cdot C(g+h) \le \sum_{a\in A} o_a^\beta l_a(g_a+h_a) \le C(g+h) + \left(1-\frac{\alpha}{\beta}\right)C(o^\beta).$$

Solving for $C(g+h)$ now gives the bound as claimed. The bound is also tight, as can be seen by considering a slightly modified Pigou instance. □

COROLLARY 5.1. *Let $\beta = 1+\sqrt{1-\alpha}$. If g is the ASCALE strategy, then $C(g+h)\le C(o^\beta)$.*

For a given instance $\mathcal{I}=(G,r,l,\alpha)$, the SCALE strategy is defined as $g=\alpha o$, where o is an optimal flow for (G,r,l). Theorem 5.2 shows that our result for ASCALE has a consequence for the SCALE strategy as well.

THEOREM 5.2. *Let $g=\alpha o$ be the SCALE strategy for instance $\mathcal{I}=(G,r,l,\alpha)$. Define a modified instance $\hat{\mathcal{I}}=(G,r,\hat{l},\alpha)$ with latency functions $\hat{l}_a(x)=l_a(x/\beta)/\beta$ for every arc a, where $\beta=1+\sqrt{1-\alpha}$, and let $\hat{C}(\cdot)$ denote the cost of a flow with respect to $\hat{l}$. Let $\hat{h}$ be the Nash flow induced by $\hat{g}=g$ in $\hat{\mathcal{I}}$. Then, $\hat{C}(\hat{g}+\hat{h})\le C(o)$.*

PROOF. Observe that the SCALE strategy for $\mathcal{I}$ can be obtained by computing the ASCALE strategy for $\mathcal{I}^{1/\beta} := (G, r/\beta, l, \alpha)$ and scaling it up by a factor of β; that is, $\hat{g}=\beta g$, where g is the ASCALE strategy for $\mathcal{I}^{1/\beta}$. Let h be the Nash flow induced by g in $\mathcal{I}^{1/\beta}$. By the variational inequality (1),

$$\sum_{a\in A} h_a l_a(g_a+h_a) \le \sum_{a\in A} y_a l_a(g_a+h_a) \tag{3}$$

for any flow y feasible for $(1-\alpha)r/\beta$. Because $l_a(x/\beta)/\beta = \hat{l}_a(x)$, we can rewrite (3) as $\sum_a(\beta h_a)\hat{l}_a(\hat{g}_a+\beta h_a) \le \sum_a(\beta y_a)\hat{l}_a(\hat{g}_a+\beta h_a)$. This implies that βh is a Nash flow induced by $\hat{g}$ in $\hat{\mathcal{I}}$. Because the cost of Nash flows is unique, $\hat{C}(\hat{g}+\beta h)=\hat{C}(\hat{g}+\hat{h})$. Finally, because $\hat{C}(\beta x)=C(x)$ for any flow x, we can conclude $\hat{C}(\hat{g}+\hat{h}) = \hat{C}(\beta(g+h)) = C(g+h) \le C(o)$ where the inequality follows from Corollary 5.1. □

A class of latency functions that are of high practical relevance are so-called *M/M/1 latency functions* (see also Roughgarden and Tardos [32]). These functions are of the form $l_a(x)=1/(u_a-x)$, where u_a intuitively represents the capacity of arc a. Theorem 5.2 has a particularly nice interpretation in this case: The modified latency functions are $\hat{l}_a(x)=l_a(x/\beta)/\beta = 1/(\beta(u_a-x/\beta)) = 1/(\beta u_a - x)$. In a purely selfish scenario, Theorem 5.2 therefore implies that to beat optimal routing, it is sufficient to double the capacity of every arc. This has been observed before by Roughgarden and Tardos [32]. In the Stackelberg scenario, Theorem 5.2 shows that it is sufficient to increase the capacities by a factor of $1+\sqrt{1-\alpha}$ if the SCALE strategy is used.

6. Bounds for specific classes of latency functions. In this section, we first present a general approach, which we call λ-*approach*, to analyze the price of anarchy of opt-restricted Stackelberg strategies. We then use the λ-approach to derive bounds on the price of anarchy of the SCALE strategy for a general class of latency functions including polynomial latency functions with nonnegative coefficients.

6.1. λ-approach. We start by proving an upper bound on the cost of the combined flow induced by an opt-restricted Stackelberg strategy.

LEMMA 6.1. *For any opt-restricted strategy g, $C(g+h) \le \sum_{a\in A} o_a l_a(g_a+h_a)$.*

PROOF. The proof follows immediately by applying the variational inequality (1) with $x = o - g$. □

For any latency function l_a and nonnegative numbers g_a, λ, we define the following nonnegative value:

$$\omega(l_a; g_a, \lambda) := \sup_{o_a, h_a \ge 0} \frac{o_a}{g_a + h_a} \cdot \frac{l_a(g_a+h_a) - \lambda l_a(o_a)}{l_a(g_a+h_a)}. \tag{4}$$

(We assume by convention $0/0 = 0$.) To bound the price of anarchy, we use the variational inequality (Lemma 6.1) and bound the cost of the induced flow on every arc by some λ-fraction of the optimal cost plus some ω-fraction of the cost of the induced flow itself:

$$C(g+h) = \sum_{a\in A} (g_a+h_a) l_a(g_a+h_a) \le \sum_{a\in A} \lambda \cdot o_a l_a(o_a) + \omega(l_a; g_a, \lambda) \cdot (g_a+h_a) l_a(g_a+h_a). \tag{5}$$

Now, the idea is to determine a λ that provides the tightest bound possible. Choosing $\lambda = 1$, the above approach resembles the one that was previously used by Correa et al. [7] to bound the price of anarchy of network routing games; however, optimizing over the parameter λ provides an additional means to obtain better bounds. The idea of introducing the scaling parameter λ was first introduced in the context of bounding the price of anarchy in atomic congestion games (see Harks [14]).

For a given opt-restricted strategy g, we further define $\omega(g, \lambda) = \max_{a\in A} \omega(l_a; g_a, \lambda)$. Before we state the main theorem, we need one additional definition.

DEFINITION 6.1. Given an opt-restricted strategy g, the *feasible region* is $\Lambda(g) := \{\lambda \in \mathbb{R}_+ \mid \omega(g, \lambda) < 1\}$.

Notice that every $\lambda \in \Lambda(g)$ induces a bound on the price of anarchy.

THEOREM 6.1. *Let $\lambda \in \Lambda(g)$. Then $C(g+h) \le (\lambda/(1-\omega(g,\lambda)))C(o)$.*

PROOF. The proof follows immediately from (5), Lemma 6.1, and the definition of $\omega(g, \lambda)$. □

6.2. Bounds for SCALE. In the following, we will analyze the SCALE strategy, defined by $g = \alpha o$.

DEFINITION 6.2. Let $\mathcal{L}_d$ be a class of continuous, nondecreasing, and standard latency functions satisfying

$$l(c\,z) \ge c^d l(z) \quad \forall\, c \in [0, 1]. \tag{6}$$

$\mathcal{L}_d$ contains, among others, polynomials with nonnegative coefficients and degree at most d. This characterization has been used before by Correa et al. [7].

6.2.1. SCALE: Latency functions in $\mathcal{L}_1$. We first consider latency functions that are in $\mathcal{L}_1$. In particular, this class contains continuous, nondecreasing, standard, and concave latencies.

LEMMA 6.2. *Assume $\lambda \in [0, 1]$ and latency functions in $\mathcal{L}_1$. Then,*

$$\omega(\alpha o, \lambda) \le \max\left\{\frac{1}{\alpha}(1-\lambda),\ \frac{1}{4\lambda}\right\}.$$

PROOF. By the definition of $\omega = \omega(l_a; \alpha o_a, \lambda)$,

$$\omega = \sup_{o_a,\, h_a \ge 0} \frac{o_a}{\alpha o_a + h_a} \cdot \frac{l_a(\alpha o_a + h_a) - \lambda l_a(o_a)}{l_a(\alpha o_a + h_a)}.$$

We consider two cases: (i) $\alpha o_a + h_a \ge o_a$ and (ii) $\alpha o_a + h_a \le o_a$. In (i), we define $\mu := o_a/(\alpha o_a + h_a) \in [0, 1]$. Then, we have

$$\omega = \sup_{o_a,\, h_a \ge 0,\, \mu\in[0,1]} \mu \cdot \frac{l_a(\alpha o_a + h_a) - \lambda l_a(\mu(\alpha o_a + h_a))}{l_a(\alpha o_a + h_a)} \le \max_{\mu\in[0,1]} \mu(1 - \lambda\mu) = \frac{1}{4\lambda},$$

where the last inequality follows from the definition of $\mathcal{L}_1$. The second case (ii) leads to

$$\omega \le \sup_{o_a, h_a \ge 0} \frac{o_a}{\alpha o_a + h_a} \cdot \frac{l_a(\alpha o_a + h_a) - \lambda l_a(\alpha o_a + h_a)}{l_a(\alpha o_a + h_a)} \le \sup_{o_a, h_a \ge 0} \frac{o_a}{\alpha o_a + h_a}(1-\lambda) \le \frac{1}{\alpha}(1-\lambda),$$

where the first inequality is valid because latencies are nondecreasing. □

We are now prepared to derive an upper bound on the price of anarchy.

THEOREM 6.2. *The price of anarchy of the SCALE strategy for latency functions in $\mathcal{L}_1$ is at most*

$$\frac{(1+\sqrt{1-\alpha})^2}{2(1+\sqrt{1-\alpha})-1}.$$

PROOF. Let $\lambda = \frac{1}{2}(1+\sqrt{1-\alpha})$. Then, by Lemma 6.2, $\omega(\alpha o, \lambda) \le 1/(2(1+\sqrt{1-\alpha})) < 1$, and thus $\lambda \in \Lambda(\alpha o)$. The proof now follows from Theorem 6.1. □

Note that the same bound has been proven by Karakostas and Kolliopoulos [18] for the special case of affine latencies. We next present a family of instances that pointwise match the upper bound of Theorem 6.2 for infinitely many values of α. More precisely, the lower bound is matched for all values of α such that $1/\sqrt{1-\alpha}$ is an integer. To the best of our knowledge, this is the first tight bound for values of $\alpha \ne 0, 1$.

THEOREM 6.3. *Let $n \ge 2$ be an integer and let $c = 1-(n-1)\alpha/n$. Then, the price of anarchy of the SCALE strategy for latency functions in $\mathcal{L}_1$ is at least*

$$\frac{nc^2 + (n-1)\alpha c}{(n-1)c + 1/n}.$$

Moreover, for all $\alpha = 1 - 1/k^2$, with k a positive integer, there exists an n such that the corresponding bound matches the upper bound of Theorem 6.2.

PROOF. We use the instance depicted in Figure 6(a). (Similar networks have been used in other constructions as well; see Babaioff et al. [1] and Roughgarden [31].) There is a single commodity (s, t) with unit demand. In the optimal flow, the demand is split evenly among the paths (s, p_i, q_i, t), $i \in [n]$. The resulting cost is $C(o) = (n-1)c + 1/n$.

The SCALE strategy sends a flow of value α/n along each direct path (s, p_i, q_i, t), $i \in [n]$. Due to the condition $c = 1-(n-1)\alpha/n$, the Nash flow is sent along the zigzag path $(s, p_1, q_1, p_2, \ldots, p_n, q_n, t)$. Thus, the cost of the combined flow $g+h$ is given by

$$C(g+h) = n\left(1 - \frac{n-1}{n}\alpha\right)^2 + (n-1)\alpha c = nc^2 + (n-1)\alpha c,$$

and the bound follows.

To see that the bound is tight when $\alpha = 1 - 1/k^2$, pick $n = k+1 = 1 + 1/\sqrt{1-\alpha}$. After substituting the expressions for n and c into the bound and appropriate rewriting, we obtain the same expression as in Theorem 6.2. □

We show that there exist instances such that no Stackelberg strategy can achieve a price of anarchy better than $(4-2\alpha+\alpha^2)/3$ for linear latency functions. That is, the upper bound on the price of anarchy of SCALE for latency functions in $\mathcal{L}_1$ (Theorem 6.2) is almost best possible (see Figure 5 for a comparison of the lower bound for arbitrary Stackelberg strategies and the upper bound of SCALE).

THEOREM 6.4. *There is an instance $\mathcal{I} = (G, r, l, \alpha)$ with linear latency functions such that if g is an arbitrary Stackelberg strategy for $\mathcal{I}$ inducing a Nash flow h and o is an optimal flow for the instance (G, r, l), then $C(g+h) \ge (4-2\alpha+\alpha^2)/3 \cdot C(o)$.*

Consider the Braess instance (Figure 6(b)) and suppose we send one unit of flow from s to t. Let g_1, g_2, and g_3 be the flow that the Stackelberg leader sends on the upper, zigzag, and lower path, respectively. Note that $g_3 = \alpha - g_1 - g_2$. Analogously, let h_1, h_2, and h_3 be the flow values on the respective paths of the selfish flow induced by g.

We first prove Lemma 6.3:

LEMMA 6.3. *Let g be an arbitrary Stackelberg strategy. The selfish flow h induced by g then satisfies $h_1 = h_3 = 0$.*

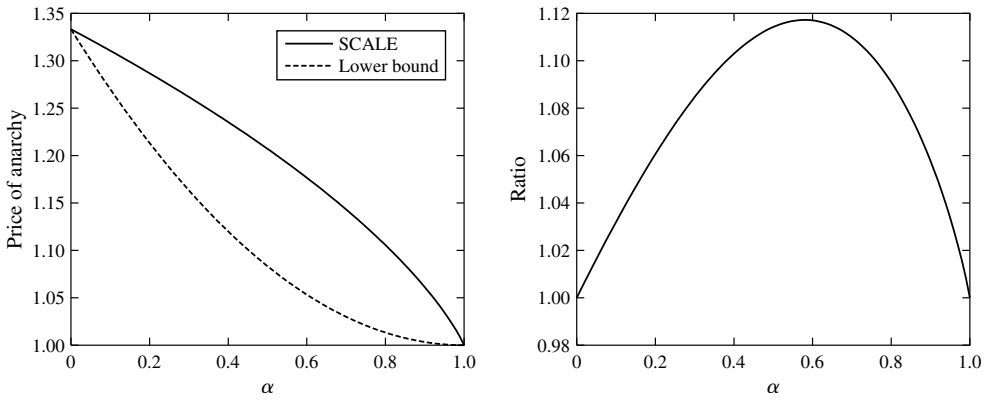

FIGURE 5. Lower bound for arbitrary Stackelberg strategies vs. upper bound of SCALE for linear latency functions (left) and the respective ratio (right).

PROOF. The latency of the zigzag path is $l_2 = g_1 + 2g_2 + g_3 + h_1 + 2h_2 + h_3 = 1 + g_2 + h_2$, where we exploit that $g_3 = \alpha - g_1 - g_2$ and $h_3 = (1-\alpha) - h_1 - h_2$. The latencies of the upper and lower paths are $l_1 = g_1 + g_2 + h_1 + h_2 + 1$ and $l_3 = 1 + g_2 + g_3 + h_2 + h_3$, respectively. Note that $l_1 \geq l_2$ and $l_3 \geq l_2$ independently of the choice of h_2. Because the selfish flow is routed on minimum latency paths, we must have $h_1 = h_3 = 0$ and $h_2 = (1-\alpha)$. □

PROOF OF THEOREM 6.4. The cost of an optimal flow o for the Braess instance is $C(o) = 3/2$. Consider the cost of the combined flow $g + h$. Using Lemma 6.3, we obtain

$$\begin{aligned} C(g+h) &= (g_1 + g_2 + (1-\alpha))^2 + (g_2 + g_3 + (1-\alpha))^2 + g_3 + g_1 \\ &= (g_1 + g_2 + (1-\alpha))^2 + (1-g_1)^2 + \alpha - g_2. \end{aligned}$$

This expression is minimized if $g_1 = \alpha/2$ and $g_2 = 0$, i.e., SCALE is the best strategy in this case. We obtain

$$\frac{C(g+h)}{C(o)} \geq \frac{2((\alpha/2 + (1-\alpha))^2 + (1-\alpha/2)^2 + \alpha)}{3} = \frac{4 - 2\alpha + \alpha^2}{3}. \quad \square$$

Because computing the best Stackelberg strategy is NP-hard (Roughgarden [28]), one may want to devise Stackelberg strategies that are efficiently computable and achieve a good approximation ratio. We say that a Stackelberg strategy g achieves an *approximation ratio of* $c \geq 1$ iff for every instance the cost of the (combined) flow induced by g is at most c times the cost of the (combined) flow induced by any other Stackelberg strategy. In this context, Corollary 6.1 follows immediately from Theorems 6.2 and 6.4.

COROLLARY 6.1. *The approximation ratio that the SCALE strategy achieves for latency functions in* $\mathcal{L}_1$ *is at most*

$$\frac{2 - \alpha + 2\sqrt{1-\alpha}}{1 + 2\sqrt{1-\alpha}} \cdot \frac{3}{4 - 2\alpha + \alpha^2} < 1.12.$$

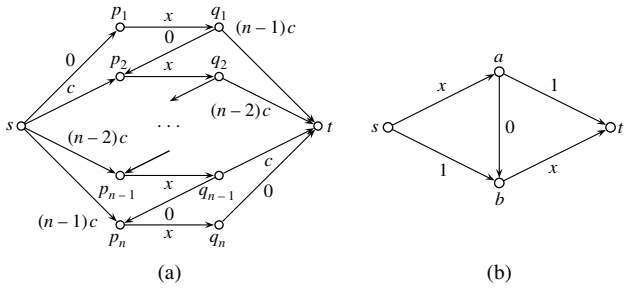

FIGURE 6. (a) Generalized Braess instance used in the proof of Theorem 6.3. (b) Braess instance. Arcs are labeled with their latency functions.

6.2.2. SCALE: Latency functions in $\mathscr{L}_d$. Next, we consider the class $\mathscr{L}_d$ of continuous, nondecreasing, and standard latency functions with $d \geq 1$. The proof of Lemma 6.4 proceeds along the same lines as the proof of Lemma 6.2.

LEMMA 6.4. *Assume $\lambda \in [0, 1]$ and latency functions in $\mathscr{L}_d$. Then,*

$$\omega(\alpha o, \lambda) \leq \max\left\{\frac{1}{\alpha}(1-\lambda),\ \frac{d}{d+1}\cdot\frac{1}{((d+1)\lambda)^{1/d}}\right\}.$$

PROOF. The proof proceeds along the same lines as the proof of Lemma 6.2. The only difference is the first part: (i) $\alpha o_a + h_a \geq o_a$. As before, we define $\mu := o_a/(\alpha o_a + h_a) \in [0, 1]$. We have

$$\omega = \sup_{o_a,\, h_a \geq 0,\, \mu \in [0,1]} \mu \cdot \frac{l_a(\alpha o_a + h_a) - \lambda l_a(\mu(\alpha o_a + h_a))}{l_a(\alpha o_a + h_a)} \leq \max_{\mu \in [0,1]} \mu(1 - \lambda\mu^d) = \frac{d}{d+1}\cdot\frac{1}{((d+1)\lambda)^{1/d}}. \quad \square$$

LEMMA 6.5. *There is a unique $\lambda \in (0, 1)$, call it λ_d, such that*

$$\frac{1}{\alpha}(1-\lambda) = \frac{d}{d+1}\cdot\frac{1}{((d+1)\lambda)^{1/d}}.$$

Then, $\lambda_d = z_d^d/(d+1)$, where $z_d \geq 1$ is the unique solution to the equation $z^{d+1} - (d+1)z + \alpha d = 0$.

PROOF. Substituting $\lambda = z_d^d/(d+1)$ in the starting equation and rewriting yields $z^{d+1} - (d+1)z + \alpha d = 0$. To verify that this equation has indeed exactly one solution larger than 1, use (for example) Descartes' rule of signs. $\square$

We are now ready to prove an upper bound for functions in $\mathscr{L}_d$.

THEOREM 6.5. *The price of anarchy of the SCALE strategy for latency functions in the class $\mathscr{L}_d$ is at most*

$$\frac{(d+1)z_d - \alpha d}{(d+1)z_d - d},$$

where $z_d \geq 1$ is the unique solution of the equation $z^{d+1} - (d+1)z + \alpha d = 0$.

PROOF. We will use Theorem 6.1 with $\lambda = \lambda_d$. However, in order to apply Theorem 6.1, we first need to upper bound $\omega(\alpha o, \lambda_d)$. Using Lemmas 6.4 and 6.5, we know that

$$\omega(\alpha o, \lambda_d) \leq \frac{d}{d+1}\cdot((d+1)\lambda_d)^{-1/d} = \frac{d}{d+1}\cdot z_d^{-1} < 1.$$

This implies $\lambda_d \in \Lambda(\alpha o)$ and we can invoke Theorem 6.1 to obtain a bound on the price of anarchy given by

$$\frac{\lambda_d}{1 - \omega(\alpha o, \lambda_d)} \leq \frac{z_d^d/(d+1)}{1 - (d/(d+1))z_d^{-1}} = \frac{z_d^{d+1}}{(d+1)z_d - d} = \frac{(d+1)z_d - \alpha d}{(d+1)z_d - d}. \quad \square$$

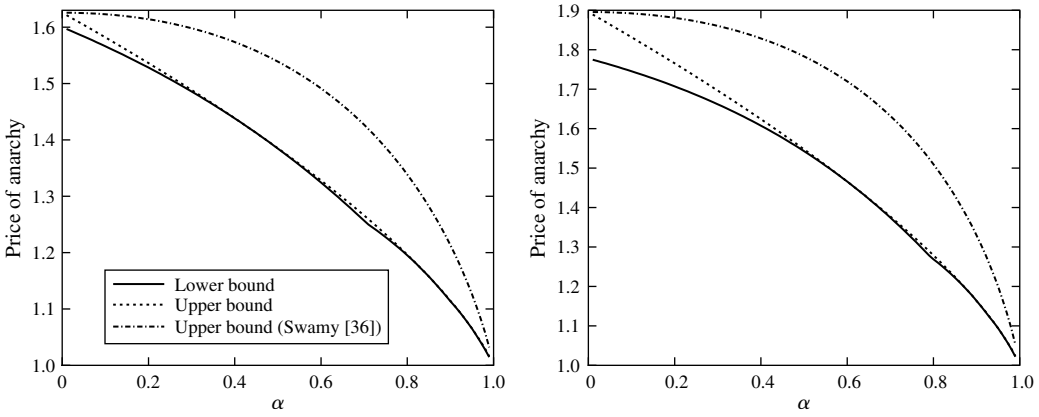

FIGURE 7. Upper vs. lower bounds for SCALE for polynomial latency functions of degree 2 (left) and 3 (right). The plots also show the previously best upper bound by Swamy [36].

A lower bound for polynomial latency functions of degree d can be obtained by generalizing the construction used in Theorem 6.3. We use again the network of Figure 6(a), except that we replace everywhere the latency function x by x^d and the constant c by $(1-(n-1)\alpha/n)^d$. The optimal flow is still split evenly on the direct paths so that, with similar arguments, we obtain the following lower bound.

THEOREM 6.6. *Let $n \geq 2$ be an integer and let $c = (1-(n-1)\alpha/n)^d$. Then, the price of anarchy of the SCALE strategy for latency functions in the class $\mathcal{L}_d$ is at least*

$$\frac{nc^{1+1/d} + (n-1)\alpha c}{(n-1)c + n^{-d}}.$$

Notice that Theorem 6.6 does not fix n so it is possible to optimize n based on α as in Theorem 6.3. For polynomial latency functions of degree 2 and 3, we compare in Figure 7 the lower bound thus obtained with the upper bound of Theorem 6.5 and also indicate the improvement with respect to the previously best bounds obtained by Swamy [36].

Acknowledgments. The research of the first author was partially supported by the FET (Future and Emerging Technologies) Unit of the European Commission (EC)—priority IST (Information Society Technologies)—within the 6th Framework Programme of EC, under Contract FP6-021235-2 (Project ARRIVAL).

References

[1] Babaioff, M., R. Kleinberg, C. H. Papadimitriou. 2007. Congestion games with malicious players. *Proc. 8th ACM Conf. Electronic Commerce*, ACM, New York, 103–112.
[2] Beckmann, M., C. B. McGuire, C. B. Winsten. 1956. *Studies in the Economics and Transportation*. Yale University Press, New Haven, CT.
[3] Braess, D. 1968. Über ein Paradoxon der Verkehrsplanung. *Unternehmenforschung* **11** 258–268.
[4] Chen, P.-A., D. Kempe. 2008. Altruism, selfishness, and spite in traffic routing. *Proc. 9th ACM Conf. Electronic Commerce*, ACM, New York, 140–149.
[5] Cole, R., Y. Dodis, T. Roughgarden. 2003. Pricing network edges for heterogeneous selfish users. *Proc. 35th ACM Sympos. Theory Comput.*, ACM, New York, 521–530.
[6] Correa, J. R., N. E. Stier-Moses. 2007. Stackelberg routing in atomic network games. Technical Report DRO-2007-03, Columbia Business School, New York.
[7] Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2004. Selfish routing in capacitated networks. *Math. Oper. Res.* **29** 961–976.
[8] Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2005. On the inefficiency of equilibria in congestion games. *Proc. 11th Internat. Conf. Integer Programming Combin. Optim.*, Springer, Berlin, 167–181.
[9] Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2007. Fast, fair, and efficient flows in networks. *Oper. Res.* **55**(2) 215–225.
[10] Dubey, P. 1986. Inefficiency of Nash equilibria. *Math. Oper. Res.* **11** 1–8.
[11] Fleischer, L. K., K. Jain, M. Mahdian. 2004. Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games. *Proc. 45th IEEE Sympos. Foundations Comput. Sci.*, IEEE, Washington, DC, 277–285.
[12] Fotakis, D. 2007. Stackelberg strategies for atomic congestion games. *Proc. 15th Eur. Sympos. Algorithms, Lecture Notes in Computer Science*, Vol. 4698. Springer, Berlin, 299–310.
[13] Gallo, G., C. Sandi, C. Sodini. 1980. An algorithm for the min concave cost flow problem. *Eur. J. Oper. Res.* **4**(4) 248–255.
[14] Harks, T. 2008. Stackelberg strategies and collusion in network games with splittable flow. *Proc. 6th Workshop Approximation Online Algorithms*, Springer, Berlin, 133–146.
[15] Hearn, D. W., M. V. Ramana. 1998. Solving congestion toll pricing models. P. Marcotte, S. Nguyen, eds. *Equilibrium and Advanced Transportation Modelling*. Springer, Berlin, 109–124.
[16] Kaporis, A. C., P. G. Spirakis. 2006. The price of optimum in Stackelberg games on arbitrary single commodity networks and latency functions. *Proc. 18th ACM Sympos. Parallelism Algorithms Architectures*. ACM, New York, 19–28.
[17] Karakostas, G., S. G. Kolliopoulos. 2004. Edge pricing of multicommodity networks for heterogeneous selfish users. *Proc. 45th IEEE Sympos. Foundations Comput. Sci.*, IEEE, Washington, DC, 268–276.
[18] Karakostas, G., S. G. Kolliopoulos. 2009. Stackelberg strategies for selfish routing in general multicommodity networks. *Algorithmica* **53**(1) 132–153.
[19] Knight, F. H. 1924. Some fallacies in the interpretation of social cost. *Quart. J. Econom.* **38**(4) 582–606.
[20] Knuth, D. E. 1997. *The Art of Computer Programming*. Addison-Wesley, Reading, MA.
[21] Korilis, Y. A., A. A. Lazar, A. Orda. 1997. Achieving network optima using Stackelberg routing strategies. *IEEE/ACM Trans. Networking* **5**(1) 161–173.
[22] Koutsoupias, E., C. H. Papadimitriou. 1999. Worst-case equilibria. C. Meines, S. Tison, eds. *Proc. 16th Sympos. Theoret. Aspects Comput. Sci.* Springer, Berlin, 404–413.
[23] Kumar, V. S. A., M. V. Marathe. 2002. Improved results for Stackelberg scheduling strategies. *Proc. 33rd Internat. Colloquium Automata, Languages Programming*, Springer, Berlin, 776–787.
[24] Lin, H., T. Roughgarden, É. Tardos, A. Walkover. 2005. Braess's paradox, Fibonacci numbers, and exponential inapproximability. *Proc. 32nd Internat. Colloquium Automata, Languages Programming*, Springer, Berlin, 497–512.
[25] Perakis, G. 2007. The "price of anarchy" under nonlinear and asymmetric costs. *Math. Oper. Res.* **32**(3) 614–628.
[26] Roughgarden, T. 2002. Selfish routing. Ph.D. thesis, Cornell University, Ithaca, NY.

[27] Roughgarden, T. 2003. The price of anarchy is independent of the network topology. *J. Comput. System Sci.* **67**(2) 341–364.
[28] Roughgarden, T. 2004. Stackelberg scheduling strategies. *SIAM J. Comput.* **33**(2) 332–350.
[29] Roughgarden, T. 2004. The maximum latency of selfish routing. *Proc. 15th ACM-SIAM Sympos. Discrete Algorithms*, SIAM, Philadelphia, 980–981.
[30] Roughgarden, T. 2005. *Selfish Routing and the Price of Anarchy*. The MIT Press, Cambridge, MA.
[31] Roughgarden, T. 2006. On the severity of Braess's paradox: Designing networks for selfish users is hard. *J. Comput. System Sci.* **72**(5) 922–953.
[32] Roughgarden, T., É. Tardos. 2002. How bad is selfish routing? *J. ACM* **49**(2) 236–259.
[33] Sharma, Y., D. Williamson. 2007. Stackelberg thresholds in network routing games or the value of altruism. *Proc. 8th ACM Conf. Electronic Commerce*, ACM, New York, 93–102.
[34] Smith, M. J. 1979a. The existence, uniqueness and stability of traffic equilibria. *Transportation Res.* **13**(4) 295–304.
[35] Smith, M. J. 1979b. The marginal cost taxation of a transportation network. *Transportation Res. Part B: Methodological* **13**(3) 237–242.
[36] Swamy, C. 2007. The effectiveness of Stackelberg strategies and tolls for network congestion games. *Proc. 18th ACM-SIAM Sympos. Discrete Algorithms*, SIAM, Philadelphia, 1133–1142.
[37] Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proc. Inst. Civil Engineers* **1**(Part II) 325–378.
[38] Yang, H., H.-J. Huang. 2004. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. *Transportation Res. Part B: Methodological* **38**(1) 1–15.