Stackelberg Routing in Arbitrary Networks*

Vincenzo Bonifaci[1,2,**], Tobias Harks[3***], and Guido Schäfer[3†]

[1] Università degli Studi dell'Aquila, Italy,
[2] Sapienza Università di Roma, Italy,
bonifaci@dis.uniroma1.it
[3] Technische Universität Berlin, Germany,
{harks,schaefer}@math.tu-berlin.de

Abstract. 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. We prove that the price of anarchy achievable via Stackelberg routing can be unbounded even for single-commodity networks. 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 which is simple, yet powerful enough to obtain (almost) tight bounds for SCALE in general networks.

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 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, which specifies the amount of flow that needs to be sent from the respective origin to the destination.

* This work was supported by the European Regional Development Fund (ERDF).

** Research supported by European Commission project ARRIVAL FP6-021235-2. This work was done while the author was a postdoctoral fellow at TU Berlin.

*** Research supported by the Federal Ministry of Education and Research (BMBF grant 03MOPAI1).

† Research supported by the DFG Research Center MATHEON "Mathematics for key technologies".

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 (non-decreasing) 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*). 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 [2, 7]. In the context of network routing games, a Nash flow in general does not minimize the total cost; or said differently, selfish behavior may cause a performance degradation in the network. Koutsoupias and Papadimitriou [13] 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 [21] 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 [16] and Correa, Schulz, and Stier-Moses [3]. (For an overview of these results, we refer to the book by Roughgarden [19].) 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 [21].

Due to 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* [12, 18]. 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, also 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.

Although Roughgarden [18] 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. As an example, Roughgarden [18] showed that for parallel-arc networks Stackelberg strategies exist that reduce 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 [23] obtained a similar result for single-commodity, series-parallel networks and Fotakis [8] for parallel-arc networks and unsplittable flows. Despite these positive results, a central question regarding the effectiveness of Stackelberg routing is still open: Does there always exist a Stackelberg strategy such that the price of anarchy is bounded? This question has been posed explicitly by Roughgarden [17, 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 [18]), which simply computes an optimal flow for the entire demand and then scales this flow by α. The currently best known bound for the price of anarchy induced by SCALE on multi-commodity networks and linear latency functions is due to Karakostas and Kolliopoulos [11]. More recently, Swamy [23] derived the first general bounds for polynomial latency functions.

Our Results. We investigate the impact of Stackelberg routing to reduce the price of anarchy in network routing games with nonatomic players. Our contribution is threefold:

1. We show that there are single-commodity networks for which every Stackelberg strategy induces a price of anarchy of at least $\Omega(n)$, where n is the number of nodes of the network. The result holds independently of the fraction $\alpha \in (0,1)$ of the centrally controlled demand. This settles the open question raised by Roughgarden [17].
2. In light of this 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 $1+\sqrt{1-\alpha}$.
3. 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).

Significance and Techniques. Our first result settles an important open question regarding the applicability of Stackelberg routing in general networks. While most existing results show that the performance degradation due to the absence of central control is *independent* of the underlying network topology, our result shows that the network topology matters in the context of Stackelberg routing. Our negative result also carries over to the unsplittable flow setting. However, due to lack of space, we omit the details from this extended abstract.

One important application of Stackelberg routing is the routing of Internet traffic within the domain of an Internet service provider, see also Sharma and Williamson [22]. 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 [21] (see Correa et al. [4] for other related results).

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 bounds by Swamy [23]. For linear latency functions, we derive an upper bound that coincides with a previous bound of Karakostas and Kolliopoulos in [11]. Their analysis is based on a (rather involved) machinery presented in [15]. However, our analysis is much simpler; in particular, we do not rely on the machinery in [15]. Moreover, we show that this bound also holds for concave latency functions. We present a generalized Braess instance that shows that for the linear case our bound is tight; a similar instance can be used to show that for higher degree polynomials our bounds are almost tight, leaving 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 also in other settings. For instance, the λ-approach can be applied to prove upper bounds when flows are unsplittable; details will be given in the full version of the paper. So far, such upper bounds for general networks are only known for linear latency functions (see Fotakis [8]).

Related Work. The idea of using Stackelberg strategies to improve the performance of a system was first proposed by Korilis, Lazar, and Orda [12]. The authors 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 [18] first formulated the problem and model considered here. He also proposed some natural Stackelberg strategies such as SCALE and Largest-Latency-First (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 Section 2 for a definition), the price of anarchy for multi-commodity networks can be unbounded [18]. However, this did not rule out the existence of effective Stackelberg strategies in general. Moreover, he also proved that it is NP-hard to compute the best Stackelberg strategy. Kumar and Marathe [14] investigated approximation schemes to compute the best Stackelberg strategy. The authors gave a PTAS for the case of parallel-arc networks.

Karakostas and Kolliopoulos [11] proved upper bounds on the price of anarchy for SCALE and LLF. Their bounds hold for arbitrary multi-commodity networks and linear latency functions. Their analysis is based on a result obtained by Perakis [15] to bound the price of anarchy for network routing games with asymmetric and non-separable latency functions. Furthermore, Karakostas and Kolliopoulos [11] showed that their analysis for SCALE is almost tight. More recently, Swamy [23] obtained upper bounds on the price of anarchy for SCALE and LLF for polynomial latency functions. Swamy also proved a bound of $1+1/\alpha$ for single-commodity, series-parallel networks with arbitrary latency functions. Fotakis [8] 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 proved that LLF still achieves an upper bound of $1/\alpha$ for arbitrary latency functions in this case.

Correa and Stier-Moses [5] 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 edge than the system optimum, does not increase the price of anarchy. Sharma and Williamson [22] 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 [10] studied a related question of finding the minimum demand that the Stackelberg leader needs to control in order to enforce an optimal flow.

2 Model

In a network routing game we are given a directed network $G=(V,A)$ and k origin-destination pairs $(s_1,t_1),\dots,(s_k,t_k)$ called *commodities*. For every commodity $i\in[k]$, a demand $r_i>0$ is given that specifies the amount of flow with origin s_i and destination t_i. Let $\mathscr{P}_i$ be the set of all paths from s_i to t_i in G and let $\mathscr{P}=\cup_i\mathscr{P}_i$. A *flow* is a function $f:\mathscr{P}\to\mathbb{R}_+$. The flow f is *feasible* (with respect to r) if for all i, $\sum_{P\in\mathscr{P}_i} f_P=r_i$. 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* denoted by $\ell_a(\cdot)$. For each $a\in A$ the latency function ℓ_a is assumed to be nonnegative, nondecreasing and differentiable. If not indicated otherwise, we also assume that ℓ_a is defined on $[0,\infty)$ and that $x\ell_a(x)$ is a convex function of x. Such functions are called *standard* [16]. 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 $\ell_P(f)=\sum_{a\in P}\ell_a(f_a)$. The triple (G,r,ℓ) is called an *instance*. The *cost* of a flow f is $C(f)=\sum_{P\in\mathscr{P}} f_P\ell_P(f)$. Equivalently, $C(f)=\sum_{a\in A} f_a\ell_a(f_a)$. The feasible flow of minimum cost is called *optimal* and denoted by o. A feasible flow f is a *Nash flow*, or *selfish flow*, if for every $i\in[k]$ and $P,P'\in\mathscr{P}_i$ with $f_P>0$, $\ell_P(f)\le\ell_{P'}(f)$. In particular, if f is a Nash flow, all s_i-t_i paths to which f assigns a positive amount of flow have equal latency. It is well-known that if f_1 and f_2 are Nash flows for the same instance, then $C(f_1)=C(f_2)$, see e.g. [21].

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* is a flow g feasible with respect to $r'=(\alpha_1r_1,\dots,\alpha_kr_k)$, for some $\alpha_1,\dots,\alpha_k\in[0,1]$ such that $\sum_{i=1}^k\alpha_ir_i=\alpha\sum_{i=1}^k r_i$. If $\alpha_i=\alpha$ for all i, g is called a *weak Stackelberg strategy*. 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 $\tilde{\ell}_a(x)=\ell_a(g_a+x)$ for all $a\in A$ and let $\tilde{r}=r-r'$. Then a flow h is *induced by g* if it is a Nash flow for the instance $(G,\tilde{r},\tilde{\ell})$.

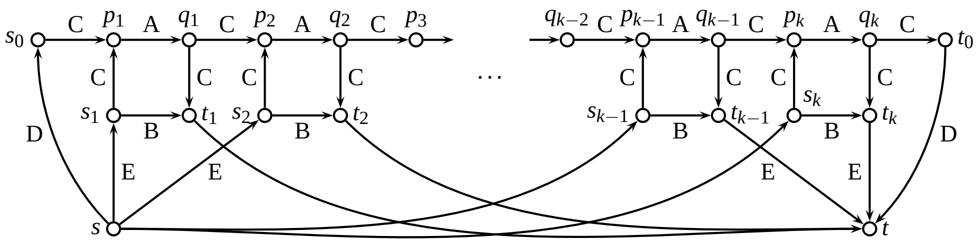

Fig. 1. The graph G_k, used in the proof of Theorem 1. Arcs are labeled with their type.

The Nash flow h can be characterized by the following *variational inequality* [6]: h is a Nash flow induced by g if and only if for all flows x feasible with respect to $\tilde{r}$,

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

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

Due to lack of space, we omit some of the proofs from this extended abstract; details will be given in the full version of the paper.

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.

Theorem 1. *Let $M > 0$ and $\alpha \in (0,1)$. Then, there exists a single-commodity instance $\mathscr{I} = (G, r, \ell, \alpha)$ such that, if g is any Stackelberg strategy for $\mathscr{I}$ 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 the theorem we use the instance $G_k = (V_k, A_k)$ depicted in Figure 1. For a positive integer k, the graph G_k has $4k+4$ nodes. 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* $\{A, B, C, D, E\}$ as indicated in Figure 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. Arcs of type A have the following latency function:

$$\ell_0(x) = \begin{cases} 0, & \text{if } x \leq r_0 \\ 1 - \frac{r_0 + r_1 - x}{r_1}, & \text{if } x > r_0. \end{cases}$$

Although $\ell_0(x)$ is not differentiable in r_0, it can be approximated with arbitrarily small error by standard functions.

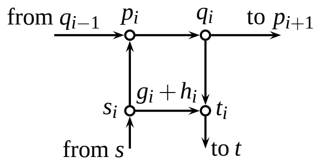

Fig. 2. The ith block of the graph G_k.

For fixed $L, \tau > 0$, 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 1. $C(o) \leq 1$.

Proof. Let P_0 be the path $(s, s_0, p_1, q_1, p_2, \ldots, p_k, q_k, t_0, t)$, and 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 0 on arcs of type A, C, D, E and 1 on arcs of type B. So $C(o) \leq C(f) = k \cdot r_1 = (1+\alpha)/2 \leq 1$. $\square$

The following lemma 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 2. *For any Stackelberg strategy g inducing a Nash flow h, the following hold:*

(i) If a is a type D arc and $g_a + h_a \geq r_0 + \delta/3k^3$, then $C(g+h) \geq M \cdot C(o)$.
(ii) If a is a type E arc and $g_a + h_a \geq r_1 + \delta/3k^3$, then $C(g+h) \geq M \cdot C(o)$.

Proof. We prove statement (i); the proof for (ii) is similar. We have $C(g+h) \geq (g_a + h_a) \cdot \ell_a(g_a+h_a) = (g_a+h_a) \cdot u_{r_0,\delta/3k^3}(g_a+h_a) \geq (r_0 + \delta/3k^3) \cdot M/(\delta/3k^3) \geq M$. The proof follows from Lemma 1. $\square$

For the remainder of the proof we assume that there is no arc satisfying the conditions of Lemma 2; otherwise the theorem follows immediately.

Lemma 3. *For any Stackelberg strategy g inducing a Nash flow h, the following hold:*

(i) For any arc $a = (q_{i-1}, p_i)$, $i \in [k]$, $g_a + h_a \geq r_0 - \delta/k$.
(ii) For any arc $a = (s, s_i)$, $i \in [k]$, $g_a + h_a \geq r_1 - \delta/k$.

We are now ready to conclude the proof of Theorem 1.

Proof (Theorem 1). For any $i \in [k]$, consider the ith block in the graph (Figure 2). Let g_i, h_i be the Stackelberg and selfish flow on the arc (s_i, t_i), respectively. We have two cases:

1. $h_i = 0$: in this case, using Lemma 3, 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 $\ell_0(r_0 + r_1 - 2\delta/k - g_i)$.

2. $h_i > 0$: 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 a Nash flow, we get $\ell_{P_i''}(g+h) \geq \ell_{P_i'}(g+h)$. 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 $\ell_{P_i''}(g+h) \geq \ell_{P_i'}(g+h)$ implies $\ell_{(p_i,q_i)}(g+h) \geq \ell_{(s_i,t_i)}(g+h) = 1$. As a consequence, $\ell_{(p_i,q_i)}(g+h) \geq 1 = \ell_0(r_0+r_1) \geq \ell_0(r_0+r_1-2\delta/k-g_i)$ since g_i and δ/k are nonnegative.

In both cases, $\ell_{(p_i,q_i)}(g+h) \geq \ell_0(r_0+r_1-2\delta/k-g_i) \geq 1 - \frac{g_i+2\delta/k}{r_1}$.

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

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

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

Choosing $\delta < (1-\alpha)/4$, we can conclude that $\ell_{P_0}(g+h) = \Omega(k)$. Together with Lemma 1 and Lemma 3, this gives

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

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

Remark 1. Suppose the Stackelberg leader (e.g., a navigation systems provider) is solely interested in minimizing the travel time of his players (customers), i.e., $C_1(g+h) = \sum_{a \in A} g_a \ell_a(g_a + h_a)$. Our result also implies that even the ratio $C_1(g+h)/C(o)$ can be unbounded, independent of the Stackelberg strategy g.

4 A Bicriteria Bound for General Latency Functions

As we have seen in the previous section, 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 $\mathscr{I} = (G, r, \ell, \alpha)$ to the cost of an optimal flow for the instance $\mathscr{I}^\beta = (G, \beta r, \ell)$ 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)*:

1. Compute an optimal flow o^β for the instance $\mathscr{I}^\beta$.
2. Define the Stackelberg flow by $g := \frac{\alpha}{\beta} o^\beta$.

We prove that the resulting flow induced by the Stackelberg strategy ASCALE satisfies $C(g+h) \leq C(o^\beta)$ if we choose $\beta = 1 + \sqrt{1-\alpha}$. This result can be seen as a generalization of the result by Roughgarden and Tardos 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 [21]. 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).

Theorem 2. *If g is the ASCALE strategy, $C(g+h) \leq \frac{1}{\beta-1} \cdot \left(1 - \frac{\alpha}{\beta}\right) \cdot C(o^\beta)$. Furthermore, this bound is tight.*

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

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

Theorem 3. *Let $g = \alpha o$ be the SCALE strategy for instance $\mathscr{I} = (G, r, \ell, \alpha)$. Define a modified instance $\hat{\mathscr{I}} = (G, r, \hat{\ell}, \alpha)$ with latency functions $\hat{\ell}_a(x) = \ell_a(x/\beta)/\beta$ for every arc a, where $\beta = 1 + \sqrt{1-\alpha}$, and let $\hat{C}(\cdot)$ denote the cost of a flow with respect $\hat{\ell}$. Let $\hat{h}$ be the Nash flow induced by $\hat{g} = g$ in $\hat{\mathscr{I}}$. Then, $\hat{C}(\hat{g} + \hat{h}) \leq C(o)$.*

5 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.

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

Lemma 4. *For any opt-restricted strategy g, $C(g+h) \leq \sum_{a \in A} o_a \ell_a(g_a + h_a)$.*

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:

$$\omega(\ell_a; g_a, \lambda) := \sup_{o_a, h_a \geq 0} \frac{o_a}{g_a + h_a} \cdot \frac{\ell_a(g_a + h_a) - \lambda \ell_a(o_a)}{\ell_a(g_a + h_a)}. \tag{2}$$

(We assume by convention $0/0 = 0$.) In order to bound the price of anarchy, we use the variational inequality (Lemma 4) 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:

$$C(g+h) \leq \sum_{a \in A} \lambda \cdot o_a \ell_a(o_a) + \omega(\ell_a; g_a, \lambda) \cdot (g_a + h_a) \ell_a(g_a + h_a). \tag{3}$$

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, Schulz, and Stier-Moses [3] 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 splittable network games (see Harks [9]).

For a given opt-restricted strategy g we further define $\omega(g,\lambda) = \max_{a\in A} \omega(\ell_a; g_a, \lambda)$. Before we state the main theorem, we need one additional definition. Given an opt-restricted strategy g, the *feasible λ-region* is defined as $\Lambda(g) := \{\lambda \in \mathbb{R}_+ \mid \omega(g,\lambda) < 1\}$. Notice that every $\lambda \in \Lambda(g)$ induces a bound on the price of anarchy.

Theorem 4. *Let $\lambda \in \Lambda(g)$. Then $C(g+h) \leq \frac{\lambda}{1-\omega(g,\lambda)} C(o)$.*

Proof. The proof follows immediately from (3), Lemma 4 and the definition of ω. □

Bounds for SCALE. In the following, we will analyze the SCALE strategy defined by $g = \alpha o$. Let $\mathscr{L}_d$, $d \geq 1$, be a class of continuous, nondecreasing, and standard latency functions satisfying $\ell(cz) \geq c^d \ell(z)$ for all $c \in [0,1]$. $\mathscr{L}_d$ contains, among others, polynomials with nonnegative coefficients and degree at most d. This characterization has been used before by Correa et al. [3].

Lemma 5. *Assume $\lambda \in [0,1]$ and latency functions in $\mathscr{L}_d$. Then, we have*

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

Proof. By the definition of $\omega = \omega(\ell_a; \alpha o_a, \lambda)$:

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

We consider two cases: (i) $\alpha\, o_a + h_a \geq o_a$. Define $\mu := \frac{o_a}{\alpha\, o_a + h_a} \in [0,1]$. We have

$$\begin{aligned}
\omega &= \sup_{o_a, h_a \geq 0, \mu \in [0,1]} \mu \cdot \frac{\ell_a(\alpha\, o_a + h_a) - \lambda \ell_a(\mu\,(\alpha\, o_a + h_a))}{\ell_a(\alpha\, o_a + h_a)} \\
&\leq \max_{\mu \in [0,1]} \mu\,(1 - \lambda\, \mu^d) = \frac{d}{d+1} \cdot \frac{1}{((d+1)\lambda)^{1/d}}.
\end{aligned}$$

where the last inequality follows from the definition of $\mathscr{L}_d$. The second case (ii) $\alpha\, o_a + h_a \leq o_a$ leads to

$$\begin{aligned}
\omega &\leq \sup_{o_a, h_a \geq 0} \frac{o_a}{\alpha\, o_a + h_a} \cdot \frac{\ell_a(\alpha\, o_a + h_a) - \lambda \ell_a(\alpha\, o_a + h_a)}{\ell_a(\alpha\, o_a + h_a)} \\
&\leq \sup_{o_a, h_a \geq 0} \frac{o_a}{\alpha\, o_a + h_a}(1-\lambda) \leq \frac{1}{\alpha}(1-\lambda),
\end{aligned}$$

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

Lemma 6. *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. $\square$

Theorem 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 4 with $\lambda = \lambda_d$. However, in order to apply the theorem, we first need to upper bound $\omega(\alpha o, \lambda_d)$. Using Lemma 5 and Lemma 6, 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 4 to obtain a bound on the price of anarchy given by

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

$\square$

The bound thus obtained gives an improvement with respect to the previously best bounds obtained by Swamy [23].

For the class of $\mathscr{L}_1$ latency functions, which, in particular, contains continuous, nondecreasing, standard, and concave latencies, the above theorem reads as stated in Corollary 2 below. The same bound has been proven by Karakostas and Kolliopoulos [11] for the special case of affine latencies.

Corollary 2. *The price of anarchy of the SCALE strategy for latency functions in $\mathscr{L}_1$ is at most $\left((1+\sqrt{1-\alpha})^2\right) / \left(2(1+\sqrt{1-\alpha})-1\right)$.*

A lower bound for polynomial latency functions of degree d can be obtained by considering generalized Braess graphs [1, 20] (details omitted).

Theorem 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 $\mathscr{L}_d$ is at least $(nc^{1+1/d} + (n-1)\alpha c)/((n-1)c + n^{-d})$.*

Note that the theorem does not fix n, so it is possible to optimize n based on α. For functions in $\mathscr{L}_1$ the stated lower bound pointwise matches the upper bound of Corollary 2 for infinitely many values of α. More precisely, the upper 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 $\alpha \neq 0, 1$.

References

1. M. Babaioff, R. Kleinberg, and C. H. Papadimitriou. Congestion games with malicious players. In *Proc. of the 8th ACM Conf. on Electronic Commerce*, pages 103–112, 2007.
2. D. Braess. Über ein Paradoxon der Verkehrsplanung. *Unternehmenforschung*, 11:258–268, 1968.
3. J. R. Correa, A. S. Schulz, and N. E. Stier-Moses. Selfish routing in capacitated networks. *Mathematics of Operations Research*, 29:961–976, 2004.
4. J. R. Correa, A. S. Schulz, and N. E. Stier-Moses. On the inefficiency of equilibria in congestion games. In *Proc. of the 11th Int. Conf. on Integer Programming and Combinatorial Optimization*, pages 167–181. Springer, 2005.
5. J. R. Correa and N. E. Stier-Moses. Stackelberg routing in atomic network games. Technical report, Columbia Business School, February 2007.
6. S. Dafermos and F. Sparrow. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards, Series B*, 73:91–118, 1969.
7. P. Dubey. Inefficiency of Nash equilibria. *Mathematics of Operations Research*, 11:1–8, 1986.
8. D. Fotakis. Stackelberg strategies for atomic congestion games. In *Proc. of the 15th European Symp. on Algorithms*, pages 299–310. Springer, 2007.
9. T. Harks. Stackelberg strategies and collusion in network games with splittable flow. In *Proc. of the 6th Workshop on Approximation and Online Algorithms (WAOA)*. Springer, 2008.
10. A. Kaporis and P. Spirakis. The price of optimum in Stackelberg games on arbitrary single commodity networks and latency functions. In *Proc. of the 18th ACM Symp. on Parallelism in Algorithms and Architectures*, pages 19–28. ACM Press, 2006.
11. G. Karakostas and S. G. Kolliopoulos. Stackelberg strategies for selfish routing in general multicommodity networks. *Algorithmica*, 2007. To appear.
12. Y. A. Korilis, A. A. Lazar, and A. Orda. Achieving network optima using Stackelberg routing strategies. *IEEE/ACM Transactions on Networking*, 5(1):161–173, 1997.
13. E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In *Proc. of the 16th Symp. on Theoretical Aspects of Computer Science*, pages 404–413. Springer, 1999.
14. V. S. A. Kumar and M. V. Marathe. Improved results for Stackelberg scheduling strategies. In *Proc. of the 33rd Int. Colloquium of Automata, Languages and Programming*, pages 776–787. Springer, 2002.
15. G. Perakis. The price of anarchy when costs are non-separable and asymmetric. In *Proc. of the 10th Int. Conference on Integer Programming and Combinatorial Optimization*, pages 46–58. Springer, 2004.
16. T. Roughgarden. The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences*, 67:341–364, 2002.
17. T. Roughgarden. *Selfish Routing*. PhD thesis, Cornell University, 2002.
18. T. Roughgarden. Stackelberg scheduling strategies. *SIAM Journal on Computing*, 33(2):332–350, 2004.
19. T. Roughgarden. *Selfish Routing and the Price of Anarchy*. The MIT Press, 2005.
20. T. Roughgarden. On the severity of Braess's paradox: Designing networks for selfish users is hard. *Journal of Computer and System Sciences*, 72(5):922–953, 2006.
21. T. Roughgarden and E. Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002.
22. Y. Sharma and D. Williamson. Stackelberg thresholds in network routing games or the value of altruism. In *Proc. of the 8th ACM Conf. on Electronic Commerce*, pages 93–102, 2007.
23. C. Swamy. The effectiveness of Stackelberg strategies and tolls for network congestion games. In *Proc. of the 18th ACM-SIAM Symp. on Discrete Algorithms*, pages 1133–1142. SIAM, 2007.