

FAST, FAIR, AND EFFICIENT FLOWS IN NETWORKS

JOSÉ R. CORREA*, ANDREAS S. SCHULZ**, AND NICOLÁS E. STIER-MOSES***

* *School of Business, Universidad Adolfo Ibáñez,*

Av. Presidente Errázuriz 3485, Las Condes, Santiago, Chile; correa@uai.cl

** *Sloan School of Management, Massachusetts Institute of Technology,*

Office E53-361, 77 Massachusetts Ave., Cambridge, MA 02139, USA; schulz@mit.edu

*** *Graduate School of Business, Columbia University,*

Uris Hall, Room 418, 3022 Broadway Ave., New York, NY 10027, USA;

stier@gsb.columbia.edu

ABSTRACT. We study the problem of minimizing the maximum latency of flows in networks with congestion. We show that this problem is NP-hard, even when all arc latency functions are linear and there is a single source and sink. Still, an optimal flow and an equilibrium flow share a desirable property in this situation: all flow-carrying paths have the same length; i.e., these solutions are “fair,” which is in general not true for optimal flows in networks with nonlinear latency functions. In addition, the maximum latency of the Nash equilibrium, which can be computed efficiently, is within a constant factor of that of an optimal solution. That is, the so-called price of anarchy is bounded. In contrast, we present a family of instances with multiple sources and a single sink for which the price of anarchy is unbounded, even in networks with linear latencies. Furthermore, we show that an s - t flow that is optimal with respect to the average latency objective is near optimal for the maximum latency objective, and it is close to being fair. Conversely, the average latency of a flow minimizing the maximum latency is also within a constant factor of that of a flow minimizing the average latency.

Key words and phrases. System Optimum, Nash Equilibrium, User Equilibrium, Selfish Routing, Price of Anarchy, Multicriteria Optimization, Multicommodity Flows.

MSC 2000: Primary: 90B10, 90B20, 91A10. Secondary: 90C35; 90C25, 90C27, 91A13, 91A43.

OR/MS classification: Primary: Networks/graphs: Multicommodity, Theory. Secondary: Games: Noncooperative; Mathematics: Combinatorics; Transportation: Models.

An extended abstract of a preliminary version of this paper appeared in the proceedings of the 10th International Conference on Integer Programming and Combinatorial Optimization (Correa, Schulz, and Stier-Moses 2004a).

1. INTRODUCTION

We study static network flow problems in which each arc possesses a latency function that describes the common delay experienced by the flow on that arc as a function of the flow rate. Load-dependent arc costs have a variety of applications in situations in which one wants to model congestion effects, which are bound to appear, e.g., in communication networks, vehicular traffic, supply chain management, or evacuation planning. In this context, a unit of flow frequently stands for a huge number of “users” (or “agents”), which might represent data packages in the Internet, drivers on a highway system, product components in a supply chain, or individuals fleeing from an area struck by disaster. Depending on the concrete circumstances, the operators or overseeing authorities of these networks can pursue a variety of system objectives. For instance, they might elect to minimize the average latency, they might aim at minimizing the maximum latency, or they might try to ensure that users having the same origin and destination experience essentially the same latency. In fact, an ideal solution might be simultaneously optimal or near-optimal with respect to all three objectives. We establish the existence of such flows.

The Model. We consider a directed graph $G = (N, A)$ together with a set of source-sink pairs $K \subseteq N \times N$. For each terminal pair $k = (s_k, t_k) \in K$, let \mathcal{P}_k be the set of directed (simple) paths in G from s_k to t_k , and let $d_k > 0$ be the demand rate associated with commodity k . Let $\mathcal{P} := \bigcup_{k \in K} \mathcal{P}_k$ be the set of all paths between terminal pairs, and let $d := \sum_{k \in K} d_k$ be the total demand. A feasible flow f assigns a nonnegative and possibly fractional value f_P to every path $P \in \mathcal{P}$ such that $\sum_{P \in \mathcal{P}_k} f_P = d_k$ for all $k \in K$. In the context of single-source single-sink instances, we will drop the subindex k and talk about s - t -flows. Each arc a has a load-dependent latency function $\ell_a(\cdot)$. We generally assume that the functions $\ell_a : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ are nondecreasing and continuous; at times, it will be convenient to assume that they are differentiable as well, which we will point out in each case. We define the latency of a path $P \in \mathcal{P}$ under a given flow f as $\ell_P(f) := \sum_{a \in P} \ell_a(\sum_{Q \in \mathcal{P}: Q \ni a} f_Q)$.

The maximum latency $L(f)$ of a feasible flow f is defined as

$$L(f) := \max\{\ell_P(f) : P \in \mathcal{P}, f_P > 0\} .$$

We call a feasible flow that minimizes the maximum latency a *min-max flow* and denote it by f^{MM} . The *maximum latency problem* consists of finding a min-max flow.

The average latency of a feasible flow f is defined as

$$C(f) := \frac{1}{d} \sum_{P \in \mathcal{P}} \ell_P(f) f_P .$$

We refer to an optimal solution with respect to this objective function as a *system optimum* and denote it by f^{SO} .

We also consider a noncooperative network game of infinitely many players where every player controls an infinitesimal amount of flow (Wardrop 1952). A feasible flow f^{NE} is at *Nash equilibrium* (is a *user equilibrium*) if for every $k \in K$ and any two paths $P_1, P_2 \in \mathcal{P}_k$ with $f_{P_1}^{\text{NE}} > 0$, $\ell_{P_1}(f^{\text{NE}}) \leq \ell_{P_2}(f^{\text{NE}})$. In other words, all flow-carrying s_k - t_k -paths have equal (and actually minimal) latency. In particular, equilibrium flows are “fair,” i.e., they have unfairness 1, where the *unfairness* of a feasible flow f is defined as

$$U(f) := \max \left\{ \frac{\ell_{P_1}(f)}{\ell_{P_2}(f)} : P_1, P_2 \in \mathcal{P}_k, f_{P_1}, f_{P_2} > 0, k \in K \right\} .$$

Main Results. For linear latency functions and a single source and sink, we prove in Section 4 the existence of a min-max flow that is fair; i.e., its unfairness is 1. Moreover, the average latency of any min-max flow is within a factor of 4/3 of that of an optimum. As attractive as such a flow might be, we also show in Section 3 that computing a min-max flow is NP-hard. Furthermore, we exhibit a surprising difference between linear and nonlinear latency functions: while min-max flows remain near-optimal with respect to the average latency, their unfairness is in general greater than 1.

It is well known that system optima are in general unfair. We establish a tight bound that quantifies the severity of this effect. Namely, we show in Section 5 that the latency of

any one user is within a constant factor of that of any other user. In particular, for single-source single-sink networks, the maximum latency of the system optimum is within the same constant factor of that of a min-max flow. This constant factor only depends on the class of allowable latency functions. For instance, its value is 2 for the case of linear latencies.

Linear latencies are sufficient for certain congestion phenomena to occur. One interesting example is Braess' paradox (1968), which refers to the fact that the addition of an arc to a network can actually increase the latency of all users if they act selfishly and independently. While the inefficiency of user equilibria and hence the severity of Braess' paradox had previously been bounded in terms of the average latency (Roughgarden and Tardos 2002), it turns out that it is also bounded with respect to the maximum latency. A user equilibrium therefore serves as another flow that is optimal or close to optimal for the three objectives introduced above.

Most of our bounds hold for more general classes of latency functions. In particular, there exist s - t flows that are simultaneously optimal or near optimal with respect to all three criteria: maximum latency, average latency, and unfairness. In fact, this property is shared by system optima, Nash equilibria, and—to some extent—min-max flows, albeit with different bounds. Table 1 presents the bounds in the single-source single-sink case with linear latencies. With one exception, the first entry in each cell presents a worst-case bound on the ratio of the value of a flow associated with the corresponding row to the value of an optimal flow for the objective function associated with the corresponding column. The exception is the value of 1 for the unfairness of a min-max flow. We only show the existence of at least one min-max flow that is fair; others might be unfair. However, we also outline an algorithm that turns any min-max flow into a fair one. The second entry refers to the theorem in this paper in which the respective result is proved. All bounds are tight, as examples provided after each theorem demonstrate. With the exception of the following two results, the remaining ones are new. Roughgarden and Tardos (2002) first proved the upper bound of $4/3$, stated in Theorem 5.1, on the ratio of the average latency of a Nash equilibrium to that of a system optimum. (See Correa, Schulz, and Stier-Moses (2004b, 2005)

TABLE 1. Summary of results for single-source single-sink networks with linear latency functions.

	maximum latency	average latency	unfairness
Min-max flow	1	$4/3$ Thm. 5.6	1^1 Thm. 4.1
System optimum	2 Thm. 5.5	1	2 Thm. 4.2
Nash equilibrium	$4/3$ Thm. 5.2	$4/3$ Thm. 5.1	1

for a simpler proof). Weitz (2001) observed that this bound carries forward to the maximum latency objective for the case of a single source and sink; we present a generalization of this observation to multicommodity flows in Theorem 5.2. Proposition 5.3 shows that, in networks with multiple sources and a single sink, the ratio of the maximum latency of a Nash equilibrium to that of a min-max flow is in general not bounded by a constant, even when latency functions are linear.

In Section 6 we analyze a fourth objective function: minimizing the maximum latency of all flow-carrying arcs. We show that optimal solutions with respect to this objective can be arbitrarily bad for the other three objectives, and vice-versa.

Related Work. Minimizing the maximum latency is common in the network routing and evacuation literature. Evacuation problems have been studied as dynamic flow problems since the seminal work of Ford and Fulkerson (1958), who proposed to minimize the time by which the network is cleared. There has recently been an increased activity in analyzing dynamic flows; see, e.g., Hoppe and Tardos (1994), Fleischer and Skutella (2003), and the surveys by Aronson (1989) and Powell, Jaillet, and Odoni (1995). Particularly relevant to our work is a paper by Jarvis and Ratliff (1982) who showed the existence of a dynamic flow that is simultaneously optimal for three objectives. A common drawback of these models is the assumption of constant traversal times, which is oftentimes not realistic. One notable exception is the work by Köhler and Skutella (2005), who considered the quickest flow problem with load-dependent transit times.

¹In contrast to the other entries in the table, this is not a worst-case, but a best-case result; there exists a min-max flow that is fair.

The maximum latency objective has also been considered in the context of telecommunication networks. The proposed models are closer in nature to the one considered here because they are static and include load-dependent delays. Koutsoupias and Papadimitriou (1999) studied the maximum latency objective in a static network consisting of multiple parallel arcs that connect a single source with a single sink. They introduced a noncooperative game with finitely many users where a user's action is to select an arc, and analyzed the inefficiency of its Nash equilibria. Papadimitriou (2001) later referred to this inefficiency as the *price of anarchy*. Refined results on the same model were obtained, among others, by Mavronicolas and Spirakis (2001), Koutsoupias, Mavronicolas, and Spirakis (2003), and Czumaj and Vöcking (2004). The main conclusion is that, under certain assumptions, the maximum latency of an equilibrium is not too large compared to that of the best coordinated solution. In particular, Czumaj, Krysta, and Vöcking (2002) showed that an equilibrium minimizes the maximum latency, provided there are infinitely many players. This is in sharp contrast with the results of this paper, in which we consider arbitrary network topologies. For more details on the parallel arc model, we refer the reader to the survey by Czumaj (2004).

Roughgarden and Tardos (2002), Roughgarden (2003b), Schulz and Stier-Moses (2003), and Correa, Schulz, and Stier-Moses (2004b, 2005) studied the price of anarchy with respect to the average travel time in general networks and for different classes of latency functions. In particular, if \mathcal{L} is the set of allowable latency functions, the ratio of the average travel time of a Nash equilibrium to that of a system optimum is bounded by $\alpha(\mathcal{L})$, where $\alpha(\mathcal{L})$ is a constant that only depends on \mathcal{L} . As shown in Table 1, if \mathcal{L} only contains linear functions, $\alpha(\mathcal{L}) = 4/3$. We will elaborate on and make use of this bound in Section 5.

Weitz (2001) observed that these price-of-anarchy results are also valid for the maximum latency objective, as long as the considered instance has a single source and sink. He also showed that Nash equilibria can be arbitrarily bad in multiple commodity networks. In Section 5 we prove that this is also true in networks with multiple sources and a single sink. Roughgarden (2004) gave a tight bound for the single-source single-sink case that depends on the size of the network.

In the context of Section 4, we should point out that there exist multiple (nonequivalent) definitions of (un)fairness. The definition we use here arises from the competition between different agents in a network game. Roughgarden (2002) introduced a less pessimistic version of unfairness, namely the ratio of the maximum latency of a system optimum to the latency of a Nash equilibrium; we later obtain his bound as a corollary to a more general result. Jahn et al. (2005) considered the definition of unfairness presented here; they looked at flows that minimize the total travel time among those with bounded unfairness.

2. AN EXAMPLE

Before we present our results we give an example to exhibit some characteristics of the different objective functions and their corresponding optimal solutions. The instance, depicted in Figure 1, has a single source and sink, and quadratic latency functions. The unit demand has to be shipped through two equal stages with two parallel arcs each. Figure 2 shows a min-max flow, a system optimum, and a Nash equilibrium for this instance.

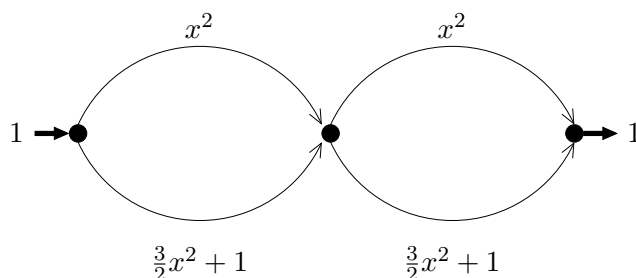


FIGURE 1. An instance with quadratic latency functions.

The min-max flow can be computed by solving a minimization problem with a single variable. Indeed, the “bottom-bottom” path is too long to carry any flow, and, making use of the symmetry of the instance, the “top-bottom” and “bottom-top” paths have to carry

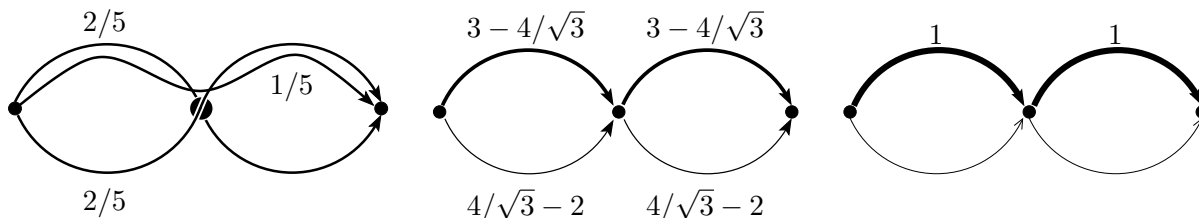


FIGURE 2. A min-max flow, a system-optimal flow, and a Nash flow for the instance in Figure 1 (from left to right).

the same flow x . Assigning $1 - 2x$ units of flow to the path “top-top,” the unique optimum occurs for $x = 2/5$. Computing a Nash equilibrium analytically is straightforward for this instance: routing all flow along the path “top-top” provides no incentive for users to deviate. As we will see in Section 4, a system optimum is an equilibrium with respect to modified latencies that incorporate the externalities. With the modified latencies in hand, we solve a system of quadratic equations and get that the flow on the top arcs equals $3 - 4/\sqrt{3} \approx 0.69$ and that of the bottom arcs equals $4/\sqrt{3} - 2 \approx 0.31$.

Table 2 displays the objective values of the three solutions. Notice that Figure 2 shows the Nash equilibrium and the system optimum as a flow on arcs: any path decomposition provides a correct solution. However, different flow decompositions can lead to different unfairness values and maximum latencies. Where needed, the table indicates two values arising from the two extreme decompositions. The min-max flow is instead given as a flow on paths. This is necessary because presenting a flow on arcs is not enough to solve the maximum latency problem, as we are going to show in Section 3.

TABLE 2. Objective values of the three solution concepts for the instance shown in Figure 1.

	maximum latency	average latency	unfairness
Min-max flow	1.6	1.424	2.222...
System optimum	{2.287..., 1.620...}	1.366...	{2.397..., 1.698...}
Nash equilibrium	2	2	1

3. COMPUTATIONAL COMPLEXITY

In our model, both a system optimum and a Nash equilibrium can be computed as solutions to convex programs. (For the system optimum, this statement is only true if we assume that $C(f)$ is convex, which is normally the case in applications. For the Nash equilibrium, monotonicity and continuity suffice; see Beckmann, McGuire, and Winsten (1956) for details.) Convex programming problems can be solved up to an arbitrarily small additive error in polynomial time; see, e.g., Vavasis (1991), Potra and Ye (1993), or Grötschel, Lovász, and Schrijver (1993) for details. One cannot hope to do better because, as the example

given in Section 2 shows, an optimal solution may need irrational numbers, and so may an equilibrium.

On the other hand, it follows from the work of Köhler and Skutella (2005) on the quickest s - t -flow problem with load-dependent transit times that the maximum latency problem considered here is NP-hard (though not necessarily in NP) when latencies include arbitrary nonlinear functions or when there are explicit arc capacities. Lemma 3.1 below implies that the maximum latency problem with linear latencies is in NP, while Theorem 3.3 establishes its NP-hardness, even in the case of a single source and a single sink, and in the absence of arc capacities.

The following observation is key to establish the membership of the maximum latency problem in NP. Interestingly, it does not follow from ordinary flow decomposition because it is not clear how to convert a flow on arcs into a path flow such that the latency of the resulting paths remains bounded; in fact, it is a consequence of Theorem 3.3 that this problem is NP-hard.

Lemma 3.1. *Let f be a feasible flow for a multicommodity flow network with load-dependent arc latencies. Then, there exists another feasible flow f' such that $L(f') \leq L(f)$, and f' uses at most $|A|$ paths for each source-sink pair.*

Proof. The proof is based on that of Carathéodory's Theorem (see, e.g., Schrijver 1998, p. 94). Consider an arbitrary commodity $k \in K$. Let P_1, \dots, P_r be s_k - t_k -paths such that $f_{P_i} > 0$ for $i = 1, \dots, r$, and $\sum_{i=1}^r f_{P_i} = d_k$. Slightly overloading notation, we let P_1, \dots, P_r also denote the arc incidence vectors of these paths. Let's assume that $r > |A|$. (Otherwise we are done.) Hence, the vectors P_1, \dots, P_r are linearly dependent and $\sum_{i=1}^r \lambda_i P_i = 0$ has a nonzero solution. Without loss of generality, $\lambda_r \neq 0$. We define a new flow f'' (not necessarily feasible) by setting $f''_{P_i} := f_{P_i} - \frac{\lambda_i}{\lambda_r} f_{P_r}$ for $i = 1, \dots, r$, and $f''_P := f_P$ for all other paths P . Notice that under f'' , the flow on arcs does not change:

$$\sum_{i=1}^r P_i f''_{P_i} = \sum_{i=1}^{r-1} P_i f_{P_i} - \sum_{i=1}^{r-1} \frac{\lambda_i}{\lambda_r} P_i f_{P_r} = \sum_{i=1}^r P_i f_{P_i} .$$

Here, we used the linear dependency for the last equality. In particular, $L(f'') \leq L(f)$. Let us consider a convex combination f' of f and f'' that is nonnegative and uses fewer paths than f . Note that such a flow always exists because $f''_{P_r} = 0$, and the flow on some other paths P_1, \dots, P_{r-1} might be negative. Moreover, $L(f') \leq L(f)$, too. If f' still uses more than $|A|$ paths between s_k and t_k , we can iterate this process so long as necessary to prove the claim. \square

We remind the reader that the decision problem of a minimization problem has a “yes” or “no” answer. Its input consists of an instance of the associated minimization problem together with some threshold value β , and the question is whether there exists a feasible solution of value at most β .

Corollary 3.2. *The decision version of the maximum latency problem with linear latency functions is in NP.*

Proof. Lemma 3.1 shows the existence of a succinct certificate. Indeed, for any flow f with maximum latency $L(f)$ there is another flow f' of smaller or equal maximum latency that uses at most $|K| \cdot |A|$ paths. Moreover, it suffices to list the paths with positive flow because the flow values can be recovered by solving a linear program similar to the one in (1) below. \square

We will now prove that the maximum latency problem is in fact NP-hard. We present a reduction from PARTITION:

Given: A set of n positive integers q_1, \dots, q_n .

Question: Is there a subset $I \subset \{1, \dots, n\}$ such that $\sum_{i \in I} q_i = \sum_{i \notin I} q_i$?

Theorem 3.3. *The decision version of the maximum latency problem is NP-complete, even when all latencies are linear functions and the network has a single source-sink pair.*

Proof. Given an instance of PARTITION, we define an instance of the maximum latency problem as depicted in Figure 3. The network consists of nodes $0, 1, \dots, n$. There is a unit demand between the source node 0 and the sink node n . For $i = 1, \dots, n$, the nodes $i-1$ and

i are connected with two arcs, namely a_i with latency $\ell_{a_i}(f_{a_i}) = q_i f_{a_i}$ and \tilde{a}_i with latency $\ell_{\tilde{a}_i}(f_{\tilde{a}_i}) = q_i$.

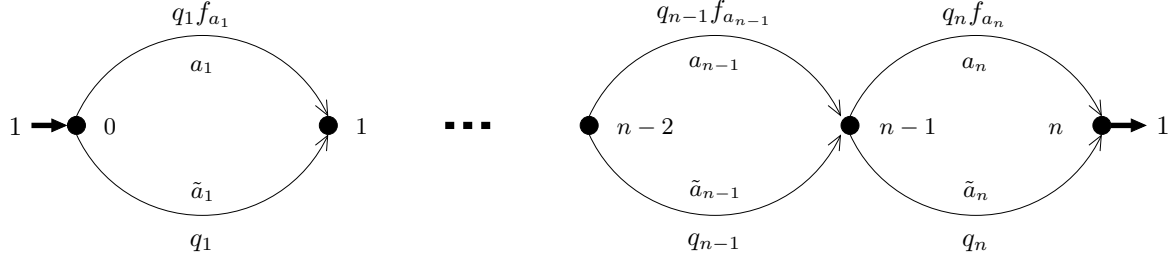


FIGURE 3. Instance used in the reduction from PARTITION.

Let $L := \frac{3}{4} \sum_{i=1}^n q_i$. Notice that any system optimum f^{SO} has cost L , and $f_a^{\text{SO}} = 1/2$ for all $a \in A$. We claim that the given instance of PARTITION is a YES-instance if and only if there is a solution to the maximum latency problem of maximum latency equal to L . Indeed, if there is a partition I , the solution that routes half a unit of flow along the 0 - n -path composed of arcs a_i , $i \in I$, and \tilde{a}_i , $i \notin I$, and the other half along the complementary path has maximum latency L .

To prove the other direction, assume that we have a feasible flow f of maximum latency equal to L . Therefore, $C(f) \leq L$, which implies $C(f) = L$ because f cannot be better than the system optimum. As the system optimum of this instance is unique (as a flow on arcs), $f_a = 1/2$ for all $a \in A$. Take any path P such that $f_P > 0$ and partition its arcs such that I contains the indices of the arcs $a_i \in P$. Then,

$$\frac{3}{4} \sum_{i=1}^n q_i = L = \ell_P(f) = \sum_{i \in I} \frac{q_i}{2} + \sum_{i \notin I} q_i,$$

and subtracting the left-hand side from the right-hand side yields $\sum_{i \in I} q_i/4 = \sum_{i \notin I} q_i/4$. \square

The following corollary states that it is difficult to compute a flow decomposition so that the lengths of the resulting paths are small.

Corollary 3.4. *Let f be a flow in an s - t -network with linear latencies. Let $(f_a)_{a \in A}$ be the associated flow on arcs. Given just $(f_a)_{a \in A}$ and $L(f)$, it is NP-hard to compute a path*

decomposition of this arc flow into a flow f' such that $L(f') \leq L(f)$. In particular, it is NP-hard to recover a min-max flow even though its arc values are given.

Note that Corollary 3.4 neither holds for the system optimum nor the Nash equilibrium. In both cases any flow derived from an ordinary flow decomposition is indeed an optimal flow or an equilibrium flow, respectively. Nevertheless, an arbitrary decomposition of a system-optimal flow need not be good with respect to the maximum latency objective. To see that, consider the instance described in the proof of Theorem 3.3. The flow that routes $1/2$ along the path a_1, a_2, \dots, a_n and $1/2$ along the path $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_n$ is indeed a system optimum, but its maximum latency is $\sum_{i=1}^n q_i$, whereas the optimal solution has value $\frac{3}{4} \sum_{i=1}^n q_i$. In Section 5, we prove a tight worst-case bound for the maximum latency of a system optimum.

Let us finally mention that Theorem 4.3 in Köhler and Skutella (2005) implies that the maximum latency problem is APX-hard when latencies can be arbitrary nonlinear functions or when there are explicit arc capacities.

4. FAIRNESS

Nash equilibria are fair by definition. Indeed, all flow-carrying paths between the same source and sink have equal latency. The next result establishes a similar property for min-max s - t -flows in the case of linear latencies: a fair min-max flow always exists. The difference between a Nash equilibrium and a fair min-max flow is that the latter may leave paths unused that are shorter than the ones carrying flow, which cannot happen in equilibrium. The following result is not true for nonlinear latencies, as we shall see later.

Theorem 4.1. *Every instance of the single-source single-sink maximum latency problem with linear latency functions has an optimal solution that is fair.*

Proof. Consider an instance with demand d and latency functions $\ell_a(f_a) = q_a f_a + r_a$, for $a \in A$. Among all min-max flows, let f^{MM} be one that uses the smallest number of paths.

Let P_1, P_2, \dots, P_u be these paths. Consider the following linear program:

$$\min \quad z \tag{1a}$$

$$\text{s.t.} \quad \sum_{a \in P_i} \left(q_a \left(\sum_{P_h \ni a} f_{P_h} \right) + r_a \right) \leq z \quad \text{for } i = 1, \dots, u, \tag{1b}$$

$$\sum_{i=1}^u f_{P_i} = d \tag{1c}$$

$$f_{P_i} \geq 0 \quad \text{for } i = 1, \dots, u. \tag{1d}$$

Note that this linear program has $u + 1$ variables. Furthermore, by construction, it has a feasible solution f with $z = L(f)$, and there is no solution with $z < L(f)$. Therefore, an optimal basic feasible solution gives a min-max flow that satisfies with equality u of the $2u$ inequalities (1b) and (1d). As $f_{P_i} > 0$ for all i because of the minimality assumption, all inequalities (1b) have to be tight. \square

A byproduct of this proof is that an arbitrary flow can be transformed into a fair one that uses a subset of its paths without increasing its maximum latency. In fact, just solve the corresponding linear program. An optimal basic feasible solution will either be fair or it will use fewer paths. In the latter case, eliminate all paths with zero flow and repeat until a fair solution is found.

With nonlinear latency functions, all min-max flows can be unfair. The instance depicted in Figure 4 features high unfairness with latencies that are polynomials of degree $p \geq 2$.

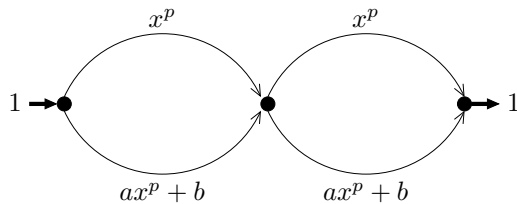


FIGURE 4. Instance with nonlinear latencies illustrating that a fair min-max flow may not exist.

When $a = (1 + \varepsilon)^{p-1}$ and $b = 2 - \left(\frac{1+\varepsilon}{2+\varepsilon}\right)^{p-1} - \delta$ for some $\varepsilon > 0$ and $\delta > 0$ such that $b > 1$, the unique min-max flow routes $\frac{1}{2+\varepsilon}$ units of flow along the “top-bottom” and “bottom-top” paths, respectively, and $\frac{\varepsilon}{2+\varepsilon}$ units of flow along the “top-top” path. It is not hard to see that this flow is optimal. Indeed, the “bottom-bottom” path is too long to carry any flow. Moreover, by symmetry, the “top-bottom” and “bottom-top” paths have to carry the same amount of flow. Letting the variable x denote the flow on the “top-top” path, the flow on both top arcs is $\frac{1+x}{2}$, and that of both bottom arcs is $\frac{1-x}{2}$. Summing along paths, we get that the latency of the “top-top” path is $2\left(\frac{1+x}{2}\right)^p$, which is always smaller than that of the other two paths, which is $\left(\frac{1+x}{2}\right)^p + a\left(\frac{1-x}{2}\right)^p + b$. Finally, we compute the optimal solution of $\min \left\{ \left(\frac{1+x}{2}\right)^p + a\left(\frac{1-x}{2}\right)^p + b : 0 \leq x < 1 \right\}$ and get $f_{\text{top-top}}^{\text{MM}} = \frac{\varepsilon}{2+\varepsilon}$, as specified before.

Let us compute the unfairness of this solution. The “top-top” path has latency equal to $2\left(\frac{1+\varepsilon}{2+\varepsilon}\right)^p$, which tends to $\left(\frac{1}{2}\right)^{p-1}$ as $\varepsilon \rightarrow 0$. The latency of the other two paths used by the optimum is equal to $2 - \delta$. Therefore, the unfairness of this min-max flow is arbitrarily close to 2^p .

A typical argument against using the system optimum in the design of route-guidance devices for traffic assignment is that, in general, it assigns some drivers to unacceptably long paths in order to use shorter paths for other drivers; see, e.g., Beccaria and Bolelli (1992). The following theorem quantifies the severity of this effect by characterizing the unfairness of the system optimum. It turns out that there is a relation to earlier work by Roughgarden (2002), who compared the maximum latency of a system optimum in a single-sink single-source network to the latency of a Nash equilibrium. He showed that for a given class \mathcal{L} of latency functions, this ratio is bounded from above by $\gamma(\mathcal{L})$, where $\gamma(\mathcal{L})$ is defined to be the smallest value that satisfies $\ell^*(x) \leq \gamma(\mathcal{L})\ell(x)$ for all $\ell \in \mathcal{L}$ and all $x \geq 0$. Here, $\ell^*(x) := \ell(x) + x\ell'(x)$ is the function that turns a system optimum for the original instance into a Nash equilibrium of an instance in which the latencies are replaced by ℓ^* (Beckmann, McGuire, and Winsten 1956). For instance, $\gamma(\{\text{polynomials of degree } p\}) = p + 1$. We prove that the unfairness of a system optimum is in fact bounded by the same constant, even for

general instances with multiple commodities. The same result was independently obtained by Roughgarden (2003a).

Theorem 4.2. *Let \mathcal{L} be a family of differentiable and nondecreasing latency functions. If f^{SO} denotes a system optimum in a multicommodity flow network with arc latency functions drawn from \mathcal{L} , then the unfairness of f^{SO} is bounded from above by $\gamma(\mathcal{L})$.*

Proof. We will prove the result for the single-source single-sink case. The extension to the general case is straightforward. As a system optimum is a Nash equilibrium with respect to latencies ℓ^* , there exists a value L^* such that $\ell_P^*(f^{\text{SO}}) = L^*$ for all paths $P \in \mathcal{P}$ with $f_P^{\text{SO}} > 0$. From the definitions of ℓ^* and $\gamma(\mathcal{L})$, we have that $\ell_a(x) \leq \ell_a^*(x) \leq \gamma(\mathcal{L})\ell_a(x)$ for all x and a . Let $P_1, P_2 \in \mathcal{P}$ be two arbitrary paths with $f_{P_1}^{\text{SO}}, f_{P_2}^{\text{SO}} > 0$. Hence, $\ell_{P_1}(f^{\text{SO}}) \leq L^*$ and $\ell_{P_2}(f^{\text{SO}}) \geq L^*/\gamma(L)$. It follows that $\ell_{P_1}(f^{\text{SO}})/\ell_{P_2}(f^{\text{SO}}) \leq \gamma(L)$. \square

An immediate corollary is that users in a system optimum f^{SO} of a network with a single source and a single sink cannot travel too long compared to any other flow. This strengthens Roughgarden's earlier bound which established that $L(f^{\text{SO}}) \leq \gamma(\mathcal{L})L(f^{\text{NE}})$, where f^{NE} is a Nash equilibrium.

Corollary 4.3. *Let \mathcal{L} be a family of nondecreasing and differentiable latency functions. If f^{SO} denotes a system optimum of a single-source single-sink network with arc latency functions drawn from \mathcal{L} , then $L(f^{\text{SO}}) \leq \gamma(\mathcal{L})L(f)$ for any feasible flow f .*

Proof. Notice that

$$L(f^{\text{SO}}) \leq \gamma(\mathcal{L}) \min\{\ell_P(f^{\text{SO}}) : P \in \mathcal{P}, f_P^{\text{SO}} > 0\} \leq \gamma(\mathcal{L})C(f^{\text{SO}}) \leq \gamma(\mathcal{L})C(f) \leq \gamma(\mathcal{L})L(f),$$

where the first inequality follows from Theorem 4.2. \square

The example shown in Figure 5 proves that the bound given in Theorem 4.2 is tight. Indeed, it is easy to see that the unique system optimum routes half of the demand along each arc, implying that the unfairness is $\ell^*(d/2)/\ell(d/2)$. Taking the supremum of that ratio over $d \geq 0$ and $\ell \in \mathcal{L}$ yields $\gamma(\mathcal{L})$.

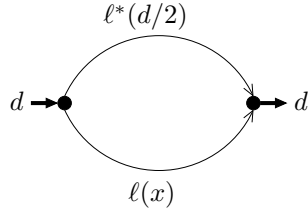


FIGURE 5. Instance showing that the bound in Theorem 4.2 is tight.

5. PRICE OF ANARCHY RESULTS

Nash equilibria in general and those of network games in particular are known to be inefficient, as evidenced by Braess’ paradox (1968). Koutsoupias and Papadimitriou (1999) suggested measuring this degradation in performance, which results from the lack of central coordination, by the worst-case ratio of the value of an equilibrium to that of an optimum. This ratio is known as the “price of anarchy,” a phrase coined by Papadimitriou (2001). It is quite appealing (especially for evacuation situations) that in the network game considered here, the price of anarchy is small; i.e., the selfishness of users actually drives the solution close to optimality. Recall that the Nash equilibrium results from everyone choosing a shortest path under the prevailing conditions.

In order to derive a bound on the price of anarchy for the maximum latency objective, we use a corresponding bound for the average latency of Nash equilibria, which was first proved for linear latency functions by Roughgarden and Tardos (2002) and then extended to different classes of latency functions by Roughgarden (2003b), and Correa, Schulz, and Stier-Moses (2004b, 2005).

Theorem 5.1 (Roughgarden 2003b; Correa, Schulz, and Stier-Moses 2005). *Consider an instance with latency functions drawn from a family \mathcal{L} of nondecreasing and continuous latency functions. Then, the ratio of the average travel time of a Nash equilibrium f^{NE} to that of a system optimum f^{SO} is bounded from above by $\alpha(\mathcal{L})$, i.e., $C(f^{\text{NE}}) \leq \alpha(\mathcal{L}) C(f^{\text{SO}})$, where*

$$\alpha(\mathcal{L}) := \left(1 - \sup_{\ell \in \mathcal{L}, 0 \leq x \leq d} \left\{ \frac{x(\ell(d) - \ell(x))}{d \ell(d)} \right\} \right)^{-1}.$$

As mentioned in the introduction, $\alpha(\mathcal{L}) = 4/3$ for linear functions. For polynomials of degree 2 with nonnegative coefficients, $\alpha(\mathcal{L})$ equals 1.626; for those with degree 3, $\alpha(\mathcal{L}) = 1.896$; in general, $\alpha(\mathcal{L}) = \Theta(p/\ln p)$ for polynomials of degree p .

Weitz (2001) observed that in networks with only one source and one sink, any upper bound on the price of anarchy for the average latency is an upper bound on the price of anarchy for the maximum latency. We include a multicommodity version of this result.

Theorem 5.2. *Consider a set \mathcal{L} of continuous and nondecreasing latency functions, and a multicommodity flow network with latency functions drawn from \mathcal{L} . Let f^{NE} be a Nash equilibrium and f^{MM} a min-max flow. For each commodity $k \in K$, $L_k(f^{\text{NE}}) \leq \frac{d}{d_k} \alpha(\mathcal{L}) L(f^{\text{MM}})$, where L_k is the maximum latency incurred by commodity k , d_k is its demand rate, and d is the total demand.*

Proof. Let f^{SO} be a system optimum. Then,

$$d_k L_k(f^{\text{NE}}) \leq d C(f^{\text{NE}}) \leq d \alpha(\mathcal{L}) C(f^{\text{SO}}) \leq d \alpha(\mathcal{L}) C(f^{\text{MM}}) \leq d \alpha(\mathcal{L}) L(f^{\text{MM}}) .$$

Here, the first inequality holds because f^{NE} is a Nash equilibrium, the second inequality follows from Theorem 5.1, the third one comes from the optimality of f^{SO} , and the last one just says that the average latency is less than the maximum latency. \square

The proof of Theorem 5.2 implies that, if for a given single-source single-sink instance an equilibrium flow f^{NE} happens to be a system optimum, then f^{NE} is also optimal for the maximum latency objective. For instance, if all latency functions are monomials of the same degree but with arc-dependent coefficients, it is well known that Nash equilibria and system optima coincide (Dafermos and Sparrow 1969). Notice that the upper bound given in Theorem 5.2 is tight as shown by the example in Figure 6, which goes back to Braess (1968). Indeed, the latency of the unique Nash equilibrium is $\ell(d)$ while the maximum latency of a min-max flow, which coincides with the system optimum, is

$$\ell(d) - \max_{0 \leq x \leq d} \left\{ \frac{x}{d} (\ell(d) - \ell(x)) \right\} .$$

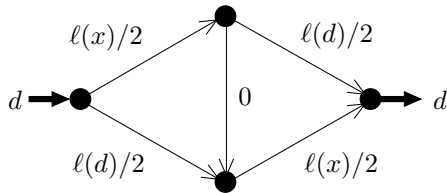


FIGURE 6. Instance showing that the bound in Theorem 5.2 is tight for single-commodity networks.

The claim follows by taking the supremum over $d \geq 0$ and $\ell \in \mathcal{L}$.

For instances with multiple sources and a single sink, the maximum latency of a Nash equilibrium is in general unbounded with respect to that of a min-max flow, even with linear latencies. In fact, we will show that the price of anarchy is $\Omega(n)$, where n is the number of nodes in the network. Weitz (2001) showed that the price of anarchy is unbounded in the case of two commodities, and Roughgarden (2004) proved that it is at most $n - 1$ if there is a common source and sink.

Proposition 5.3. *The price of anarchy in a single-commodity network with multiple sources and a single sink is $\Omega(n)$, even if all latencies are linear functions.*

Proof. Fix a constant $\varepsilon > 0$ and consider the instance presented in Figure 7. Nodes $n, n - 1, \dots, 1$ are the sources while node 0 is the sink. Nodes i and $i - 1$ are connected with two arcs: a_i with constant latency equal to 1 and \tilde{a}_i with latency equal to x/ε^i . Let the demand entering node $i > 0$ be ε^i . A Nash equilibrium of this instance routes the flow along paths of the form $\tilde{a}_i, a_{i-1}, \dots, a_1$ and has maximum latency n . (Note that all Nash equilibria have the same flow values on arcs in this instance; moreover, all path decompositions have the same maximum latency.) To show the claim, it suffices to exhibit a good solution. For instance, for origin i , let its demand flow along the path $a_i, \tilde{a}_{i-1}, \dots, \tilde{a}_1$. Under this flow, the load of \tilde{a}_i is equal to $\varepsilon^{i+1} + \dots + \varepsilon^n$ and its traversal time is $\varepsilon^1 + \dots + \varepsilon^{n-i}$. Hence, we can bound the maximum latency from above by $1 + n\varepsilon/(1 - \varepsilon)$, which tends to 1 when $\varepsilon \rightarrow 0$. \square

In contrast to the authors' earlier results on the total latency objective, the previous theorem implies that all Nash equilibria can be arbitrarily bad with respect to the maximum

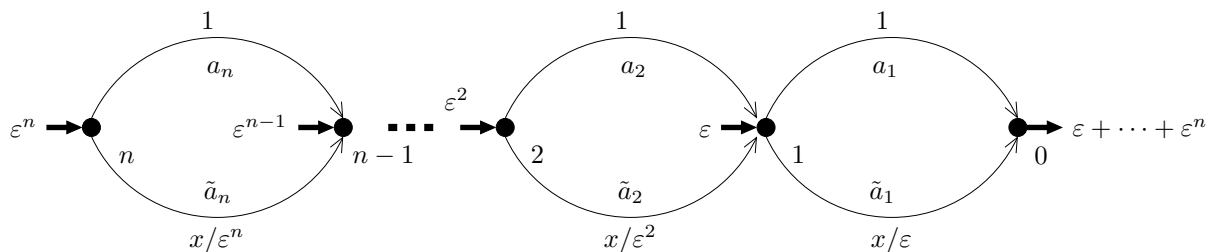


FIGURE 7. Instance showing that Nash equilibria can be arbitrarily bad for the maximum latency objective in networks with multiple sources and a single sink.

latency objective in single-source single-sink networks with capacities. A Nash equilibrium with capacities is a Nash equilibrium in the same instance without capacities, but where players experience infinite disutilities when their actions would result in infeasible solutions (Correa, Schulz, and Stier-Moses 2004b).

Corollary 5.4. *Consider single-source single-sink networks with explicit arc capacities. The worst-case ratio of the maximum latency of a best Nash equilibrium to that of a min-max flow is unbounded, even if all latencies are linear functions.*

Proof. We modify the instance from the proof of Proposition 5.3 by adding a super-source $n + 1$ and arcs $(n + 1, i)$, for $i = n, \dots, 1$. Arc $(n + 1, i)$ has capacity ε^i , and all nodes have supply zero except for $n + 1$, whose supply is $\varepsilon^n + \dots + \varepsilon$. The latencies of all arcs $(n + 1, i)$ are identical zero, for $i = n, \dots, 1$.

Because all feasible flows in the new instance saturate every arc $(n + 1, i)$ for $i = n, \dots, 1$, flows in the original and in the new instances are in one-to-one correspondence. In particular, the extension of a min-max flow is a min-max flow in the new instance, and the extension of a Nash equilibrium is a Nash equilibrium with capacities. The result follows because the best equilibrium with capacities of the new instance has the same maximum latency as an arbitrary equilibrium of the original instance. \square

In the single-source single-sink case, Nash equilibria are not the only good approximations to the maximum latency problem; an immediate corollary of Theorem 4.2 is that system optima are also close to optimality with respect to the maximum latency objective.

Theorem 5.5. *Let \mathcal{L} be a family of differentiable and nondecreasing latency functions. For single-source single-sink instances with latency functions drawn from \mathcal{L} , the maximum latency of a system optimum is bounded by $\gamma(\mathcal{L})$ times that of a min-max flow.*

Proof. The result follows just by using a min-max flow f^{MM} in Corollary 4.3. \square

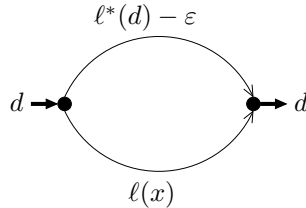


FIGURE 8. Instance showing that the bound in Theorem 5.5 is tight.

The bound given in Theorem 5.5 is best possible. To see this, consider the instance depicted in Figure 8. The min-max flow routes the entire demand along the lower arc, for a small enough $\varepsilon > 0$. On the other hand, the unique system optimum has to satisfy $\ell^*(x) = \ell^*(d) - \varepsilon$, where x is the flow along the lower arc. Therefore, the upper arc has positive flow and the maximum latency is $\ell^*(d) - \varepsilon$. The ratio between the maximum latencies of the two solutions is arbitrarily close to $\ell^*(d)/\ell(d)$. Taking the supremum over $d \geq 0$ and $\ell \in \mathcal{L}$ shows that the bound in Theorem 5.5 is tight.

To complete Table 1, let us prove that the average latency of a min-max flow is not too far from that of a system optimum.

Theorem 5.6. *Consider a set \mathcal{L} of continuous and nondecreasing latency functions. Let f^{MM} be a min-max flow and let f^{SO} be a system optimum for an instance with a single source, a single sink and latencies drawn from \mathcal{L} . Then, $C(f^{\text{MM}}) \leq \alpha(\mathcal{L})C(f^{\text{SO}})$.*

Proof. Note that $C(f^{\text{MM}}) \leq L(f^{\text{MM}}) \leq L(f^{\text{NE}}) = C(f^{\text{NE}}) \leq \alpha(\mathcal{L})C(f^{\text{SO}})$, where f^{NE} is a Nash equilibrium of the instance. \square

Again, the guarantee given in the previous theorem is tight. To show this, it is enough to note that the equilibrium flow and the min-max flow coincide in the example of Figure 9,

and their average latency is $\ell(d)$. Moreover, the average latency of the system optimum is arbitrary close to

$$\ell(d) - \max_{0 \leq x \leq d} \left\{ \frac{x}{d} (\ell(d) - \ell(x)) \right\} .$$

Taking the supremum of the ratio of these two values over $d \geq 0$ and $\ell \in \mathcal{L}$ completes the argument.

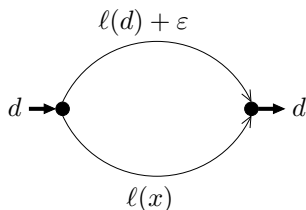


FIGURE 9. Instance showing that the bound in Theorem 5.6 is tight.

6. THE BOTTLENECK OBJECTIVE

In this section, we consider a fourth objective. It arises from a different interpretation of the problem considered by Koutsoupias and Papadimitriou (1999). As explained in the introduction, they considered the maximum latency in a network in which all paths consist of just a single arc. Instead of generalizing that objective to the maximum latency of a path, we could as well consider the problem of minimizing the maximum latency of the arcs. Given an arbitrary instance with multiple commodities, a *bottleneck flow* is a feasible flow f^{BN} that minimizes the maximum latency among those arcs with positive flow. This problem, referred to as the *bottleneck problem*, is of interest to telecommunication network service providers because they typically use routing schemes that minimize arc loads (see, e.g., Qiu et al. 2003). Indeed, providers seek to have spare capacity available, so, in the event of an arc failure, it can be used for rerouting traffic. Another application can be found in the operation of server farms for which a bottleneck optimum effectively balances the load among the servers (Czumaj, Krysta, and Vöcking 2002).

If the arc latency functions are convex, a bottleneck flow can be obtained by solving $|A|$ convex programs; thus, in contrast to a min-max flow, it can be approximated arbitrarily well in polynomial time. Let us outline an algorithm. For $a \in A$, let $A^a := \{a' \in A : \ell_{a'}(0) \leq$

$\ell_a(0)$ be the set of arcs that can be used in subproblem a , and $d_k(s_k) := d_k$, $d_k(t_k) := -d_k$, and $d_k(v) := 0$ for all $v \in V \setminus \{s_k, t_k\}$, where $k \in K$. Consider the following convex program.

$$Z_a = \min \quad z \quad (2a)$$

$$\text{s.t.} \quad \ell_{a'}(f_{a'}) \leq z \quad a' \in A^a, \quad (2b)$$

$$\sum_{k \in K} f_{a'}^k = f_{a'} \quad a' \in A^a, \quad (2c)$$

$$\sum_{a' \in \delta^+(v) \cap A^a} f_{a'}^k - \sum_{a' \in \delta^-(v) \cap A^a} f_{a'}^k = d_k(v) \quad v \in V, k \in K, \quad (2d)$$

$$f_{a'}^k \geq 0 \quad a' \in A^a, k \in K, \quad (2e)$$

where $\delta^+(v)$ (resp. $\delta^-(v)$) represents the outgoing (resp. incoming) arcs from (to) v , and the superindex k represents the commodity. When the convex program is infeasible we set the corresponding objective function value to ∞ .

Theorem 6.1. *The value of a bottleneck flow equals $Z = \min\{Z_a : a \in A\}$.*

Proof. Let f^{BN} be a bottleneck flow, and let $v^* := \max\{\ell_a(f_a^{\text{BN}}) : f_a^{\text{BN}} > 0 \text{ for } a \in A\}$ be its value. Moreover, let f^a be an optimal solution to the convex program corresponding to the value Z_a . Clearly, if $Z_a < \infty$ then f^a is a feasible flow in the original network. Therefore $Z \geq v^*$.

To see the other inequality, first note that if $v^* \geq \ell_{a_1}(0) := \max\{\ell_a(0) : a \in A\}$, then $v^* = Z_{a_1} \geq Z$. Assume now that $v^* < \ell_{a_1}(0)$, and let $\ell_{a_2}(0) := \max\{\ell_a(0) : a \in A \text{ and } \ell_a(0) \leq v^*\}$. Note that f^{BN} is not using any arc a with $\ell_a(0) > v^*$; hence, $v^* = Z_{a_2} \geq Z$. \square

Next we study the quality of bottleneck flows with respect to the maximum latency objective, the average latency objective, and their unfairness. We show that bottleneck flows can be arbitrarily far from optimal for each of these three objectives. In turn, min-max flows, system-optimal flows, and Nash flows can be of arbitrarily poor quality with respect to the bottleneck objective.

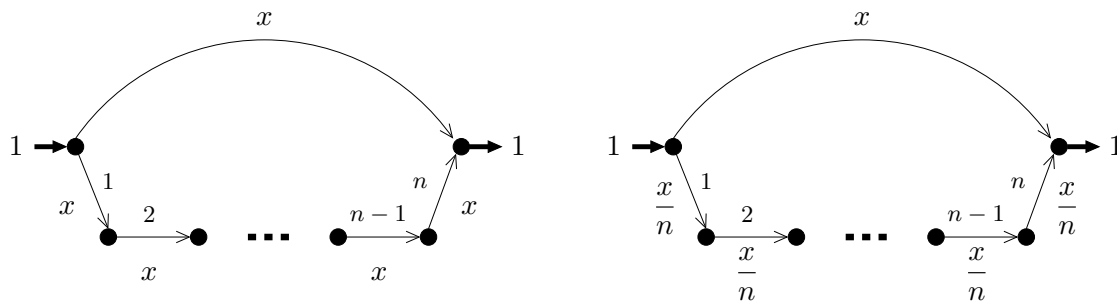


FIGURE 10. Examples for the bottleneck objective.

To see the first part, consider an instance with unit demand and two nodes connected with two paths as shown on the left-hand side of Figure 10. The first path consists of a single arc, and the second path is given by a chain of n arcs. The latency function associated with each arc is $\ell_a(x) = x$. The unique bottleneck flow has value $1/2$ on all arcs. The system optimum, the Nash equilibrium, and the min-max flow coincide (and are unique). They all route $\frac{1}{n+1}$ units of flow along the chain and $\frac{n}{n+1}$ units on the path consisting of the single arc.

On the other hand, consider the instance on the right side of Figure 10. Now, the only bottleneck flow routes $\frac{n}{n+1}$ units along the arcs on the chain and $\frac{1}{n+1}$ units on the path consisting of the single arc. The system optimum, the Nash equilibrium, and the min-max flow coincide again; they route $1/2$ units of flow on all arcs.

7. CONCLUSION

We have shown that computing a flow of minimum maximum latency is NP-hard, even in the single-source single-sink case with linear latency functions. Still, the problem admits a solution that is fair. We have proved tight worst-case bounds between the different solutions and with respect to different objectives. For instance, we have shown that two standard solutions to network problems give constant-factor approximations for the maximum latency problem. On the one hand, the maximum latency of any Nash equilibrium is within a factor of $\alpha(\mathcal{L})$ of that of a min-max flow, and Nash flows are fair. On the other hand, the ratio of the maximum latency of any system optimum to that of a min-max flow is at most $\gamma(\mathcal{L})$, and its unfairness is also bounded by $\gamma(\mathcal{L})$. Table 3 summarizes the findings for single-

TABLE 3. Overview of approximation guarantees for single-source single-sink networks when latencies belong to a given set \mathcal{L} . All bounds are tight. The “?” indicates that no upper bound is known; recall from the example depicted in Figure 4 that 2^p is a lower bound for polynomials of degree p , for $p \geq 2$.

	maximum latency	average latency	unfairness
min-max flow	1	$\alpha(\mathcal{L})$?
system optimum	$\gamma(\mathcal{L})$	1	$\gamma(\mathcal{L})$
Nash equilibrium	$\alpha(\mathcal{L})$	$\alpha(\mathcal{L})$	1

source single-sink networks with latencies drawn from a given class \mathcal{L} of allowable latency functions. We do not include the bottleneck objective and the bottleneck flow because we have shown that all interesting entries would be infinity. Correa, Schulz, and Stier-Moses (2005) extended our study of the maximum latency objective to nonatomic congestion games and gave pseudoapproximation results. Lin et al. (2005) showed that the maximum latency of an equilibrium of general multicommodity instances can be exponentially larger than that of an optimal solution. And Chakrabarty et al. (2005) presented results similar to ours for atomic congestion games and a different notion of fairness.

Acknowledgements. The authors are grateful to three anonymous referees whose constructive comments helped to improve the presentation of this paper.

REFERENCES

- Aronson, J. E. (1989). A survey of dynamic network flows. *Annals of Operations Research* 20, 1–66.
- Beccaria, G. and A. Bolelli (1992). Modelling and assessment of dynamic route guidance: The MARGOT project. In L. Olausson and E. Helli (Eds.), *Proceedings of the 3rd IEEE Vehicle Navigation & Information Systems Conference*, Oslo, Norway, pp. 117–126.
- Beckmann, M. J., C. B. McGuire, and C. B. Winsten (1956). *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.
- Braess, D. (1968). Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung* 12, 258–268. An English translation appeared in *Transportation Science* 39, 446–450, 2005.

- Chakrabarty, D. Mehta, V. Nagarajan, and V. Vazirani (2005). Fairness and optimality in congestion games. In J. Riedl, M. J. Kearns, and M. K. Reiter (Eds.), *Proceedings of the 6th ACM Conference on Electronic Commerce (EC)*, Vancouver, BC, Canada, pp. 52–57.
- Correa, J. R., A. S. Schulz, and N. E. Stier-Moses (2004a). Computational complexity, fairness, and the price of anarchy of the maximum latency problem. In D. Bienstock and G. Nemhauser (Eds.), *Proceedings of the 10th Conference on Integer Programming and Combinatorial Optimization (IPCO)*, New York, NY, Volume 3064 of *Lecture Notes in Computer Science*, pp. 59–73. Springer, Heidelberg.
- Correa, J. R., A. S. Schulz, and N. E. Stier-Moses (2004b). Selfish routing in capacitated networks. *Mathematics of Operations Research* 29(4), 961–976.
- Correa, J. R., A. S. Schulz, and N. E. Stier-Moses (2005). On the inefficiency of equilibria in congestion games. In M. Jünger and V. Kaibel (Eds.), *Proceedings of the 11th Conference on Integer Programming and Combinatorial Optimization (IPCO)*, Berlin, Germany, Volume 3509 of *Lecture Notes in Computer Science*, pp. 167–181. Springer, Heidelberg.
- Czumaj, A. (2004). Selfish routing on the Internet. In J. Leung (Ed.), *Handbook of Scheduling: Algorithms, Models, and Performance Analysis*, Volume 1 of *Chapman & Hall/CRC Computer & Information Science Series*. CRC Press, Boca Raton, FL.
- Czumaj, A., P. Krysta, and B. Vöcking (2002). Selfish traffic allocation for server farms. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, Montreal, Canada, pp. 287–296. ACM Press, New York, NY.
- Czumaj, A. and B. Vöcking (2004). Tight bounds for worst-case equilibria. *Journal of Algorithms*. To appear. A preliminary version appeared in SODA 2002.
- Dafermos, S. C. and F. T. Sparrow (1969). The traffic assignment problem for a general network. *Journal of Research of the U.S. National Bureau of Standards* 73B, 91–118.
- Fleischer, L. and M. Skutella (2003). Quickest flows over time. IBM Research Report RC22833 (W0306-175).

- Ford, L. R. and D. R. Fulkerson (1958). Constructing maximal dynamic flows from static flows. *Operations Research* 6, 419–433.
- Grötschel, M., L. Lovász, and A. Schrijver (1993). *Geometric Algorithms and Combinatorial Optimization*. Springer, Berlin.
- Hoppe, B. and É. Tardos (1994). Polynomial time algorithms for some evacuation problems. In *Proceedings of the 5th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Arlington, VA, pp. 433–441. SIAM, Philadelphia, PA.
- Jahn, O., R. H. Möhring, A. S. Schulz, and N. E. Stier-Moses (2005). System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations Research* 53(4), 600–616.
- Jarvis, J. J. and H. D. Ratliff (1982). Some equivalent objectives for dynamic network flow problems. *Management Science* 28(1), 106–109.
- Köhler, E. and M. Skutella (2005). Flows over time with load-dependent transit times. *SIAM Journal on Optimization* 15(4), 1185–1202.
- Koutsoupias, E., M. Mavronicolas, and P. Spirakis (2003). Approximate equilibria and ball fusion. *Theory of Computing Systems* 36(6), 683–693.
- Koutsoupias, E. and C. H. Papadimitriou (1999). Worst-case equilibria. In C. Meinel and S. Tison (Eds.), *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, Trier, Germany, Volume 1563 of *Lecture Notes in Computer Science*, pp. 404–413. Springer, Heidelberg.
- Lin, H., T. Roughgarden, E. Tardos, and A. Walkover (2005). Braess’s paradox, Fibonacci numbers, and exponential inapproximability. In L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi, and M. Yung (Eds.), *Automata, Languages and Programming: Proceedings of the 32nd International Colloquium (ICALP)*, Lisboa, Portugal, Volume 3580 of *Lecture Notes in Computer Science*. Springer, Heidelberg.
- Mavronicolas, M. and P. Spirakis (2001). The price of selfish routing. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC)*, Hersonissos, Greece, pp. 510–519. ACM Press, New York, NY.

- Papadimitriou, C. H. (2001). Algorithms, games, and the Internet. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC)*, Hersonissos, Greece, pp. 749–753. ACM Press, New York, NY.
- Potra, F. and Y. Ye (1993). A quadratically convergent polynomial algorithm for solving entropy optimization problems. *SIAM Journal on Optimization* 3(4), 843–860.
- Powell, W. B., P. Jaillet, and A. Odoni (1995). Stochastic and dynamic networks and routing. In M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser (Eds.), *Networks*, Volume 4 of *Handbook in Operations Research and Management Science*, pp. 141–295. Elsevier Science, Amsterdam.
- Qiu, L., Y. R. Yang, Y. Zhang, and S. Shenker (2003). On selfish routing in Internet-like environments. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM)*, Karlsruhe, Germany, pp. 151–162. ACM Press, New York, NY.
- Roughgarden, T. (2002). How unfair is optimal routing? In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, San Francisco, CA, pp. 203–204. SIAM, Philadelphia, PA.
- Roughgarden, T. (2003a). Personal communication.
- Roughgarden, T. (2003b). The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences* 67, 341–364.
- Roughgarden, T. (2004). The maximum latency of selfish routing. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New Orleans, LA, pp. 973–974. SIAM, Philadelphia, PA.
- Roughgarden, T. and É. Tardos (2002). How bad is selfish routing? *Journal of the ACM* 49, 236–259.
- Schrijver, A. (1998). *Theory of Linear and Integer Programming*. J. Wiley & Sons, New York.

- Schulz, A. S. and N. E. Stier-Moses (2003). On the performance of user equilibria in traffic networks. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Baltimore, MD, pp. 86–87. SIAM, Philadelphia, PA.
- Vavasis, S. A. (1991). *Nonlinear Optimization: Complexity Issues*. Oxford University Press, New York, NY.
- Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II, Vol. 1*, 325–378.
- Weitz, D. (2001). The price of anarchy. Unpublished manuscript.