Symmetry in Network Congestion Games: Pure Equilibria and Anarchy Cost*

Dimitris Fotakis[1], Spyros Kontogiannis[2,3], and Paul Spirakis[2]

[1] Dept. of Information and Communication Systems Engineering, University of the Aegean, 83200 Samos, Greece
fotakis@aegean.gr

[2] Research Academic Computer Technology Institute, 26500 Patras, Greece
{kontog, spirakis}@cti.gr

[3] Dept. of Computer Science, University of Ioannina, 45110 Ioannina, Greece

Abstract. We study computational and coordination efficiency issues of Nash equilibria in symmetric network congestion games. We first propose a simple and natural greedy method that computes a pure Nash equilibrium with respect to traffic congestion in a network. In this algorithm each user plays only once and allocates her traffic to a path selected via a shortest path computation. We then show that this algorithm works for series-parallel networks when users are identical or when users are of varying demands but have the same best response strategy for any initial network traffic. We also give constructions where the algorithm fails if either the above condition is violated (even for series-parallel networks) or the network is not series-parallel (even for identical users). Thus, we essentially indicate the limits of the applicability of this greedy approach.

We also study the price of anarchy for the objective of maximum latency. We prove that for any network of m uniformly related links and for identical users, the price of anarchy is $\Theta(\frac{\log m}{\log \log m})$.

1 Introduction

Network congestion games provide a sound model for selfish routing of unsplittable traffic and have recently been the subject of intensive research. The prevailing questions in recent work have to do with the performance degradation due to lack of users' coordination (e.g., [23,12,10,1,3]) and the efficient computation of pure Nash equilibria (e.g., [8,11,10]).

A natural greedy approach for computing a pure Nash equilibrium (PNE) is Greedy Best Response (GBR). Let us consider a dynamic setting with new users arriving in the network. The users play only once and irrevocably choose their strategy upon arrival. Each new user routes her traffic on the minimum delay path given the paths of the users currently in the network. Hopefully the existing

* This work was partially supported by the EU within the Future and Emerging Technologies Programme under contract IST-2001-33135 (CRESCCO) and within the 6th Framework Programme under contract 001907 (DELIS).

T. Erlebach and G. Persiano (Eds.): WAOA 2005, LNCS 3879, pp. 161–175, 2006.

users are not affected by the new one and the network configuration remains at a PNE without any defections taking place. This approach is not only intuitive and computationally efficient, but also resembles how things work in practice. A natural question is whether there are any interesting classes of networks for which Greedy Best Response maintains a PNE.

Greedy Best Response can be regarded as a generalization of Graham's LPT algorithm [13]. The restriction of GBR to parallel-link networks is known to maintain a PNE for arbitrary non-decreasing latency functions and weighted users arriving in non-increasing order of weights [17,9]. In this work, we prove that GBR maintains a PNE for symmetric congestion games in series-parallel networks. This result is extended to weighted congestion games with a certain notion of symmetry, namely that the users have the same best response strategies for any initial network traffic.

The second important research direction has to do with the inefficiency of Nash equilibria. The *coordination ratio* or *price of anarchy* was introduced in [16] for measuring the performance degradation due to lack of users' coordination in resource sharing. The price of anarchy is the worst-case ratio between the cost of a Nash equilibrium and the cost of an optimal solution. For network congestion games, there are two natural notions of cost for defining the price of anarchy: the *total* and the *maximum* latency. As the price of anarchy for non-atomic congestion games becomes well-understood (e.g., [23,21] for total latency and [22,4] for maximum latency), the interest moves to the atomic setting (e.g. [18,12,10,1,3].) In both settings, the case of linear latencies is prominent and has been the focus of most of the previous work.

In this paper, we study the price of anarchy relative to the objective of maximum latency for symmetric network congestion games and latency functions $d_e(x) = a_e x, a_e \geq 0$. This corresponds to uniformly related links, with the coefficient a_e denoting the inverse speed of link e. We show that the price of anarchy for any network of m links is $\Theta(\frac{\log m}{\log\log m})$.

Related Work. Rosenthal [20] initiated the study of congestion games and proved that their PNE correspond to the local optima of a natural potential function. Therefore, the best response dynamics converges to a PNE. On the other hand, it is PLS-complete to find a PNE in symmetric (not necessarily network) and non-symmetric network congestion games [8]. On the positive side, [8] shows that in symmetric *network* congestion games, a PNE can be found by a min-cost flow computation. For weighted congestion games, [11] considers the case of identical parallel links and restricted assignments and shows how to compute a PNE in strongly-polynomial time. [10] shows that weighted congestion games with linear latencies admit a weighted potential function. Thus, the best response dynamics converges to a PNE in pseudo-polynomial time.

In a seminal paper, Koutsoupias and Papadimitriou [16] introduce the price of anarchy and consider the objective of maximum latency for a weighted congestion game on m uniformly related parallel links. The price of anarchy for that game is $\Theta(\frac{\log m}{\log\log m})$ if either the users or the links are identical [19,15,5] and $\Theta(\frac{\log m}{\log\log\log m})$ otherwise [5]. For uniformly related parallel links, identical users, and the objective of total latency, the price of anarchy is $2 - o(1)$ for the general case of mixed equilibria and $4/3$ for pure equilibria [18,12].

Similar results have been obtained recently for network congestion games with linear latency functions. The price of anarchy for the objective of total latency is $\frac{3+\sqrt{5}}{2}$ if weighted congestion games and mixed equilibria are considered [1]. This drops to $5/2$ for the special case of identical users and pure equilibria ([1] and independently in [3]). The price of anarchy for maximum latency is also $5/2$ for pure Nash equilibria and symmetric games (with identical users) and becomes $\Theta(\sqrt{n})$ for non-symmetric games [3].

On the other hand, the price of anarchy for m identical links and the objective of maximum latency is $\Omega(\frac{\log m}{\log \log m})$ if mixed Nash equilibria are considered [16,19]. [10] studies weighted single-commodity congestion games in layered networks with m identical links and shows that the price of anarchy for maximum latency remains $\Theta(\frac{\log m}{\log \log m})$ for the general case of mixed Nash equilibria.

The bounds above apply to the *atomic* setting, with users controlling a non-negligible amount of traffic demand, and consider both pure and mixed equilibria (with the exception of the results in [3] on maximum latency). Improved bounds can be obtained in the *non-atomic* setting, where each user controls a negligible amount of demand and pure and mixed equilibria are equivalent. [23] initiates the study of the price of anarchy for the objective of total latency in the non-atomic setting and shows that the price of anarchy for linear latencies is $4/3$. [21] proves that the price of anarchy depends on the class of latency functions and not on the network topology and gives a tight bound for every class.

As for the objective of maximum latency in the non-atomic setting, the upper bounds for total latency also apply to maximum latency in single-commodity networks [4]. For multi-commodity networks, the price of anarchy is $\Omega(|V|)$ even for linear latencies [4]. On the other hand, the price of anarchy for maximum latency is at most $|V| - 1$ in single-commodity networks [22].

Contribution. If the users are identical, GBR behaves as an online algorithm. For weighted users, GBR is the most natural greedy algorithm since it determines a fixed order in which the users are considered and each user makes an irrevocable greedy choice given the choices of the previous users.

In this paper, we essentially characterize the class of network congestion games for which GBR maintains a PNE. More specifically, we prove that GBR maintains a PNE for symmetric congestion games in series-parallel networks. This is extended to weighted congestion games with the *common best response* property. This property requires that the users have the same best response strategies for any initial network traffic. In addition to symmetric network congestion games, this class includes weighted congestion games in layered networks with identical edges (i.e., edge delays are given by a common linear latency function). We also prove that the restriction to series-parallel networks and games with the common best response property is essentially necessary for GBR to maintain a PNE.

For the price of anarchy, we focus on the objective of maximum latency. We consider symmetric network congestion games and linear latencies with no additive term, thus extending to arbitrary networks the widely-studied setting of identical users and uniformly related parallel links (e.g., [16,9,19,5]). We consider the general case of mixed equilibria and show that the price of anarchy remains $\Theta(\frac{\log m}{\log\log m})$ for identical users and networks of m links. The setting of identical users and arbitrary networks is orthogonal to the setting of [10] where the network has a notion of symmetry, namely all paths have the same length and consist of identical edges, and the users have different weights.

The results on the price of anarchy were obtained independently of results of [1,3]. Our approach is fundamentally different and may be of independent interest. It is based on a natural correspondence between mixed strategies and fractional $s-t$ flows (see also [10]). To motivate the approach, we first show that the optimal solution of a quadratic program corresponds to a symmetric mixed Nash equilibrium. We use quadratic programming duality and show that the expected cost of any user in a (pure or mixed) Nash equilibrium is at most 3 times the optimal maximum latency. A Chernoff-Hoeffding bound yields that the expected maximum latency is $O(\frac{\log m}{\log\log m})$ times the optimal maximum latency.

2 Definitions and Preliminaries

The Model. A *network congestion game* is a tuple $(N, G, (d_e)_{e\in E})$, where $N = \{1,\dots,n\}$ is the set of users controlling a unit of traffic demand each, $G(V,E)$ is a directed graph representing the communication network, and d_e is the latency function associated with edge $e \in E$. We assume that d_e's are non-negative and non-decreasing functions of the edge loads. If the edge delays are given by a *common linear* latency function, we say that the edges are *identical*. For identical edges, we assume wlog. that the edge delays are given by the identity function, i.e. $\forall e \in E, d_e(x) = x$. We restrict our attention to *single-commodity* network congestion games, where the network G has a single source s and destination t and the set of users' strategies is the set of $s-t$ paths, denoted $\mathcal{P}$. Wlog. we assume that G is connected and every vertex of G lies on a directed $s-t$ path.

We also consider *weighted* single-commodity network congestion games, where user i controls w_i units of traffic demand[1]. The users are indexed in non-increasing order of weights, i.e., $w_1 \geq w_2 \geq \dots \geq w_n$. Single-commodity network congestion games are *symmetric*[2]. However, weighted games are non-symmetric in general because the users' cost functions are different and non-symmetric due to different user weights.

A vector $P = (p_1,\dots,p_n)$ consisting of an $s-t$ path p_i for each user i is a *pure strategies profile.* Let $\ell_e(P) \equiv \sum_{i:e\in p_i} w_i$ denote the load of edge e in P. The cost $\lambda^i_p(P)$ of user i for routing her demand on path p in the profile P is

$$\lambda^i_p(P) \equiv \textstyle\sum_{e\in p\cap p_i} d_e(\ell_e(P)) + \sum_{e\in p\setminus p_i} d_e(\ell_e(P) + w_i)$$

The cost $\lambda^i(P)$ of user i in P is $\lambda^i_{p_i}(P)$, namely the total delay along her path.

[1] In (unweighted) congestion games, $w_1 = w_2 = \dots = w_n = 1$.

[2] A game is *symmetric* if all users have the same strategy set and the users' costs are given by identical symmetric functions of other users' strategies. In congestion games, the users are identical and a common strategy set implies symmetry.

A vector $Q = (q_1, \ldots, q_n)$ consisting of a probability distribution q_i over $\mathcal{P}$ for each user i is a *mixed strategies profile*. For each path p, $q_i(p)$ denotes the probability that user i routes her demand on p. Let $\ell_p(Q) \equiv \sum_{j=1}^{n} q_j(p)w_j$ be the expected load routed on path p in Q, and let $\ell_p(Q^{-i}) \equiv \ell_p(Q) - q_i(p)w_i$ be the expected load on p excluding the contribution of user i. Similarly, let $\ell_e(Q) \equiv \sum_{p:e\in p} \ell_p(Q)$ and $\ell_e(Q^{-i}) \equiv \sum_{p:e\in p} \ell_p(Q^{-i})$ be the expected load on edge e with and without user i respectively. The cost $\lambda_p^i(Q)$ of user i for routing her demand on path p in the mixed strategies profile Q is the expectation according to Q^{-i} of $\lambda_p^i(P^{-i} \oplus p)$ over all pure strategies profiles[3] P^{-i}. The cost $\lambda^i(Q)$ of user i in Q is the expectation according to Q of $\lambda^i(P)$ over all pure strategies profiles P.

For a strategies profile Q, let $\lambda^{\max}(Q) \equiv \max_{i\in N}\{\lambda^i(Q)\}$ be the maximum user cost in Q.

In this paper, we consider mixed strategies profiles only for identical users and linear latency functions $d_e(x) = a_e x$. Then, simply $\lambda_p^i(Q) \equiv \sum_{e\in p} a_e(\ell_e(Q^{-i})+1)$ and $\lambda^i(Q) \equiv \sum_{p\in\mathcal{P}} q_i(p)\lambda_p^i(Q)$ by linearity of expectation.

A mixed (in general) strategies profile Q is a *Nash equilibrium* if for every user i and every $p, p' \in \mathcal{P}$ with $q_i(p) > 0$, $\lambda_p^i(Q) \le \lambda_{p'}^i(Q)$. Therefore, if Q is a Nash equilibrium, $\lambda_p^i(Q) = \lambda_{p'}^i(Q) = \lambda^i(Q)$ for every user i and every $p, p' \in \mathcal{P}$ with both $q_i(p), q_i(p') > 0$.

We evaluate strategies profiles using the objective of *maximum latency*. The maximum latency $L(P)$ of a pure strategies profile P is the maximum user cost in P, $L(P) \equiv \lambda^{\max}(P)$. The maximum latency $L(Q)$ of a mixed strategies profile Q is the expectation according to Q of $L(P)$ over all pure strategies profiles P, $L(Q) \equiv \sum_{P\in\mathcal{P}^n} \mathbb{P}(P,Q)L(P)$, where $\mathbb{P}(P,Q) = \prod_{i=1}^{n} q_i(p_i)$ is the occurrence probability of P in Q. The optimal solution, denoted P^*, corresponds to a pure strategies profile and the optimal cost is $L(P^*)$. The price of anarchy is defined as worst-case ratio $L(Q)/L(P^*)$ over all Nash equilibria Q.

Flows. A feasible flow is a function $f : \mathcal{P} \mapsto \mathbb{R}_{\ge 0}$ such that $\sum_{p\in\mathcal{P}} f_p = \sum_{i=1}^{n} w_i$. We also use f to denote the $|\mathcal{P}|$-dimensional vector corresponding to the flow f. A flow is *unsplittable* if each user's demand is routed on a single path and *splittable* otherwise. Let $f_e \equiv \sum_{p:e\in p} f_p$ denote the flow on edge e.

Greedy Best Response. GBR considers the users one-by-one in non-increasing order of weight. Each user adopts her best response strategy given the strategies of previous users. The choice is irrevocable since no user can change her strategy in the future. In simple words, each user plays only once and selects its best response strategy at the moment she is considered by the algorithm.

Formally, let p_i be the path of user i, and let $P^i = (p_1, \ldots, p_i)$ be the pure strategies profile for users $1, \ldots, i$. Then, the path p_{i+1} of user $i+1$ is

$$p_{i+1} = \arg\min_{p\in\mathcal{P}}\{\textstyle\sum_{e\in p} d_e(\ell_e(P^i) + w_{i+1})\} \qquad (1)$$

We say that GBR *succeeds* if every profile P^i is a Nash equilibrium.

[3] For a n-dimensional vector X, $X^{-i} \equiv (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ and $X^{-i} \oplus x \equiv (x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)$.

Common Best Response. The single-commodity network congestion game $((w_i)_{i\in N}, G, (d_e)_{e\in E})$ has the *common best response* property if for every initial flow f (not necessarily feasible), all users have the same set of best response strategies wrt the edge loads induced by f. In other words, if a path p is a best response wrt f for some user, then the following inequality holds for all users j and all paths p':

$$\sum_{e\in p'} d_e(f_e + w_j) \geq \sum_{e\in p} d_e(f_e + w_j)$$

Furthermore, every segment π of a best response path p is a best response for routing the demand of any user between π's endpoints. We should highlight that in the definition above, best responses are computed without taking into account that some users may already contribute to the initial flow f. The common best response property requires a notion of symmetry between the users, namely that all of them have the same topmost preferences for any initial traffic conditions. This notion of symmetry is weaker than that of a symmetric game but still strong enough to make GBR work in series-parallel networks.

Layered and Series-Parallel Graphs. A directed (multi)graph $G(V,E)$ with a distinguished source s and destination t is *layered* if all directed $s-t$ paths have exactly the same length and each vertex lies on a directed $s-t$ path. A multigraph is *series-parallel with terminals* (s,t) if it is either a single edge (s,t) or can be obtained from two series-parallel graphs with terminals (s_1,t_1) and (s_2,t_2) connected either in *series* or in *parallel*. In a series connection, t_1 is identified with s_2, s_1 becomes s, and t_2 becomes t. In a parallel connection, s_1 is identified with s_2 and becomes s, and t_1 is identified with t_2 and becomes t. A directed graph with terminals (s,t) is series-parallel if and only if it does not contain a θ-graph with degree-2 terminals as a topological minor (Fig. 1.b) [7].

Proposition 1. *Let $G(V,E)$ be a series-parallel graph with terminals (s,t), and let vertices u,v connected by two disjoint paths, denoted π and π', only sharing their endpoints. Every $s-t$ path having at least one edge in common with π' contains both u and v.*

3 Greedy Best Response in Series-Parallel Networks

We first show that GBR succeeds if the network is series-parallel and the game has the common best response property.

Theorem 1. *If G is a series-parallel graph with terminals (s,t) and the game $((w_i)_{i\in N}, G, (d_e)_{e\in E})$ has the common best response property,* GBR *succeeds and computes a pure Nash equilibrium in time $O(nm\log m)$.*

Proof. The proof is by induction on the number of users considered by the algorithm. The claim holds for the first user, since she adopts her best response strategy and is the only user in the network. We inductively assume that after user i has been considered, $P^i = (p_1,\ldots,p_i)$ is a Nash equilibrium. Let p_{i+1}

be the path chosen by user $i+1$ according to (1). To reach a contradiction, we assume that $P^{i+1} = (p_1, \ldots, p_i, p_{i+1})$ is not a Nash equilibrium.

Consequently, there is a user j, $j \leq i$, preferring another path p to her path p_j. Let u be a *split* point where p departs from p_j (u may be s). Any pair of different paths has at least one split point because they have a common source. Let v be the first *merge* point after u where p joins p_j again (v may be t). Each split point is followed by a merge point because the paths have a common destination.

For simplicity of notation, let π and π_j denote the segments of p and p_j respectively between u and v. By the definition of v, π and π_j are edge disjoint and have only their endpoints u and v in common.

Since j wants to defect from π_j in P^{i+1} but not in P^i, it is p_{i+1} that shares some edges with π_j and makes it inferior to π for user j. Since p_{i+1} and π_j have at least one edge in common, p_{i+1} contains both u and v by Proposition 1. Let π_{i+1} be the segment of p_{i+1} between u and v (Fig. 1.a).

The path p_{i+1} is a best response for user $i+1$ wrt the flow induced by P^i. Since the game has the common best response property, p_{i+1} is also a best response for user j wrt the flow induced by P^i (ignoring that w_j already contributes to the flow). Therefore, the path segment π_{i+1} is a best response wrt P^i for routing the demand of user j from u to v:

$$\sum_{e \in \pi} d_e(\ell_e(P^i) + w_j) \geq \sum_{e \in \pi_{i+1}} d_e(\ell_e(P^i) + w_j) \tag{2}$$

Since j prefers π to π_j after user $i+1$ routing her traffic on π_{i+1},

$$\begin{aligned} \sum_{e \in \pi_j \setminus \pi_{i+1}} d_e(\ell_e(P^i)) + \sum_{e \in \pi_j \cap \pi_{i+1}} d_e(\ell_e(P^i) + w_{i+1}) > \\ \sum_{e \in \pi} d_e(\ell_e(P^i) + w_j) \geq \sum_{e \in \pi_{i+1}} d_e(\ell_e(P^i) + w_j) \geq \\ \sum_{e \in \pi_{i+1} \setminus \pi_j} d_e(\ell_e(P^i) + w_j) + \sum_{e \in \pi_{i+1} \cap \pi_j} d_e(\ell_e(P^i) + w_{i+1}) \end{aligned}$$

The second inequality follows from Ineq. (2). The last inequality holds because the latency functions are non-decreasing and $w_j \geq w_{i+1}$.

If $\pi_j = \pi_{i+1}$, the contradiction is immediate. If $\pi_j \neq \pi_{i+1}$, user j prefers the path segment $\pi_{i+1} \setminus \pi_j$ to the path segment $\pi_j \setminus \pi_{i+1}$ even in P^i:

$$\lambda^j_{\pi_j \setminus \pi_{i+1}}(P^i) = \sum_{e \in \pi_j \setminus \pi_{i+1}} d_e(\ell_e(P^i)) > \sum_{e \in \pi_{i+1} \setminus \pi_j} d_e(\ell_e(P^i) + w_j) = \lambda^j_{\pi_{i+1} \setminus \pi_j}(P^i)$$

This contradicts to the inductive hypothesis that P^i is a Nash equilibrium. Therefore, p and p_j does not have any split points and p coincides with p_j. Consequently, P^{i+1} is a Nash equilibrium.

GBR performs n $s-t$ shortest path computations in a graph of m edges. This can be done in time $O(nm \log m)$ using Dijkstra's algorithm. □

Single-commodity network congestion games with identical users have the common best response property because the users' cost functions are identical functions of the edge loads. We are also aware of a class of weighted single-commodity network congestion games with the common best response property.

Proposition 2. *A weighted single-commodity congestion game in a layered network with identical edges has the common best response property for any set of user weights.*

Corollary 1. GBR *succeeds for single-commodity congestion games in series-parallel networks:*

1. *if the users are* identical *(for arbitrary non-decreasing edge delays).*
2. *if the graph is* layered *and the edges are* identical *(for arbitrary user weights).*

GBR has a natural distributed implementation based on a leader election algorithm. There is a process corresponding to each player. We assume that the processes know the network and the edge latency functions. We also assume a message passing subsystem and an underlying synchronization mechanism (e.g. logical timestamps) allowing a distributed algorithm to proceed in logical rounds.

Initially, all processes are active. In each round, they run a leader election algorithm and determine the active process of largest weight. This process routes its demand on its best response path, announces its strategy to the remaining active processes, and becomes passive. Notice that all processes can compute their best responses locally. In the offline setting, the algorithm terminates as soon as there are no active processes. In the online setting, new users/processes may enter the system at any point in time.

We conclude the study of GBR by providing some simple examples demonstrating that GBR may not succeed in maintaining a PNE if either the network is not series-parallel or the game does not have the common best response property. Hence both conditions of Theorem 1 are necessary for GBR to succeed.

If the network is not series-parallel, the simplest *symmetric game* for which GBR fails consists of two identical users and the 3-layered equivalent of the θ-graph with identical edges (Fig. 1.b). The pure Nash equilibrium assigns one user to π_1 and the other to π_3. If GBR assigns the first user to π_2, there is no strategy for the second user that yields a Nash equilibrium. We can force GBR to assign the first user to π_2 by slightly decreasing the latency function of the second edge to $(1-\epsilon)x$, where ϵ is a small positive constant.

The common best response property is also necessary for series-parallel networks other than a sequence of parallel-link graphs connected in series[4]. For example, let us consider the 2-layered series-parallel graph of Fig. 1.c and three users of weights $w_1 = 100$, $w_2 = 10$, and $w_3 = 4$. The corresponding congestion game does not have the common best response property. GBR assigns the first user to the path π_1, the second user to π_2, and the third user to π_3, while in every pure Nash equilibrium the first two users are assigned to π_1.

[4] If the network consists of bunches of parallel-link connected in series, a pure Nash equilibrium can be computed by independently applying GBR to each bunch of parallel links.

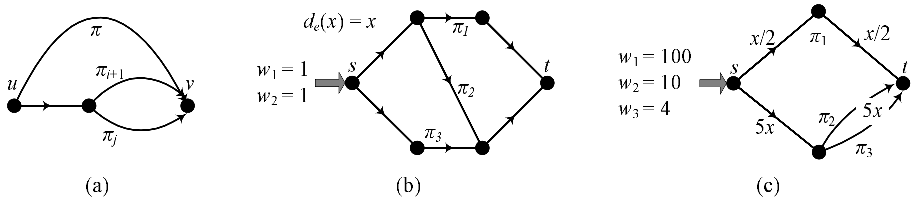

Fig. 1. (a) The graph in the proof of Theorem 1. GBR may fail if (b) the network is not series-parallel (even if the game is symmetric) and (c) the game does not have the common best response property (even if the network is series-parallel).

4 The Price of Anarchy in Networks of Uniformly Related Links

We proceed to bound the price of anarchy in symmetric network congestion games with uniformly related links. We consider n identical users routing their (unit) traffic demands on a directed graph $G(V, E)$ with a unique source s and destination t, and $m \equiv |E|$ edges. There is a linear latency function $d_e(x) = a_e x$, $a_e \geq 0$, associated with each edge e. We regard a_e as the inverse speed of edge e. For each path $p \in \mathcal{P}$, let $a_p \equiv \sum_{e \in p} a_e$ denote the inverse speed of p.

Flows and Mixed Strategies Profiles. A feasible flow is a function $f : \mathcal{P} \mapsto \mathbb{R}_{\geq 0}$ such that $\sum_{p \in \mathcal{P}} f_p = n$. Let $\theta_p(f) \equiv \sum_{e \in p} a_e f_e$ denote the total delay along the path p wrt f. We map a mixed (in general) strategies profile $Q = (q_1, \ldots, q_n)$ to a feasible flow f^Q as follows: For each $s - t$ path $p \in \mathcal{P}$, $f_p^Q \equiv \ell_p(Q)$. In other words, we handle the expected load routed on p in Q as a splittable flow, where user i routes a fraction $q_i(p)$ of her demand on p. If Q is a pure strategies profile, the corresponding flow is unsplittable.

We say that a feasible flow f^Q corresponding to a strategies profile Q is at Nash equilibrium with the understanding that actually Q is a Nash equilibrium. For every Nash equilibrium Q and the corresponding flow f^Q,

$$\lambda^{\max}(Q) \leq \min_{p \in \mathcal{P}} \{\theta_p(f^Q) + a_p\} \equiv \delta^{\min}(f^Q) \tag{3}$$

Otherwise, a user of cost $\lambda^{\max}(Q)$ in Q could improve her cost by switching to the path minimizing $\theta_p(f^Q) + a_p$. Furthermore, for any path $p \in \mathcal{P}$ with $f_p^Q > 0$,

$$\max\{\theta_p(f^Q), a_p\} \leq \lambda^{\max}(Q) \leq \delta^{\min}(f^Q) \tag{4}$$

For simplicity, we drop the superscript of Q from its corresponding flow f^Q when the strategies profile is clear from the context.

Total Latency and Total Load. A flow f can be evaluated by its *total latency* defined as

$$C(f) \equiv \sum_{p \in \mathcal{P}} f_p \theta_p(f) = \sum_{e \in E} a_e f_e^2$$

In addition, a flow f can be evaluated by its *total load* defined as

$$W(f) \equiv \sum_{e \in E} a_e f_e = \sum_{e \in E} a_e \sum_{p: e \in p} f_p = \sum_{p \in \mathcal{P}} a_p f_p$$

We sometimes use $W(P^*)$ to denote the total load of the flow corresponding to the optimal solution P^*.

Proposition 3. *Let f be a feasible flow at Nash equilibrium. Then $C(f) \leq n\, \delta^{\min}(f)$.*

Proof. By Ineq. (4), for every path $p \in \mathcal{P}$, $f_p\, \theta_p(f) \leq f_p\, \delta^{\min}(f)$. Summing over all paths, we conclude that $C(f) \leq n\, \delta^{\min}(f)$. □

Let M be the $|\mathcal{P}| \times |\mathcal{P}|$ square matrix defined as $M[p, p'] \equiv \sum_{e \in p \cap p'} a_e$ for each pair of paths $p, p' \in \mathcal{P}$. By definition, M is a symmetric matrix. For every flow f, Mf is the $|\mathcal{P}|$-dimensional vector with coordinates $\theta_p(f)$. Thus, the total latency of f can be expressed as $C(f) = f^T M f$. Since $C(f) = f^T M f = \sum_{e \in E} a_e f_e^2$ the matrix M is positive semi-definite[5].

Let A be the $|\mathcal{P}|$-dimensional vector with $A[p] \equiv a_p$ for each path $p \in \mathcal{P}$. The total load of every flow f can be expressed as $W(f) = A^T f$.

The *maximum latency* of an unsplittable flow f is $L(f) \equiv \max_{p: f_p > 0}\{\theta_p(f)\}$. Notice that for every pure strategies profile P and its corresponding unsplittable flow f^P, $L(P) = L(f^P)$.

4.1 Computing a Symmetric Nash Equilibrium

We next prove that the flow minimizing $\frac{n-1}{2n} C(f) + W(f)$ corresponds to a symmetric Nash equilibrium. Formally, let $\hat{f}$ be the optimal fractional solution to the following quadratic program $\min\{\frac{n-1}{2n} f^T M f + A^T f : \mathbf{1}^T f \geq n, f \geq \mathbf{0}\}$, where $\mathbf{1}$ (resp. $\mathbf{0}$) denotes the $|\mathcal{P}|$-dimensional vector with 1 (resp. 0) in each coordinate. We observe that $\hat{f}$ is a splittable flow of value n.

Lemma 1. *Let Q be the mixed strategies profile where each user i routes its demand on every path p with probability $q_i(p) = \hat{f}(p)/n$. Then, Q is a symmetric Nash equilibrium.*

Proof. The mixed strategies profile Q is symmetric by definition. We only have to show that Q is a Nash equilibrium. By construction, for every user i and every path p, $\ell_p(Q^{-i}) = \frac{n-1}{n} \hat{f}_p$. Therefore, for every user i and every edge e, $\ell_e(Q^{-i}) = \frac{n-1}{n} \hat{f}_e$. Thus, the cost of a user i routing her demand on a path p in the mixed strategies profile Q is

$$\lambda_p^i(Q) = \sum_{e \in p} a_e(\ell_e(Q^{-i}) + 1) = \sum_{e \in p} a_e(\tfrac{n-1}{n} \hat{f}_e + 1) = \tfrac{n-1}{n} \theta_p(\hat{f}) + a_p$$

[5] A $n \times n$ matrix M is positive semi-definite if for every vector $x \in \mathbb{R}^n$, $x^T M x \geq 0$.

The flow $\hat{f}$ minimizes the convex function $\sum_{e\in E}(\frac{n-1}{2n}a_e f_e^2 + a_e f_e)$. Therefore, for every $p, p' \in \mathcal{P}$ with $\hat{f}_p > 0$, the following inequality holds (e.g., [2], [23, Lemma 2.5]):

$$\tfrac{n-1}{n}\theta_p(\hat{f}) + a_p = \sum_{e\in p}(\tfrac{n-1}{n}a_e\hat{f}_e + a_e) \le \sum_{e\in p'}(\tfrac{n-1}{n}a_e\hat{f}_e + a_e) = \tfrac{n-1}{n}\theta_{p'}(\hat{f}) + a_{p'}$$

Consequently, for every user i and every $p, p' \in \mathcal{P}$ with $q_i(p) = \hat{f}_p/n > 0$,

$$\lambda_p^i(Q) = \tfrac{n-1}{n}\theta_p(\hat{f}) + a_p \le \tfrac{n-1}{n}\theta_{p'}(\hat{f}) + a_{p'} = \lambda_{p'}^i(Q)$$

and the mixed strategies profile Q is a Nash equilibrium. □

Remark. If the network consists of m uniformly related parallel links, the equilibrium of Lemma 1 is identical to the *generalized fully mixed* Nash equilibrium of [9, Theorem 5].

4.2 Bounding the Price of Anarchy

We first apply the Chernoff-Hoeffding bound and prove that for every Nash equilibrium Q with $\lambda^{\max}(Q) \le \alpha\, L(P^*)$ for some constant $\alpha \ge 1$, $L(Q) = \alpha\, O(\frac{\log m}{\log\log m})\, L(P^*)$ (Lemma 2). We then prove that for every Nash equilibrium Q, $\lambda^{\max}(Q) \le L(P^*) + \frac{2}{n}\, W(P^*)$ (Theorem 2). The proof is based on Dorn's Theorem [6] establishing strong duality in quadratic programming. As an immediate consequence, we obtain that for every Nash equilibrium Q, $\lambda^{\max}(Q) \le 3\, L(P^*)$ (Corollary 2). The results in this section can be extended to symmetric (not necessarily network) congestion games with identical users, resource set E, strategy set $\mathcal{P}$, and resource costs $d_e(x) = a_e x$, $a_e \ge 0$.

Lemma 2. *Let Q be any strategies profile at Nash equilibrium. If there exists some constant $\alpha \ge 1$ such that $\lambda^{\max}(Q) \le \alpha\, L(P^*)$, $L(Q) \le \alpha\, O(\frac{\log m}{\log\log m})\, L(P^*)$.*

Proof. For every edge e and every user i, let $X_{e,i}$ be the random variable describing the actual load routed on e by i. The random variable $X_{e,i}$ is 1 if i routes its demand on a path containing e and 0 otherwise. The expectation of $X_{e,i}$ is $\mathbb{E}[X_{e,i}] = \sum_{p: e\in p} q_i(p)$. Since the users select their paths independently, for every edge e, the random variables $\{X_{e,i},\ i \in N\}$ are mutually independent.

For each edge e, let $X_e = a_e \sum_{i=1}^n X_{e,i}$ be the random variable that describes the actual delay incurred by any user traversing e. Multiplying each $X_{e,i}$ by a_e, we can regard X_e as the sum of n independent random variables with values in $\{0, a_e\}$. By linearity of expectation,

$$\mathbb{E}[X_e] = a_e \sum_{i=1}^n \mathbb{E}[X_{e,i}] = a_e \sum_{p: e\in p}\sum_{i=1}^n q_i(p) = a_e \sum_{p: e\in p} \ell_p(Q) = a_e \ell_e(Q)$$

The Hoeffding bound[6] for $w = a_e$ and $t = \mathrm{e}\,\kappa\, a_e \max\{\ell_e(Q), 1\}$, yields that for every $\kappa \geq 1$,

$$\mathbb{P}[X_e \geq \mathrm{e}\,\kappa\, a_e \max\{\ell_e(Q), 1\}] \leq \kappa^{-\mathrm{e}\,\kappa}$$

Applying the union bound, we conclude that

$$\mathbb{P}[\exists e \in E : X_e \geq \mathrm{e}\,\kappa\, a_e \max\{\ell_e(Q), 1\}] \leq m\kappa^{-\mathrm{e}\,\kappa} \tag{5}$$

For every path $p \in \mathcal{P}$ with $\ell_p(Q) > 0$, we define the random variable $X_p = \sum_{e\in p} X_e$ describing the actual delay along p. The maximum latency of Q cannot exceed the expected maximum delay among paths p with $\ell_p(Q) > 0$. Formally,

$$L(Q) \leq \mathbb{E}[\max_{p:\ell_p(Q)>0}\{X_p\}]$$

Let us assume that for all edges $e \in E$, $X_e < \mathrm{e}\,\kappa\, a_e \max\{\ell_e(Q), 1\}$. Let p be any path with $\ell_p(Q) > 0$, and let i be any user with $q_i(p) > 0$. Then,

$$\begin{aligned} X_p = \textstyle\sum_{e\in p} X_e &< \mathrm{e}\,\kappa \textstyle\sum_{e\in p} a_e \max\{\ell_e(Q), 1\} \leq \mathrm{e}\,\kappa \textstyle\sum_{e\in p} a_e(\ell_e(Q^{-i}) + 1) \\ &= \mathrm{e}\,\kappa\, \lambda^i(Q) \leq \mathrm{e}\,\kappa\, \lambda^{\max}(Q) \leq \mathrm{e}\,\kappa\,\alpha\, L(P^*) \end{aligned}$$

The second inequality follows from $\max\{\ell_e(Q), 1\} \leq \ell_e(Q^{-i}) + 1$ which holds for every edge $e \in p$ and every user i with $q_i(p) > 0$. Since $q_i(p) > 0$ and Q is a Nash equilibrium, $\lambda^i(Q) = \sum_{e\in p} a_e(\ell_e(Q^{-i}) + 1)$ and the next equality follows. The third inequality follows from the definition of $\lambda^{\max}(Q)$ and the last inequality by hypothesis. Therefore, using Ineq. (5), we conclude that

$$\mathbb{P}[\max_{p:\ell_p(Q)>0}\{X_p\} \geq \mathrm{e}\,\kappa\,\alpha\, L(P^*)] \leq m\kappa^{-\mathrm{e}\,\kappa}$$

In words, the probability that the actual maximum delay caused by Q exceeds the optimal maximum delay by a factor greater than $\mathrm{e}\,\kappa\,\alpha$ is at most $m\kappa^{-\mathrm{e}\,\kappa}$. Therefore, for every $\kappa_0 \geq 2$,

$$\begin{aligned} L(Q) &\leq \mathbb{E}[\max_{p:\ell_p(Q)>0}\{X_p\}] \leq \mathrm{e}\,\alpha\, L(P^*)\big(\kappa_0 + \textstyle\sum_{k=\kappa_0}^{\infty} m k^{-\mathrm{e}\,k}\big) \\ &\leq \mathrm{e}\,\alpha\, L(P^*)\big(\kappa_0 + 2m\kappa_0^{-\mathrm{e}\,\kappa_0}\big) \end{aligned}$$

For $\kappa_0 = \frac{2\log m}{\log\log m}$, we obtain that $L(Q) \leq 2\,\mathrm{e}\,\alpha\,(\frac{\log m}{\log\log m} + 1)\, L(P^*)$. □

Theorem 2. *For every strategies profile Q at Nash equilibrium, $\lambda^{\max}(Q) \leq L(P^*) + \frac{2}{n}\, W(P^*)$.*

[6] We use the standard version of Hoeffding bound [14]: Let $X_1, X_2, \ldots, X_n$ be independent random variables with values in the interval $[0, w]$. Let $X = \sum_{i=1}^{n} X_i$ and let $\mathbb{E}[X]$ denote its expectation. Then, $\forall t > 0$, $\mathbb{P}[X \geq t] \leq (\frac{\mathrm{e}\,\mathbb{E}[X]}{t})^{t/w}$.

Proof. Let $\overline{f}$ denote the optimal fractional solution of the following quadratic program: QP $\equiv \min\{f^T(\frac{1}{2}M)f + A^T f : \mathbf{1}^T f \geq n, f \geq \mathbf{0}\}$. Notice that $\overline{f}$ is a splittable flow of value n. We first prove that $n\lambda^{\max}(Q) \leq C(\overline{f}) + 2\,W(\overline{f})$.

We use Dorn's Theorem [6], which establishes strong duality in quadratic programming[7], and prove that for every flow f,

$$n\,\delta^{\min}(f) - \tfrac{1}{2}C(f) \leq \tfrac{1}{2}C(\overline{f}) + W(\overline{f}) \tag{6}$$

The quadratic program QP $\equiv \min\{f^T(\frac{1}{2}M)f + A^T f : \mathbf{1}^T f \geq n, f \geq \mathbf{0}\}$ is always feasible and its optimal value is $\frac{1}{2}C(\overline{f}) + W(\overline{f})$. The Dorn's dual of QP is DP $\equiv \max\{z \cdot n - f^T(\frac{1}{2}M)f : Mf + A \geq \mathbf{1}z, z \geq 0\}$ (e.g., [6], [2, Chapter 6]). Every flow f becomes a feasible solution to DP by setting $z = \min_{p\in\mathcal{P}}\{\theta_p(f) + a_p\} \equiv \delta^{\min}(f)$. Hence, both the primal and the dual programs are feasible. Since the matrix M is symmetric and positive semi-definite, by Dorn's Theorem, the objective value of the optimal dual solution is exactly $\frac{1}{2}C(\overline{f}) + W(\overline{f})$[8].

Consequently, for every flow f, $(f, \delta^{\min}(f))$ is a feasible solution to DP and

$$n\,\delta^{\min}(f) - \tfrac{1}{2}C(f) \leq \tfrac{1}{2}C(\overline{f}) + W(\overline{f})$$

Let f^Q be the flow corresponding to the strategies profile Q. Since Q is a Nash equilibrium, $C(f^Q) \leq n\,\delta^{\min}(f^Q)$ by Proposition 3. Hence, $n\,\delta^{\min}(f^Q) \leq C(\overline{f}) + 2\,W(\overline{f})$. Using $\lambda^{\max}(Q) \leq \delta^{\min}(f^Q)$ by Ineq. (3), we obtain that $n\lambda^{\max}(Q) \leq C(\overline{f}) + 2\,W(\overline{f})$.

To conclude the proof, let f^* be the unsplittable flow corresponding to the pure strategies profile P^*, namely the optimal solution wrt. the objective of maximum latency. Then,

$$\begin{aligned} n\lambda^{\max}(Q) &\leq 2\,[\tfrac{1}{2}C(\overline{f}) + W(\overline{f})] \leq 2\,[\tfrac{1}{2}C(f^*) + W(f^*)] \\ &\leq nL(f^*) + 2W(f^*) = nL(P^*) + 2W(P^*) \end{aligned}$$

The second inequality holds because f^* is a feasible solution to QP. The third inequality holds because the average latency of f^* cannot exceed its maximum latency. For the last equality, since P^* is a pure strategies profile, its maximum latency and total load coincide with those of f^*. □

[7] Let $\min\{x^T Mx + c^T x : Ax \geq b, x \geq \mathbf{0}\}$ be the primal quadratic program. The Dorn's dual of this program is $\max\{-y^T My + b^T u : A^T u - 2My \leq c, u \geq \mathbf{0}\}$. Dorn [6] proved strong duality when the matrix M is symmetric and positive semi-definite. Thus, if M is symmetric and positive semi-definite and both the primal and the dual programs are feasible, their optimal solutions have the same objective value.

[8] The optimal dual solution is obtained from $\overline{f}$ by setting $z = \delta^{\min}(\overline{f})$. Since $\overline{f}$ is an optimal solution to the primal program, we can use Karush-Kuhn-Tucker optimality conditions (e.g. [2]) and prove that for any $s-t$ path p with $\overline{f}_p > 0$, $\theta_p(\overline{f}) + a_p = \delta^{\min}(\overline{f})$. Multiplying this equality by $\overline{f}_p$ and summing over all $p \in \mathcal{P}$, we obtain that

$$z \cdot n = \delta^{\min}(\overline{f}) \textstyle\sum_{p\in\mathcal{P}} \overline{f}_p = \sum_{p\in\mathcal{P}} \overline{f}_p(\theta_p(f) + a_p) = C(\overline{f}) + W(\overline{f})$$

Therefore, the dual objective value of $(\overline{f}, \delta^{\min}(\overline{f}))$ is exactly $\frac{1}{2}C(\overline{f}) + W(\overline{f})$.

Corollary 2. *For every strategies profile Q at Nash equilibrium, $\lambda^{\max}(Q) \leq 3\,L(P^*)$.*

Proof. We observe that $W(P^*) \leq n\,L(P^*)$ because P^* is a pure strategies profile. The corollary follows from Theorem 2. □

Theorem 3. *The price of anarchy for single-commodity network congestion games with identical users and latencies $d_e(x) = a_e x$ is at most $6\,\mathrm{e}\,(\frac{\log m}{\log\log m}+1)$.*

Acknowledgements. We wish to thank Burkhard Monien for suggesting the significance and the possibility of obtaining stronger results on the efficient computation of PNE in series-parallel networks.

References

1. B. Awerbuch, Y. Azar, and A. Epstein. The Price of Routing Unsplittable Flow. *STOC '05*, pp. 57–66, 2005.
2. M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming: Theory and Algorithms (2nd edition).* John Wiley and Sons, Inc., 1993.
3. G. Christodoulou and E. Koutsoupias. The Price of Anarchy of Finite Congestion Games. *STOC '05*, pp. 67–73, 2005.
4. J.R. Correa, A.S. Schulz, and N.E. Stier Moses. Computational Complexity, Fairness, and the Price of Anarchy of the Maximum Latency Problem. *IPCO '04*, LNCS 3064, pp. 59–73, 2004.
5. A. Czumaj and B. Vöcking. Tight Bounds for Worst-case Equilibria. *SODA '02*, pp. 413–420, 2002.
6. W.S. Dorn. Duality in Quadratic Programming. *Quarterly of Applied Mathematics*, 18(2):155–162, 1960.
7. R.J. Duffin. Topology of Series-Parallel Networks. *J. of Mathematical Analysis and Applications*, 10:303–318, 1965.
8. A. Fabrikant, C. Papadimitriou, and K. Talwar. The Complexity of Pure Nash Equilibria. *STOC '04*, pp. 604–612, 2004.
9. D. Fotakis, S. Kontogiannis, E. Koutsoupias, M. Mavronicolas, and P. Spirakis. The Structure and Complexity of Nash Equilibria for a Selfish Routing Game. *ICALP '02*, LNCS 2380, pp. 123–134, 2002.
10. D. Fotakis, S. Kontogiannis, and P. Spirakis. Selfish Unsplittable Flows. *ICALP '04*, LNCS 3142, pp. 593–605, 2004.
11. M. Gairing, T. Lücking, M. Mavronicolas, and B. Monien. Computing Nash Equilibria for Scheduling on Restricted Parallel Links. *STOC '04*, pp. 613–622, 2004.
12. M. Gairing, T. Lücking, M. Mavronicolas, B. Monien, and M. Rode. Nash Equilibria in Discrete Routing Games with Convex Latency Functions. *ICALP '04*, LNCS 3142, pp. 645–657, 2004.
13. R. L. Graham. Bounds on Multiprocessing Timing Anomalies. *SIAM Journal of Applied Mathematics*, 17(2):416–429, 1969.
14. W. Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *J. of the American Statistical Association*, 58(301):13–30, 1963.
15. E. Koutsoupias, M. Mavronicolas, and P. Spirakis. Approximate Equilibria and Ball Fusion. *Theory of Computing Systems*, 36:683–693, 2003.

16. E. Koutsoupias and C. Papadimitriou. Worst-case Equilibria. *STACS '99*, LNCS 1563, pp. 404–413, 1999.
17. L. Libman and A. Orda. Atomic Resource Sharing in Noncooperative Networks. *Telecommunication Systems*, 17(4):385–409, 2001.
18. T. Lücking, M. Mavronicolas, B. Monien, and M. Rode. A New Model for Selfish Routing. *STACS '04*, LNCS 2996, pp. 547–558, 2004.
19. M. Mavronicolas and P. Spirakis. The Price of Selfish Routing. *STOC '01*, pp. 510–519, 2001.
20. R.W. Rosenthal. A Class of Games Possessing Pure-Strategy Nash Equilibria. *International Journal of Game Theory*, 2:65–67, 1973.
21. T. Roughgarden. The Price of Anarchy is Independent of the Network Topology. *STOC '02*, pp. 428–437, 2002.
22. T. Roughgarden. The Maximum Latency of Selfish Routing. *SODA '04*, pp. 980–981, 2004.
23. T. Roughgarden and É. Tardos. How Bad is Selfish Routing? *J. of the ACM*, 49(2):236–259, 2002.