The Maximum Latency of Unsplittable Flows in Capacitated Networks

D. Fotakis, S. Kontogiannis, P. Spirakis

2004
The Maximum Latency of Unsplittable Flows in Capacitated Networks

Dimitris Fotakis\textsuperscript{1,2}, Spyros Kontogiannis\textsuperscript{1,3}, and Paul Spirakis\textsuperscript{1}

\textsuperscript{1} Research Academic Computer Technology Institute, Riga Feraiou 61, 26221 Patras, Greece. \{kontog,spirakis\}@cti.gr
\textsuperscript{2} Dept. of Mathematical, Physical and Computational Sciences, Aristotle University of Thessaloniki, 54006 Thessaloniki, Greece. fotakis@auth.gr
\textsuperscript{3} Dept. of Computer Science, University of Ioannina, 45110 Ioannina, Greece.

Abstract. We study the price of anarchy in (unweighted) single-commodity network congestion games relative to the objective of maximum latency. We prove that for any network of \(m\) edges and edge delays \(d_e(x) = a_e x\), the price of anarchy is \(\Theta(\frac{\log m}{\log \log m})\).

1 Introduction

We study the performance degradation due to lack of coordination in a simple selfish routing game where unsplittable traffic is routed on a capacitated network. We consider a network \(G(V,E)\) with a unique source \(s\) and destination \(t\) and some users willing to route their identical, unsplittable traffic demands from \(s\) to \(t\). For each edge \(e \in E\), the delay per user is given by a linear function \(d_e(x) = a_e x\), where \(x\) is the number of users routing their traffic on \(e\) and \(a_e\) denotes the inverse capacity of \(e\). We assume that the users are selfish and do not cooperate. Each of them selects the routing strategy minimizing his total delay given the routing strategies of other users. We are interested in situations that the users have reached a Nash equilibrium: a "stable point" from which no user is willing to deviate unilaterally. We measure the performance degradation due to lack of users' coordination using the price of anarchy \cite{2}, that is the worst-case ratio between the cost of a Nash equilibrium and the cost of an optimal solution.

In this paper, we measure the price of anarchy relative to the maximum latency incurred by the traffic. We show that the price of anarchy is \(\Theta(\frac{\log m}{\log \log m})\). In other words, the worst case instance (with respect to the price of anarchy) among unweighted congestion games in networks with \(m\) edges is essentially the parallel edges game introduced in \cite{2}.

The Model. We consider a communication network \(G(V,E)\) with a unique source \(s \in V\) and destination \(t \in V\) and a set \(N \equiv [n]\) of users of identical traffic demands willing to route their traffic from \(s\) to \(t\). Without loss of generality, we assume that each user controls a unit of traffic.

Let \(\mathcal{P}\) denote the set of paths from the source \(s\) to the destination \(t\) and \(m \equiv |E|\) denote the number of edges in the network. There is a linear delay function \(d_e(x) = a_e x\), \(a_e \geq 0\), associated with each edge \(e \in E\). Alternatively, we can regard \(a_e\) as the inverse capacity of edge \(e\). For each \(s-t\) path \(\pi \in \mathcal{P}\), \(a_\pi \equiv \sum_{e \in \pi} a_e\) denotes the sum of inverse capacities of its edges.

Each user routes his traffic on a single path \(\pi \in \mathcal{P}\). A tuple \(\varpi = (\varpi_1, \ldots, \varpi_n) \in \times_{i=1}^n \mathcal{P} \equiv \mathcal{P}^n\) is called a pure strategies profile, or simply a user configuration. A real vector \(p = (p_1, \ldots, p_n)\) such that each \(p_i : \mathcal{P} \mapsto [0,1]\) is a probability distribution over \(\mathcal{P}\) is called a mixed strategies profile.
For any configuration \( \varpi \) and edge \( e \), let \( \theta_e(\varpi) = a_e \{ i \in N : e \in \varpi_i \} \) be the delay of \( e \) with respect to \( \varpi \). The cost \( \lambda^i(\varpi) \) of user \( i \) for routing his traffic on path \( \varpi_i \) in a given configuration \( \varpi \) is equal to the total delay along \( \varpi_i \):

\[
\lambda^i(\varpi) = \lambda_{\varpi_i}(\varpi) = \sum_{e \in \varpi_i} \theta_e(\varpi)
\]

On the other hand, for a mixed strategies profile \( p \), the expected cost of user \( i \) for routing his traffic on path \( \varpi_i \) is:

\[
\lambda^{i, p}(\varpi) = \sum_{\varpi^{-i}} P(\varpi^{-i}, \varpi^{-i}) \cdot \sum_{e \in \varpi_i} \theta_e(\varpi^{-i} \oplus \varpi_i)
\]

where \( \varpi^{-i} \) is a configuration of all the users except for user \( i \), \( p^{-i} \) is the mixed strategies profile of all users except for \( i \), \( \varpi^{-i} \oplus \varpi_i \) is the new configuration with user \( i \) choosing strategy \( \varpi_i \), and \( P(\varpi^{-i}, \varpi^{-i}) \equiv \prod_{j \in N \setminus \{i\}} p_j(\varpi_j) \) is the occurrence probability of \( \varpi^{-i} \).

We say that a (mixed in general) strategies profile \( p \) is a Nash Equilibrium (NE) if and only if \( \forall i \in N, \forall \varpi_i, \pi_i \in P, p_i(\varpi_i) > 0 \Rightarrow \lambda^{i, p}(\varpi) \leq \lambda^{i, p}(\varpi) \).

We measure the cost of a strategies profile using the objective of maximum latency. The social cost \( SC(p) \) of a strategies profile \( p \) is

\[
SC(p) = \sum_{\varpi \in \mathcal{P}^n} P(\varpi, \varpi) \cdot \max_{i \in N} \{ \lambda_{\varpi_i}(\varpi) \}
\]

where \( P(\varpi, \varpi) \equiv \prod_{i=1}^n p_i(\varpi_i) \) is the occurrence probability of configuration \( \varpi \) in the mixed strategies profile \( p \). The social optimum with respect to the objective of maximum latency is

\[
OPT = \min_{\varpi \in \mathcal{P}^n} \{ \max_{i \in N} [\lambda_{\varpi_i}(\varpi)] \}
\]

The price of anarchy is then defined as

\[
\mathcal{R} = \max_{p \text{ is a NE}} \left\{ \frac{SC(p)}{OPT} \right\}
\]

Related Work. In the seminal paper [12] the price of anarchy (or coordination ratio) was introduced as a means for measuring the performance degradation due to lack of coordination. In this work it was proved that the price of anarchy is 3/2 for two capacitated parallel edges, while for \( m \) parallel edges and users of varying demands, the price of anarchy is between \( \Omega(\frac{\log m}{\log \log m}) \) and \( O(\sqrt{m \log m}) \). For \( m \) identical parallel edges, [13] proved that \( \mathcal{R} = \Theta(\frac{\log m}{\log \log m}) \) for the fully-mixed NE, while for the case of \( m \) identical parallel edges and users of varying demands it was shown in [11] that \( \mathcal{R} = \Theta(\sqrt{m \log m}) \).

In [5] it was finally shown that \( \mathcal{R} = \Theta(\frac{\log m}{\log \log m}) \) for the general case of capacitated parallel edges and users of varying demands. [4] presents a thorough study of the case of general, monotone delay functions on parallel edges, with emphasis on delay functions from queuing theory. Unlike the case of linear cost functions, they show that the price of anarchy for non-linear delay functions in general is far worse and often unbounded.

[15] initiated the study of the price of anarchy in multi-commodity network congestion games\(^1\) with infinitely many users each controlling a negligible amount of traffic (non-atomic setting). The

\(^1\) We call a network single-commodity if there is a single source-destination pair and multi-commodity otherwise.
price of anarchy is relative to the objective of total latency in [15]. For linear edge delays, the price of anarchy is at most $4/3$. For general delays, the total latency of any Nash flow is at most the total latency of an optimal flow routing twice as much traffic. [16] proves that it is actually the class of edge delay functions and not the network topology that determines the price of anarchy for the total latency in the non-atomic setting.

[17, 3, 18] study the price of anarchy relative to the maximum latency in a single-commodity network congestion game with infinitely many users (non-atomic setting). [18] observes that in a single-commodity network, any upper bound on the price of anarchy for the total latency is also an upper bound on the price of anarchy for the maximum latency. Hence, the price of anarchy for linear delays is $4/3$. This is not true for multi-commodity networks, where the price of anarchy is $\Omega(n)$ even for linear delays [18, 3]. On the other hand, [17] proves that in single-commodity networks, the price of anarchy for maximum latency is at most $|V| - 1$ even if the price of anarchy for total latency is much worse. [17] conjectures that the bound of $|V| - 1$ holds for multi-commodity networks.

[9] studies the price of anarchy for the maximum latency in a weighted single-commodity network congestion game with a finite number of users, each controlling a non-negligible amount of traffic (atomic setting). [9] focuses on $\ell$-layered graphs with edge delays equal to the edge loads and users of varying demands\(^2\) and proves that the price of anarchy remains $\Theta(\frac{\log m}{\log \log m})$. The proof is based on a natural correspondence between mixed strategies profiles and splittable flows in the underlying network. The idea is to use quadratic programming duality and establish that the total delay along the best $s-t$ path in any Nash equilibrium is bounded from above by a convex combination of the equilibrium’s average latency and the optimal average latency.

**Contribution.** In this paper, we generalize the approach of [9] and prove that the price of anarchy in (unweighted) single-commodity network congestion games with edge delays $d_e(x) = a_e x$ remains $\Theta(\frac{\log m}{\log \log m})$.

**Organization.** In Section 2, we present a natural correspondence between mixed strategies profiles and splittable $s-t$ flows. Section 3 comprises the main part of the paper. In Section 3.1, we motivate our approach by establishing that the optimal solution to a quadratic program corresponds to a mixed Nash equilibrium. In Section 3.2, we use quadratic programming duality and show that the maximum expected latency of any Nash equilibrium is within a small constant factor of the optimal maximum latency. Finally, in Section 3.3, we apply a Chernoff-Hoeffding bound and prove that the expected maximum latency of any Nash equilibrium is $O\left(\frac{\log m}{\log \log m}\right)$ times the optimal maximum latency.

## 2 Flows and Mixed Strategies Profiles

Given a network $G(V, E)$ with a source $s$ and a destination $t$, and $n$ users of unit traffic demands, a *feasible* flow is a function $\rho : \mathcal{P} \rightarrow \mathbb{R}_{\geq 0}$ such that $\sum_{\pi \in \mathcal{P}} \rho(\pi) = n$, i.e., all users’ demands are actually met. We sometimes use $\rho$ to denote the $|\mathcal{P}|$-dimensional vector corresponding to the flow $\rho$. Let $\rho(e) \equiv \sum_{\pi : e \in \pi} \rho(\pi)$ denote the amount of flow routed on edge $e \in E$, $\theta_e(\rho) \equiv d_e(\rho(e)) = a_e \rho(e)$ denote the delay of $e$ with respect to the flow $\rho$, and $\theta_\pi(\rho) \equiv \sum_{e \in \pi} \theta_e(\rho)$ denote the total delay of path $\pi \in \mathcal{P}$ with respect to the flow $\rho$.

\(^2\)This generalizes the case of identical parallel edges and users of varying demands considered in [12, 13, 11, 5].
We distinguish between unsplitable and splittable (feasible) flows. A flow is unsplitable if each user’s traffic demand is routed on a single \(s - t\) path. A flow is splittable if the traffic demand of some user is routed on several \(s - t\) paths.

We map a mixed strategies profile \(p = (p_1, p_2, \ldots, p_n)\) to a flow \(\rho_p\) as follows: For each \(s - t\) path \(\pi \in \mathcal{P}\), \(\rho_p(\pi) = \sum_{i \in [n]} p_i(\pi)\). That is, we handle the expected load routed on \(\pi\) by \(p\) as a splittable flow, where user \(i\) routes a fraction of \(p_i(\pi)\) of his (unit) demand on \(\pi\). Observe that if \(p\) is a pure strategies profile, the corresponding flow is then unsplitable. For each edge \(e \in E\),

\[
\theta_e(p) \equiv a_e \sum_{i=1}^{n} \sum_{\pi : e \in \pi} p_i(\pi) = a_e \sum_{\pi : e \in \pi} \rho_p(\pi) = a_e \rho_p(e) = d_e(\rho_p(e)) \equiv \theta_e(\rho_p)
\]
denotes the expected delay of \(e\) with respect to \(p\). As for the expected total delay of a path \(\pi \in \mathcal{P}\) with respect to \(p\), this is

\[
\theta_\pi(p) = \sum_{e \in \pi} \theta_e(p) = \sum_{e \in \pi} a_e \sum_{\pi' : e \in \pi'} \rho_p(\pi') = \sum_{\pi' \in \mathcal{P}} \rho_p(\pi') \sum_{e \in \pi' \cap \pi} a_e \equiv \theta_\pi(\rho_p)
\]

For simplicity, we drop the subscript of \(p\) from its corresponding flow \(\rho_p\), when the mixed strategies profile is clear by the context.

For a feasible flow \(\rho\), let \(a(\rho) \equiv \max_{\pi : \rho(\pi) > 0} \{a_\pi\}\) denote the inverse capacity of the “slowest” least capacity path used by \(\rho\), and let \(d_{\min}(\rho) \equiv \min_{\pi \in \mathcal{P}} \{\theta_\pi(\rho) + a_\pi\}\).

**Maximum Latency, Total Latency, and Total Load.** A flow \(\rho\) can be evaluated by the objective of maximum latency, which is defined as

\[
L(\rho) \equiv \max_{\pi : \rho(\pi) > 0} \{\theta_\pi(\rho)\} = \max_{\pi : \exists i, \rho(\pi) > 0} \{\theta_\pi(p)\} \equiv L(p)
\]

In other words, \(L(\rho)\) is the maximum expected delay incurred by the users with respect to \(p\). From now on, we use \(\rho^*\) to denote the optimal unsplitable flow for the objective of maximum latency.

In addition, a feasible flow \(\rho\) can be evaluated by the objective of total latency, which is defined as

\[
C(\rho) \equiv \sum_{\pi \in \mathcal{P}} \rho(\pi)\theta_\pi(\rho) = \sum_{e \in E} \theta_e(\rho)p(e) = \sum_{e \in E} a_e\rho^2(e) \equiv C(p)
\]

The second equality is obtained by summing over the edges of \(\pi\) and reversing the order of the summation.

Moreover, a feasible flow \(\rho\) can be evaluated by the objective of total load, which is defined as

\[
W(\rho) \equiv \sum_{e \in E} a_e\rho(e) = \sum_{e \in E} a_e \sum_{\pi : e \in \pi} \rho(\pi) = \sum_{\pi \in \mathcal{P}} a_e\rho(\pi) \equiv W(p)
\]

The third equality is obtained by reversing the order of the summation and using \(a_\pi \equiv \sum_{e \in \pi} a_e\).

Let \(Q\) be the square matrix consisting of the sum of inverse capacities of the edges shared by each pair of paths. Formally, \(Q\) is a \(|\mathcal{P}| \times |\mathcal{P}|\) matrix and for every \(\pi, \pi' \in \mathcal{P}\), \(Q[\pi, \pi'] \equiv \sum_{e \in \pi \cap \pi'} a_e\). By definition, \(Q\) is a symmetric matrix. For a feasible flow \(\rho\), \(Q\rho\) is the \(|\mathcal{P}|\)-dimensional vector with coordinates \(\theta_\pi(\rho)\) for each \(\pi \in \mathcal{P}\). Furthermore, the total latency of \(\rho\) can be expressed as \(C(\rho) = \rho^T Q\rho\). Since

\[
C(\rho) = \rho^T Q\rho = \sum_{\pi \in \mathcal{P}} \theta_\pi(\rho)\rho(\pi) = \sum_{e \in E} a_e\rho^2(e),
\]
the matrix $Q$ is positive semi-definite$^3$.

Let $A$ be the $|P|$-dimensional vector with coordinates $a_\pi$ for each $s$–$t$ path $\pi$. Formally, for each $\pi \in \mathcal{P}$, $A[\pi] \equiv a_\pi = \sum_{e \in \pi} a_e$. We can express the total load of $\rho$ as $W(\rho) = A^T \rho$.

**Flows at Nash Equilibrium.** Let $p$ be a mixed strategies profile, and let $\rho$ be the corresponding flow. We sometimes write that the flow $\rho$ is at Nash equilibrium with the understanding that it is actually $p$ which is at Nash equilibrium.

The cost of user $i$ on a path $\pi$ is $\lambda^i_\pi(p) = \theta^{-i}_\pi(p) + a_\pi$ (recall that each user controls a unit of traffic), where $\theta^{-i}_\pi(p)$ is the expected delay along the path $\pi$ if the demand of user $i$ was removed from the system:

$$
\theta^{-i}_\pi(p) = \sum_{\pi' \in \mathcal{P}} Q[\pi, \pi'] \sum_{j \neq i} p_j(\pi') = \theta_\pi(p) - \sum_{\pi' \in \mathcal{P}} Q[\pi, \pi'] p_i(\pi')
$$

Thus, $\lambda^i_\pi(p) = \theta_\pi(p) + a_\pi - \sum_{\pi' \in \mathcal{P}} Q[\pi, \pi'] p_i(\pi')$.

Let $\rho$ be a flow at Nash equilibrium and $\pi \in \mathcal{P}$ be any path with $\rho(\pi) > 0$. Then,

$$
\max \{\theta_\pi(\rho), a_\pi\} \leq d^{\min}(\rho) \equiv \min_{\pi \in \mathcal{P}} \{\theta_\pi(\rho) + a_\pi\}
$$

(4)

Otherwise, the users routing their traffic on the path $\pi$ could improve their delay by defecting to a path minimizing $\theta_\pi(\rho) + a_\pi$. The following useful proposition is an immediate consequence of Ineq. (4).

**Proposition 1.** Let $\rho$ be a flow at Nash equilibrium. For any $\alpha \in [0, 1],

$$
\alpha C(\rho) + (1 - \alpha)W(\rho) \leq n d^{\min}(\rho)
$$

**Proof.** By Ineq. (4), for any path $\pi \in \mathcal{P}$ with $\rho(\pi) > 0$, $\rho(\pi)\theta_\pi(\rho) \leq \rho(\pi)d^{\min}(\rho)$. Summing up over all paths $\pi$ with $\rho(\pi) > 0$, we conclude that $C(\rho) \leq n d^{\min}(\rho)$. Similarly, we prove that $W(\rho) \leq n d^{\min}(\rho)$.

$\square$

### 3 The Price of Anarchy in Capacitated Networks

Our approach is motivated by the fact that $[C(\rho) + W(\rho)]/2$ is an exact potential function for our network congestion game (e.g. [14], [9, Theorem 1]). Hence, any unsplittable flow $\rho$ being a local optimum with respect to $[C(\rho) + W(\rho)]/2$ is a pure Nash equilibrium.

We first prove that any feasible flow at Nash equilibrium is a good approximation to the feasible (splittable) flow minimizing $C(\rho)/2 + W(\rho)$ (Lemma 1). The proof is based on Dorn’s Theorem [7] establishing strong duality in quadratic programming and Proposition 1. On the other hand, the objective value of the optimal flow with respect to the previous objective cannot exceed $\frac{3}{2}C(\rho^*)$ (Lemma 2). Then, we apply Lemma 1, Lemma 2, and Ineq. (4) and show that for any flow $\rho$ at Nash equilibrium, $\max \{L(\rho), a(\rho)\} \leq 6 L(\rho^*)$ (Lemma 3). Finally, a standard application of the Chernoff-Hoeffding bound yields that the price of anarchy is $O\left(\frac{\log n}{\log \log n}\right)$ (Lemma 4).

$^3$ A $n \times n$ matrix $Q$ is positive semi-definite if for every vector $x \in \mathbb{R}^n$, $x^T Q x \geq 0$. 
3.1 Computing a Mixed Nash Equilibrium

In this section, we prove that the feasible flow minimizing $\frac{n-1}{2n}C(\rho) + W(\rho)$ corresponds to a mixed Nash equilibrium. Formally, let $\hat{\rho}$ be the optimal fractional solution to the following quadratic program

$$\min \left\{ \frac{n-1}{2n} \rho^T Q \rho + A^T \rho : 1^T \rho \geq n, \rho \geq 0 \right\},$$

where $1$ (resp. $0$) denotes the $|P|$-dimensional vector with $1$ (resp. $0$) in each coordinate. We observe that $\hat{\rho}$ routes exactly $n$ units of traffic (i.e., satisfies the first constraint with equality).

**Proposition 2.** Let $p$ be the mixed strategies profile where every user routes its traffic on each path $\pi$ with probability $\hat{\rho}(\pi)/n$. Then, $p$ is at Nash equilibrium.

**Proof.** By construction, the expected path loads corresponding to $p$ are equal to the values of $\hat{\rho}$ on these paths. Since all users follow exactly the same strategy and route their demand on path $\pi$ with probability $\hat{\rho}(\pi)/n$, for each user $i$,

$$\lambda_i^\pi(p) = \theta_i^\pi(p) = \frac{1}{n} \sum_{\pi' \in P} Q[\pi, \pi'][\hat{\rho}(\pi')] = \frac{n-1}{n} \theta_i(p) = \frac{n-1}{n} \theta_i(\hat{\rho})$$

By definition, $\hat{\rho}$ is the feasible flow minimizing the convex function $\sum_{e \in E} \left[ \frac{n-1}{2n} a_e \rho^2(e) + a_e \rho(e) \right]$. Therefore, for every $\pi_1, \pi_2 \in P$ with $\hat{\rho}(\pi_1) > 0$,

$$\sum_{e \in \pi_1} \left[ \frac{n-1}{n} a_e \hat{\rho}(e) + a_e \right] = \theta^{-1}_{\pi_1}(p) + a_{\pi_1} \leq \sum_{e \in \pi_2} \left[ \frac{n-1}{n} a_e \hat{\rho}(e) + a_e \right] = \theta^{-1}_{\pi_2}(p) + a_{\pi_2}$$

(e.g., [2, 6], [15, Lemma 2.5]).

Therefore, for every user $i$ and every $\pi_1, \pi_2 \in P$ such that the user $i$ routes his (unit) traffic demand on $\pi_1$ with positive probability,

$$\lambda_{\pi_1}^i(p) = \theta_{\pi_1}^{-1}(p) + a_{\pi_1} \leq \theta_{\pi_2}^{-1}(p) + a_{\pi_2} = \lambda_{\pi_2}^i(p)$$

Consequently, $p$ is at Nash equilibrium. \qed

**Remark.** If the network consists of $m$ parallel capacitated edges, the mixed Nash equilibrium of Proposition 2 is identical to the generalized fully mixed equilibrium of [8, Theorem 5].

3.2 An Upper Bound on the Expected Latency of Nash Equilibria

Let $\overline{p}$ be the (splittable) flow corresponding to the optimal solution of the following quadratic program: $QP \equiv \min \{ \rho^T (\frac{1}{2} Q) \rho + A^T \rho : 1^T \rho \geq n, \rho \geq 0 \}$. Since no flow of value strictly greater than $n$ can be an optimal solution to QP, it must be $\sum_{\pi \in P} \overline{p}(\pi) = n$. The following lemma establishes that any feasible flow at Nash equilibrium is a $4$-approximation to the optimal solution of QP.

**Lemma 1.** For every feasible flow $\rho$ corresponding to a mixed strategies profile at Nash equilibrium, $C(\rho) + W(\rho) \leq 4 \left[ \frac{1}{2} C(\overline{p}) + W(\overline{p}) \right]$. 

---

6
Proposition 1 implies the lemma. This concludes the proof of Ineq. (5).

By setting $z = \min_{\pi \in \mathcal{P}} \{ \theta_{\pi}(\rho) + a_\pi \} \equiv d_{\text{min}}(\rho)$. Hence, both the primal and the dual programs are feasible. Since the matrix $\frac{1}{2}Q$ is symmetric and positive semi-definite, by Dorn’s Theorem [7], the objective value of the optimal dual solution is exactly $\frac{1}{2}C(\overline{\rho}) + W(\overline{\rho})$.

Since for any flow $\rho$, the solution $(\rho, d_{\text{min}}(\rho))$ is a feasible solution to DP, it must be

$$n d_{\text{min}}(\rho) - \frac{1}{2}C(\rho) \leq \frac{1}{2}C(\overline{\rho}) + W(\overline{\rho})$$

This concludes the proof of Ineq. (5).

Applying Proposition 1 with $\alpha = \frac{2}{n}$, we conclude that for every feasible flow $\rho$ at Nash equilibrium,

$$\frac{1}{2}[C(\rho) + W(\rho)] \leq n d_{\text{min}}(\rho) - \frac{1}{2}C(\rho) \leq \frac{1}{2}C(\overline{\rho}) + W(\overline{\rho})$$

On the other hand, the objective value of $\overline{\rho}$ (wrt the objective of QP) cannot exceed 1.5 times the total latency of the optimal unsplittable flow $\rho^*$.

Lemma 2. Let $\overline{\rho}$ be the optimal solution to QP, and let $\rho^*$ be the optimal unsplittable flow wrt the maximum latency objective. Then, $\frac{1}{2}C(\overline{\rho}) + W(\overline{\rho}) \leq \frac{3}{2}C(\rho^*)$.

Proof. Since $\rho^*$ is an unsplittable flow, for every edge $e \in E$, either $\rho^*(e) = 0$ or $\rho^*(e) \geq 1$. Therefore, the total load of $\rho^*$ can never exceed its total latency. Formally,

$$W(\rho^*) = \sum_{e \in E} a_e \rho^*(e) \leq \sum_{e \in E} a_e \rho^*(e)^2 = C(\rho^*)$$

\[4\] Let $\min\{x^T Q x + c^T x : A x \geq b, x \geq 0\}$ be the primal quadratic program. The Dorn’s dual of this program is $\max\{-y^T Q y + b^T u : A^T u - 2Qy \leq c, u \geq 0\}$. Dorn [7] proved strong duality when the matrix $Q$ is symmetric and positive semi-definite. Thus, if $Q$ is symmetric and positive semi-definite and both the primal and the dual programs are feasible, their optimal solutions have the same objective value.

The optimal dual solution is actually obtained from $\overline{\rho}$ by setting $z = d_{\text{min}}(\overline{\rho})$. Since $\overline{\rho}$ is an optimal solution to the primal program, we can use Karush-Kuhn-Tucker optimality conditions (e.g. [1]) and prove that for any $s \rightarrow t$ path $\pi$ with $\overline{\rho}(\pi) > 0$, $\theta_{\pi}(\overline{\rho}) + a_\pi = d_{\text{min}}(\overline{\rho})$. Multiplying this equality by $\overline{\rho}(\pi)$ and summing over all $\pi \in \mathcal{P}$, we obtain that

$$z \cdot n = d_{\text{min}}(\overline{\rho}) \sum_{\pi \in \mathcal{P}} \overline{\rho}(\pi) = \sum_{\pi \in \mathcal{P}} \overline{\rho}(\pi)[\theta_{\pi}(\rho) + a_\pi] = C(\overline{\rho}) + W(\overline{\rho})$$

Therefore, the dual objective value of the solution $(\overline{\rho}, d_{\text{min}}(\overline{\rho}))$ is exactly $\frac{1}{2}C(\overline{\rho}) + W(\overline{\rho})$ and hence $\overline{\rho}$ also corresponds to the optimal solution of the dual program.
Therefore,

$$\frac{1}{2}C(\overline{f}) + W(\overline{f}) \leq \frac{1}{2}C(\rho^*) + W(\rho^*) \leq \frac{3}{2}C(\rho^*),$$

because $\overline{f}$ is the feasible flow minimizing $\frac{1}{2}C(\overline{f}) + W(\overline{f}).$ \hfill $\square$

We are now ready to establish the main technical lemma of this section.

**Lemma 3.** For every feasible flow $\rho$ corresponding to a mixed strategies profile at Nash equilibrium, $\max\{L(\rho), a(\rho)\} \leq 6 L(\rho^*).$

**Proof.** We first show that $L(\rho) \leq 6 L(\rho^*).$ To reach a contradiction, we assume that there exists a path $\pi \in \mathcal{P}$ with $\rho(\pi) > 0$ such that $\theta_{\pi}(\rho) > 6 L(\rho^*).$ By Ineq. (4), for any $\pi' \in \mathcal{P},$

$$6 L(\rho^*) < \theta_{\pi}(\rho) \leq \theta_{\pi'}(\rho) + a_{\pi'}$$

Multiplying by $\rho(\pi')$ and summing over all $\pi' \in \mathcal{P},$ we obtain that $6 n L(\rho^*) < C(\rho) + W(\rho).$ Thus,

$$6 C(\rho^*) \leq 6 n L(\rho^*) < C(\rho) + W(\rho) \leq 4 \left[ \frac{1}{2}C(\overline{f}) + W(\overline{f}) \right] \leq 6 C(\rho^*),$$

which is a contradiction. The first inequality holds because the average latency of any feasible flow cannot exceed its maximum latency and hence, $C(\rho^*) \leq n L(\rho^*).$ The third inequality follows from Lemma 1 since $\rho$ is at Nash equilibrium. The last inequality follows from Lemma 2.

The proof that $a(\rho) \leq 6 L(\rho^*)$ is identical. \hfill $\square$

### 3.3 Bounding the Price of Anarchy

Next we derive an upper bound on the social cost of any strategy profile corresponding to a flow $\rho$ with $\max\{L(\rho), a(\rho)\} \leq \alpha L(\rho^*).$

**Lemma 4.** Let $\rho^*$ be the optimal unsplittable flow, let $p$ be a mixed strategies profile and let $\rho$ be the feasible flow corresponding to $p.$ If there exists some constant $\alpha \geq 1$ such that $\max\{L(\rho), a(\rho)\} \leq \alpha L(\rho^*),$ then

$$\text{SC}(p) \leq 2\alpha O\left(\frac{\log m}{\log \log m}\right) L(\rho^*),$$

where $m = |E|$ denotes the number of edges in the network.

**Proof.** For each edge $e \in E$ and each user $i,$ let $X_{e,i}$ be the random variable describing the actual load routed through $e$ by $i.$ The random variable $X_{e,i}$ is equal to 1 if $i$ routes its demand on a path $\pi$ including $e$ and 0 otherwise. Consequently, the expectation of $X_{e,i}$ is equal to $\mathbb{E}[X_{e,i}] = \sum_{\pi:e \in \pi} p_i(\pi).$ Since each user selects its path independently, for every fixed edge $e,$ the random variables $\{X_{e,i}, i \in [n]\},$ are independent from each other.

For each edge $e \in E,$ let $X_e = a_e \sum_{i=1}^n X_{e,i}$ be the random variable that describes the actual delay paid 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,

$$\mathbb{E}[X_e] = a_e \sum_{i=1}^n \mathbb{E}[X_{e,i}] = a_e \sum_{i=1}^n \sum_{\pi:e \in \pi} p_i(\pi) = a_e \sum_{\pi:e \in \pi} \sum_{i=1}^n p_i(\pi) = a_e \sum_{\pi:e \in \pi} \rho(\pi) = \theta_e(\rho)$$
By applying the standard Hoeffding bound with \( w = a_e \) and \( t = e \kappa \max \{ \theta_e(\rho), a_e \} \), we obtain that for every \( \kappa \geq 1 \),
\[
P[X_e \geq e \kappa \max \{ \theta_e(\rho), a_e \}] \leq \kappa^{-e \kappa}.
\]

For \( m \equiv |E| \), by applying the union bound we conclude that
\[
P[\exists e \in E : X_e \geq e \kappa \max \{ \theta_e(\rho), a_e \}] \leq m \kappa^{-e \kappa} \quad (6)
\]

For each path \( \pi \in \mathcal{P} \) with \( \rho(\pi) > 0 \), we define the random variable \( X_\pi = \sum_{e \in \pi} X_e \) describing the actual delay along \( \pi \). The social cost of \( p \), which is equal to the expected maximum delay experienced by some user, cannot exceed the expected maximum delay among paths \( \pi \) with \( \rho(\pi) > 0 \). Formally,
\[
SC(p) \leq E[\max_{\pi: \rho(\pi) > 0} \{ X_\pi \}].
\]

If for all \( e \in E \), \( X_e \leq e \kappa \max \{ \theta_e(\rho), a_e \} \), then for every path \( \pi \in \mathcal{P} \) with \( \rho(\pi) > 0 \),
\[
X_\pi = \sum_{e \in \pi} X_e \leq e \kappa \sum_{e \in \pi} \max \{ \theta_e(\rho), a_e \}
\leq e \kappa \sum_{e \in \pi} (\theta_e(\rho) + a_e)
\leq e \kappa (\theta_\pi(\rho) + a_\pi)
\leq e \kappa (L(\rho) + a(\rho))
\leq 2 e \kappa L(\rho^*)
\]

The third equality follows from \( \theta_\pi(\rho) = \sum_{e \in \pi} \theta_e(\rho) \) and \( a_\pi = \sum_{e \in \pi} a_e \), the fourth inequality from \( \theta_\pi(\rho) \leq L(\rho) \) and \( a_\pi \leq a(\rho) \) since \( \rho(\pi) > 0 \), and the last inequality from the hypothesis that \( \max \{ L(\rho), a(\rho) \} \leq \alpha L(\rho^*) \). Therefore, using Ineq. (6), we conclude that
\[
P[\max_{\pi: \rho(\pi) > 0} \{ X_\pi \} \geq 2 e \kappa L(\rho^*)] \leq m \kappa^{-e \kappa}.
\]

In other words, the probability that the actual maximum delay caused by \( p \) exceeds the optimal maximum delay by a factor greater than \( 2 e \kappa \) is at most \( m \kappa^{-e \kappa} \). Therefore, for every \( \kappa_0 \geq 2 \),
\[
SC(p) \leq E[\max_{\pi: \rho(\pi) > 0} \{ X_\pi \}] \leq 2 e \kappa L(\rho^*) \left( \kappa_0 + \sum_{k=\kappa_0}^{\infty} k \kappa^{-e \kappa} \right)
\leq 2 e \kappa L(\rho^*) \left( \kappa_0 + 2 \kappa_0^{-e \kappa_0+1} \right)
\]

If \( \kappa_0 = \frac{2 \log m}{\log \log m} \), then \( \kappa_0^{-e \kappa_0+1} \leq m^{-1}, \forall m \geq 4 \). Thus, \( SC(p) \leq 4 e \kappa (\frac{\log m}{\log \log m} + 1) L(\rho^*) \).

The following theorem is an immediate consequence of Lemma 3 and Lemma 4.

**Theorem 1.** The price of anarchy for unweighted single-commodity network congestion games with edge delays \( d_e(x) = a_e x, a_e \geq 0 \), is at most 24 \( e (\frac{\log m}{\log \log m} + 1) \).

---

\* We use the standard version of Hoeffding bound ([10]): 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 \( E[X] \) denote its expectation. Then, \( \forall t > 0 \), \( P[X \geq t] \leq (\frac{E[X]}{t})^{1/w} \).
References