Theoretical and Computational Aspects of Resource Allocation Games

Dr. Tobias Harks

Kumulative Habilitationsschrift an der
Fakultät II – Mathematik und Naturwissenschaften
der Technischen Universität Berlin

Lehrgebiet:
Mathematik

Eröffnung des Verfahrens: 26.09.2011
Verleihung der Lehrbefähigung: 28.06.2012
Ausstellung der Urkunde: 09.07.2012
Aushändigung der Urkunde:

Gutachter:
Prof. Dr. Susanne Albers
Prof. Dr. Patrice Marcotte
Prof. Dr. Rolf H. Möhring
Prof. Dr. Berthold Vöcking

Tag der wissenschaftlichen Aussprache: 04.06.2012

Berlin 2012

D 83

List of Publications

The following nine papers are contained in this cumulative habilitation thesis.

(1) Vincenzo Bonifaci, Tobias Harks and Guido Schäfer
Stackelberg Routing in Arbitrary Networks
Math. Oper. Res. **35** (2010), no. 2, pp. 330–346

(2) Tobias Harks
Stackelberg Strategies and Collusion in Network Games with Splittable Flow
Theory of Computing Systems **48** (2011), no. 4, pp. 781–802

(3) Tobias Harks and Konstantin Miller
The Worst-Case Efficiency of Cost Sharing Methods in Resource Allocation Games
Oper. Res. **56** (2011), no. 6, pp.1491–1503

(4) Philipp von Falkenhausen and Tobias Harks
Optimal Cost Sharing Protocols for Scheduling Games
In Proc. of the 12th ACM Conference on Electronic Commerce (EC), 2011, pp. 285-294
Full version to appear in Mathematics of Operations Research

(5) Tobias Harks, Max Klimm and Rolf H. Möhring
Characterizing the Existence of Potential Functions in Weighted Congestion Games
Theory of Computing Systems **49** (2011), no. 1, pp. 46–70

(6) Tobias Harks and Max Klimm
On the Existence of Pure Nash Equilibria in Weighted Congestion Games
Math. Oper. Res. **37** (2012), no. 3, pp. 419-436

(7) Tobias Harks, Max Klimm and R. H. Möhring
Strong Nash Equilibria in Games with the Lexicographical Improvement Property
In Proceedings of the 5th International Workshop On Internet And Network Economics (WINE), LNCS 5929, Springer, pp. 455-462, 2009
Full version to appear in International Journal of Game Theory

(8) Tobias Harks, Martin Hoefer, Max Klimm and Alexander Skopalik
Computing Pure and Strong Nash Equilibria in Bottleneck Congestion Games
In Proc. of the 18th Annual European Symposium on Algorithms (ESA), 2010 LNCS 6347, Springer, pp. 29-38, 2010
Full version to appear in Mathematical Programming

(9) Tobias Harks, Jannik Matuschke and Felix G. König
Approximation Algorithms for Capacitated Location Routing
Transportation Science, to appear

Contents

Introduction

This collection contains the following nine papers that I submit for obtaining the habilitation at the Technische Universität Berlin, Fakultät II – Mathematik und Naturwissenschaften.

(1) *Stackelberg Routing in Arbitrary Networks*
(2) *Stackelberg Strategies and Collusion in Network Games with Splittable Flow*
(3) *The Worst-Case Efficiency of Cost Sharing Methods in Resource Allocation Games*
(4) *Optimal Cost Sharing Protocols for Scheduling Games*
(5) *Characterizing the Existence of Potential Functions in Weighted Congestion Games*
(6) *On the Existence of Pure Nash Equilibria in Weighted Congestion Games*
(7) *Strong Nash Equilibria in Games with the Lexicographical Improvement Property*
(8) *Computing Pure and Strong Nash Equilibria in Bottleneck Congestion Games*
(9) *Approximation Algorithms for Capacitated Location Routing*

The papers form a cross-section through my research in algorithmic game theory and combinatorial optimization. They can be grouped into five topics:

- Stackelberg Routing (Papers 1 and 2)
- Design of Cost Sharing Methods (Papers 3 and 4)
- Existence of Pure Nash Equilibria and Potentials (Papers 5 and 6)
- Existence and Computability of Strong Equilibria (Papers 7 and 8)
- Approximation Algorithms for Location Routing (Paper 9)

In the following, I will outline these topics and the main ideas of the papers.

Note. The only changes I made with respect to the published journal papers concern the unified layout, e.g., renumbering of theorems and minor reformulations necessary for the modified presentation and the updating of some references. The published conference papers appear in this thesis as *full* versions (including all proofs that are sometimes omitted in the originally published conference versions).

1. Stackelberg Routing

It is a well known fact that selfish behavior results in outcomes that are inefficient in general. A prime example is the rush-hour phenomenon observed in urban road traffic. Since every traffic participant solely aims at minimizing her individual travel time, the overall outcome is less efficient, e.g., in terms of the total average travel time, as if everybody would have been routed according to a centrally coordinated routing scheme. With the increasing number of traffic participants, traffic regulation becomes an increasingly important issue.

In this context, the Papers 1 and 2 study the concept of *Stackelberg routing* as a means to reduce the worst-case inefficiency of selfish routing (this measure is also termed the *price of anarchy*). The basic model is based on the classical work of Wardrop modeling the interaction between the selfish network users as a *noncooperative game*. We are given a directed graph with latency functions on the arcs and a set of origin-destination pairs, called *commodities*. Every commodity has a *demand* associated with it, which specifies the amount of flow that needs to be sent from the respective origin to the destination. The demands represent a large population of players, each controlling an infinitesimal small amount of flow of the entire demand (such players are also called *nonatomic*). The latency that a player experiences to traverse an arc is given by a (non-decreasing) function of the total flow on that arc. It is assumed that every player acts selfishly and routes his flow along a minimum-latency path from its origin to the destination; under mild assumptions on the latency functions this corresponds to a common solution concept for noncooperative games, that of a *Nash equilibrium* or *Wardrop flow*. In a Wardrop flow no player can improve his own latency by unilaterally switching to another path.

Because Wardrop flows can be very *inefficient* in the sense that the price of anarchy is unbounded, a prominent approach to reduce this inefficiency is that of *Stackelberg routing*, see the papers for detailed references. In this setting, it is assumed that a fraction $\alpha \in [0, 1]$ of the entire demand is controlled by a central authority, termed *Stackelberg leader*, while the remaining demand is controlled by the selfish nonatomic players, also called the *followers*. In a *Stackelberg game*, the Stackelberg leader first routes the centrally controlled flow according to a predetermined policy, called the *Stackelberg strategy*, and then the remaining demand is routed by the selfish followers. The aim is to devise Stackelberg strategies so as to minimize the worst-case inefficiency of the resulting combined flow with respect to the optimal solution for the entire demand.

As the main result of Paper 1, we show that there exists a family of single-commodity networks parameterized by $k \in \mathbb{N}$ for which every Stackelberg strategy induces a worst-case inefficiency of $\Omega(k)$, where the parameter k represents the size of the network. By increasing the size of the network, we can thus show that the worst-case inefficiency is unbounded. This result holds independently of the fraction $\alpha \in (0, 1)$ of the centrally controlled demand. In Paper 2, I study Stackelberg routing for atomic splittable network games, where players may control a discrete amount of demand. The main results provide bounds on the worst-case inefficiency of Stackelberg routing for restricted network topologies (parallel arcs) and for restricted classes of latency functions (affine latencies).

2. Designing Cost Sharing Methods

In Papers 3 and 4 we study the design of cost sharing methods in resource allocation games as a means to reduce the price of anarchy and the price of stability. Resource allocation games are based on congestion games and have applications in several areas, including traffic networks, telecommunication networks and economics. In a congestion game, there is a set of resources and a pure strategy of a player consists of a subset of resources. The cost of a resource depends only on the number of players choosing the resource, and the private cost of a player is the sum of the costs of the chosen resources. The term *resource allocation games* as used in Paper 3 refers to a variant of congestion games where players have a *variable demand* for the resources and can assign their demand *fractionally* over the set of subsets of the resources. The basic idea now is to design cost sharing rules that distribute the cost of a resource among those players using this resource. The space of feasible cost sharing methods (sometimes called protocols) is defined via certain axioms which in turn are motivated by requirements arising in practice. Informally, these axioms require the existence of pure Nash equilibria of the resulting strategic game, budget balance or cost covering, and separability of the cost sharing method. While the first two properties are self-explanatory, the separability requirement demands that the cost share of a player on a resource only depends on the set of players using this resource. Given such a space of feasible cost sharing methods, the overall goal is to design a cost sharing method that minimizes the resulting price of anarchy and the price of stability.

In Papers 3 and 4 we study cost sharing protocols for two variants of congestion games (resource allocation games in Paper 3 and scheduling games in Paper 4). As the main result of Paper 4, we obtain tight bounds on the achievable price of anarchy and price of stability for the protocol space informally defined above. In Paper 3, we obtain tight bounds on the price of anarchy for two well-known cost sharing methods: *average cost sharing* and *marginal cost pricing*.

3. Existence of Pure Nash Equilibria

While the existence of pure Nash equilibria in classical congestion games follows by the seminal result of Rosenthal, the existence problem becomes much more delicate if each player is associated with a nonnegative weight. In a weighted congestion game, every player has a demand $d_i \in \mathbb{R}_{>0}$ that she places on the chosen resources and the cost of a resource is a function of its total load. In Papers 5 and 6 we consider such weighted congestion games and we ask for the maximal set of cost functions that guarantee (i) the existence of potential functions (see Paper 5 for a formal definition) and (ii) the existence of pure Nash equilibria (see Paper 6). Under mild assumptions on the set of cost functions (continuity), we give in Paper 5 a complete characterization of the existence of exact and weighted potential functions. In Paper 6 we derive a complete characterization of the existence of pure Nash equilibria for twice continuously differentiable cost functions. In that paper, we also study weighted congestion games with restricted strategy spaces such as weighted network congestion games and weighted singleton congestion games.

4. Existence and Computability of Pure and Strong Equilibria

In Papers 7 and 8 we study a class of congestion games from a perspective of cooperative game theory. Specifically, we consider *strong* equilibria (introduced by Aumann in 1959) as solution concept. In a strong equilibrium, no coalition (of any size) can deviate and strictly improve the utility of each of its members. Every strong equilibrium is a pure Nash equilibrium, but the converse does not always hold. Thus, even though strong equilibrium may rarely exist, they form a very robust and appealing stability concept.

The main result of Paper 7 establishes the existence of pure Nash and strong equilibria for so called *bottleneck congestion games*. While in standard congestion games the private cost of a player is given by the *sum* over the costs of the resources in the strategy, in a bottleneck congestion game the private cost function of a player is equal to the cost of the *most expensive* resource that she uses. A prominent application of bottleneck congestion games is routing in computer networks. The throughput of a stream of packets in a communication network is usually determined by the available bandwidth or the capacity of the weakest links. A model that captures this aspect more realistically are in fact bottleneck congestion games in which the individual cost of a player is the maximum (instead of the sum) of the delays in her strategy.

Given the existence of pure Nash and strong equilibria, we study in Paper 8 the complexity of computing pure Nash and strong equilibria. We give a detailed study of the computational complexity of exact and approximate pure Nash and strong equilibria in bottleneck congestion games. We identify three classes of games where it is possible to compute a strong equilibrium in polynomial time: single-commodity networks, branchings, and matroids. For general games, we show that the problem of computing a SE is NP-hard, even in two-commodity networks.

5. Approximation Algorithms for Capacitated Location Routing

The last paper (Paper 9) of this thesis is devoted to designing approximation algorithms for hard combinatorial optimization problems arising in logistics. An *approximation algorithm* for an optimization problem runs in polynomial time for all instances and is guaranteed to deliver solutions with bounded optimality gap. We derive approximation algorithms for different variants of *capacitated location routing*, an important generalization of vehicle routing where the cost of opening the depots from which vehicles operate is taken into account. We also derive results to further generalizations of both problems, including a *prize-collecting* variant, a *group* version, and a variant where *cross-docking* is allowed. We finally present a computational study of our approximation algorithm for capacitated location routing on benchmark instances and large-scale randomly generated instances. Our study is an outcome of a joint research project with the industry partner 4flow AG, Berlin, Germany.

Paper 1

Stackelberg Routing in Arbitrary Networks

Vincenzo Bonifaci, Tobias Harks and Guido Schäfer
Stackelberg Routing in Arbitrary Networks
Math. Oper. Res. **35** (2010), no. 2, pp. 330–346

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

1. Introduction

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

It is well known that Nash equilibria can be *inefficient* in the sense that they need not achieve socially desirable objectives [3, 10]. In the context of network routing games, a Nash flow in general does not minimize the total cost; or said differently, selfish behavior may cause a performance degradation in the network. Koutsoupias and Papadimitriou [22] initiated the investigation of the efficiency loss caused by selfish behavior. They introduced a measure to quantify the inefficiency of Nash equilibria which they termed the *price of anarchy*. The price of anarchy is defined as the worst-case ratio of the cost of a Nash equilibrium over the cost of a system optimum. In recent years, considerable progress has been made in quantifying the degradation in network performance caused by the selfish behavior of noncooperative network users. In a seminal work, Roughgarden and Tardos [32] showed that the price of anarchy for network routing games with nonatomic players and linear latency functions is $4/3$; in particular, this bound holds independently of the underlying network topology. The case of more general families of latency functions has been studied by Roughgarden [27] and Correa, Schulz, and Stier-Moses [6]. (For an overview of these results, we refer to the book by Roughgarden [30].) Despite these bounds for specific classes of latency functions, it is known that the price of anarchy for general latency functions is unbounded even on simple parallel-arc networks [32].

Due to this large efficiency loss, researchers have proposed different approaches to reduce the price of anarchy in network routing games. One of the most prominent approaches is the use of *Stackelberg routing* [21, 29]. In this setting, it is assumed that a fraction $\alpha \in [0, 1]$ of the entire demand is controlled by a central authority, termed *Stackelberg leader*, while the remaining demand is controlled by the selfish nonatomic players, also called the *followers*. In a *Stackelberg game*, the Stackelberg leader first routes the centrally controlled flow according to a predetermined policy, called the *Stackelberg strategy*, and then the remaining demand is routed by the selfish followers. The aim is to devise Stackelberg strategies so as to minimize the price of anarchy of the resulting combined flow with respect to the optimal solution for the entire demand.

Although Roughgarden [29] showed that computing the *best* Stackelberg strategy, i.e., one that minimizes the price of anarchy of the induced flow, is NP-hard even for parallel-arc networks and linear latency functions, several advances have been made recently in proving that Stackelberg routing can indeed significantly reduce the price of anarchy in network routing games. A well-studied Stackelberg strategy is the *Largest-Latency-First (LLF)* strategy. Intuitively, LLF tries to save the part of an optimal flow that is unattractive for the selfish followers by sending flow along paths of large latencies. More precisely, LLF computes an optimal flow for the entire demand and orders the paths that carry a positive amount of flow by non-increasing latencies. According to this order, it then iteratively sends as much flow as possible along these paths (not exceeding the optimal flow value) until an α fraction of the demand has been routed.

Roughgarden [29] showed that for parallel-arc networks the Largest-Latency-First strategy reduces the price of anarchy to $1/\alpha$, *independently* of the latency functions. That is, even if the Stackelberg leader controls only a small constant fraction of the overall demand, the price of anarchy reduces to a constant (while it is unbounded in the absence of any centralized control).

More recently, Swamy [36] obtained a similar result for single-commodity, series-parallel networks and Fotakis [12] for parallel-arc networks and unsplittable flows. Despite these positive results, a central question regarding the effectiveness of Stackelberg routing was still open: Does every routing game admit a Stackelberg strategy inducing a bounded price of anarchy? More precisely, is there a function $g(\cdot)$ such that, for any Stackelberg routing game, there is a Stackelberg strategy inducing a flow with cost at most $g(\alpha)$ times the cost of the optimal flow? This question has been posed explicitly by Roughgarden [26, Open Problem 4].

Besides these efforts, researchers have also tried to characterize the effectiveness of easy-to-implement Stackelberg strategies for specific classes of latency functions. One of the simplest Stackelberg strategies is SCALE (see also [29]), which simply computes an optimal flow for the entire demand and then scales this flow down by α. The currently best known bound for the price of anarchy induced by SCALE on multi-commodity networks and linear latency functions is due to Karakostas and Kolliopoulos [18]. More recently, Swamy [36] derived the first general bounds for polynomial latency functions.

1.1. Our Results

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

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

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

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

(4) We give upper bounds on the efficiency of SCALE for a general class of latency functions which, among others, contains polynomial latency functions with nonnegative coefficients. We also derive the first tight lower bounds for SCALE. Our bound is tight for concave latency functions; for higher degree polynomials our bounds are almost tight (though there remains a small gap for small values of α). Our results also imply that for concave latency functions and general networks SCALE achieves an approximation guarantee of less than 1.12 with respect to the best Stackelberg strategy (which is NP-hard to compute).

1.2. Significance

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

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

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

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

There are numerous applications that can be interpreted as a Stackelberg routing game. Here, we focus on highlighting only one of them: the routing of Internet traffic within the domain of an Internet service provider, see also Sharma and Williamson [33]. Here, the Internet service provider centrally controls a fraction of the overall traffic traversing its domain. In this setting, our second result provides the Internet service provider with an efficient algorithm to route the centrally controlled traffic. The performance of this routing algorithm is characterized by a smooth trade-off curve that scales between the absence of centralized control (doubling the demands is sufficient) and completely centralized control (no scaling is necessary). Additionally, our result has a nice interpretation for the class of (practical relevant) M/M/1-latency functions that model arc-capacities: In order to beat the cost of an optimal flow, it is sufficient to scale all arc capacities by $1 + \sqrt{1-\alpha}$. Our bound is a natural generalization of the bicriteria bound by Roughgarden and Tardos [32] for the entirely selfish setting (see Correa et al. [7] for other related results).

1.3. Techniques

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

We also show that the Largest-Latency-First strategy induces a price of anarchy that is bounded by $O(\alpha^{-1} \cdot b(n, m, k))$, where $b(n, m, k)$ is a function that depends on the number of vertices, arcs and commodities of the network. In order to prove this, we bound the price of anarchy of LLF in terms of the worst-case ratio between the maximum latency that a selfish follower experiences if the followers are routed according to a Nash flow and the maximum latency that a follower experiences if they are routed according to an arbitrary flow. To the best of our knowledge, this relation has not been observed before and might be of independent interest. Our upper bounds then simply follow from existing results characterizing the ratio of the largest latency in a Nash flow and that of a flow that minimizes the maximum latency [8, 24, 28].

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

1.4. Related Work

The idea of using Stackelberg strategies to improve the performance of a system was first proposed by Korilis, Lazar, and Orda [21]. The authors identified necessary and sufficient conditions for the existence of Stackelberg strategies that induce a system optimum; their model differs from the one discussed here. Roughgarden [29] first formulated the problem and model considered here. He also proposed some natural Stackelberg strategies such as SCALE and Largest-Latency-First. For parallel-arc networks he showed that the price of anarchy for LLF is bounded by $4/(3 + \alpha)$ and $1/\alpha$ for linear and arbitrary latency functions, respectively. Both bounds are tight. He also showed that for certain types of Stackelberg strategies, which he termed *weak* strategies (see Section 2 for a definition), the price of anarchy for multi-commodity networks can be unbounded [29]. However, this did not rule out the existence of effective Stackelberg strategies in general. Moreover, he also proved that it is NP-hard to compute the best Stackelberg strategy. Kumar and Marathe [23] investigated approximation schemes to compute the best Stackelberg strategy. The authors gave a a polynomial-time approximation scheme for the case of parallel-arc networks.

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

Correa and Stier-Moses [9] proved, besides some other results, that the use of *opt-restricted strategies*, i.e., strategies in which the Stackelberg leader sends no more flow on every arc than the system optimum, does not increase the price of anarchy. Sharma and Williamson [33] considered the problem of determining the smallest value of α such that the price of anarchy can be improved. They obtained results for parallel-arc networks and linear latency functions. Kaporis and Spirakis [16] studied a related question of finding the minimum demand that the Stackelberg leader needs to control in order to enforce an optimal flow.

Another prominent way to reduce the price of anarchy in nonatomic network routing games is the use of non-negative tolls on arcs of the network. In the area of transportation networks, this concept has been called *congestion toll pricing*, see for example Knight [19], Beckmann et al. [2], Smith [35], and Hearn and Ramana [15]. This mechanism assigns tolls to certain arcs of the network which are charged to those users that decide to take routes through them. If users value latency relative to toll the same, Beckmann et al. [2] showed that charging users the difference between the marginal cost and the real cost in the socially optimal solution (marginal cost pricing) leads to an equilibrium flow which is optimal. Cole et al. [5] considered the case of heterogeneous users, that is, users value latency relative to cost differently. For single-commodity networks, the authors showed the existence of tolls that induce an optimal flow as Nash flow. Finally, Fleischer et al. [11], Karakostas and Kolliopoulos [17], and Yang and Huang [38] proved that there are tolls inducing an optimal flow for heterogenous users even in general networks.

2. Model and Notation

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

Moreover, each arc $a \in A$ has an associated variable *latency* $\ell_a : \mathbb{R}_+ \to \mathbb{R}_+$. For each $a \in A$ the latency function ℓ_a is assumed to be nondecreasing and differentiable. If not indicated

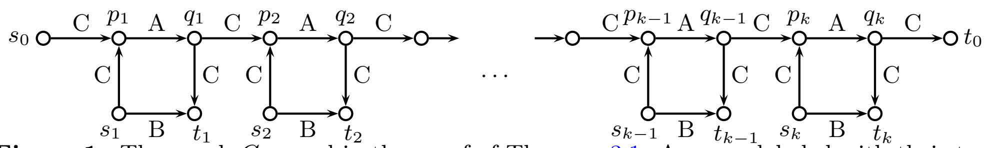


Figure 1. The graph G_k, used in the proof of Theorem 3.1. Arcs are labeled with their type.

otherwise, we assume that $x\ell_a(x)$ is a convex function of x. Such functions are called *standard* [27]. The latency of a path P with respect to a flow f is defined as the sum of the latencies of the arcs in the path, denoted by $\ell_P(f) = \sum_{a\in P} \ell_a(f_a)$. The triple (G, r, ℓ) is called an *instance*.

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

In a Stackelberg network game we are given, in addition to G, r and ℓ, a parameter $\alpha \in (0, 1)$. A *(strong) Stackelberg strategy* [29] is a flow g feasible with respect to $r' = (\alpha_1 r_1, \ldots, \alpha_k r_k)$, for some $\alpha_1, \ldots, \alpha_k \in [0, 1]$ such that $\sum_{i=1}^k \alpha_i r_i = \alpha \sum_{i=1}^k r_i$. If $\alpha_i = \alpha$ for all i, g is called a *weak Stackelberg strategy* [30]. Thus, both strong and weak strategies route a fraction α of the overall traffic, but a strong strategy can choose how much flow of each commodity is centrally controlled. For single-commodity networks the two definitions coincide. A Stackelberg strategy g is called *opt-restricted* if $g_a \leq o_a$ for all $a \in A$.

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

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

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

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

3. Limits of Stackelberg Routing

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

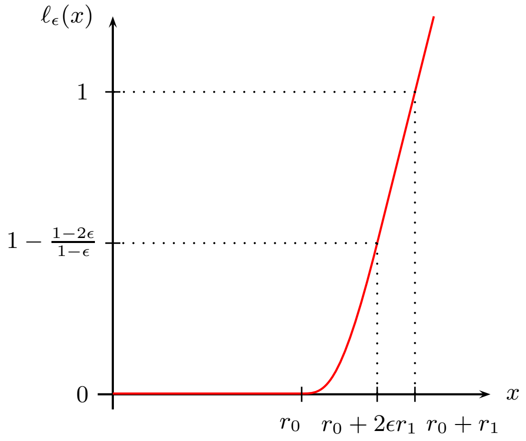


Figure 2. The latency function $\ell_\epsilon(x)$ used in the proof of Theorem 3.1.

3.1. Multi-Commodity Networks

We first show this claim for multi-commodity networks. In this case, such a result was already known to hold for weak Stackelberg strategies [30]; here we prove that it also holds for strong Stackelberg strategies.

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

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

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

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

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

Let us first bound the cost of the optimal flow.

Lemma 3.2. $C(o) \leq 1$.

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

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

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

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

The last inequality follows from $\sum_i g_i \leq \alpha$, and the last equality from $r_1 = (1+\alpha)/2k$. Since $\epsilon < \frac{1-\alpha}{1+\alpha}$, we conclude that $\ell_{P_0}(g+h) = \Omega(k)$. Together with Lemma 3.2, we obtain

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

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

Remark 3.3. We remark that the above proof also works for undirected networks. In these networks, flow can be sent across an edge in both directions and the aggregated flow of an edge is defined as the sum of the flows traversing that edge (in either direction). To see that the lower bound proof still holds, observe that the selfish flow of commodity $i \in [k]$ is still routed along the (s_i, p_i, q_i, t_i) path. The selfish flow sent from s_0 to t_0 now has potentially more paths available than in the directed case. However, it is easy to see that this flow is sent along the $(s_0, p_1, q_1, \ldots, p_k, q_k, t_0)$ path and thus the proof goes through without change. We do not know, however, whether the lower bound for single-commodity networks presented in the next section can be extended to undirected networks.

3.2. Single-Commodity Networks

We use the insights gained in the previous section to prove the following, stronger result:

Theorem 3.4. Let $M > 0$ and $\alpha \in (0,1)$. There is a single-commodity instance $\mathcal{I} = (G, r, \ell, \alpha)$ such that, if g is any strong Stackelberg strategy for $\mathcal{I}$ inducing a Nash flow h, and o is an optimal flow for the instance (G, r, ℓ), then $C(g+h) \geq M \cdot C(o)$.

Theorem 3.4 extends Theorem 3.1 to single-commodity networks. The main idea behind the proof is to simulate the instance used in Theorem 3.1 by creating a supersource s and a supersink t and connecting them to the sources and sinks of the original network (see also Figure 3). If somehow we were able to enforce the s-t flow to split according to the demand vector of the multi-commodity instance, the result would easily follow as in the proof of Theorem 3.1. In order to achieve this, we use latency functions that simulate capacities on the arcs connecting the supersource to the sources and the sinks to the supersink. Although these "capacities" might be exceeded, we will make sure that if the excess flow is too large, the price of anarchy will already be large enough for our purposes.

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

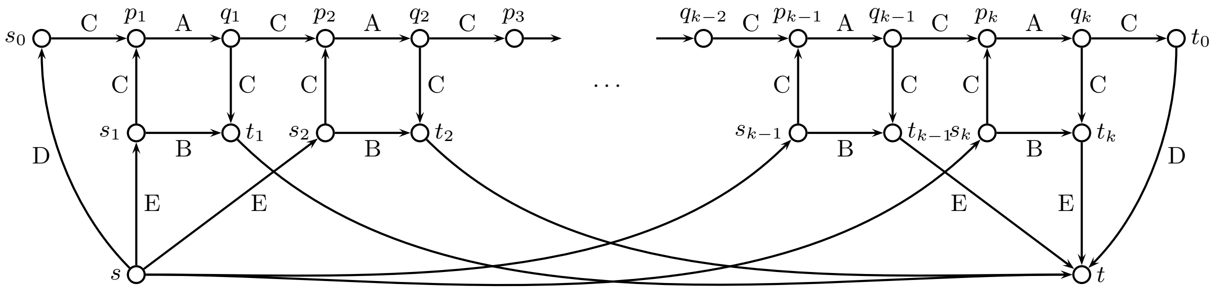


Figure 3. The graph G_k, used in the proof of Theorem 3.4. Arcs are labeled with their type.

of an arc is determined by its type. Type B arcs have constant latency 1, and type C arcs have constant latency 0. Arcs of type A have the following latency function:

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

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

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

Lemma 3.5. $C(o) \le 1$.

Proof. Let P_0 be the path $(s, s_0, p_1, q_1, p_2, \ldots, p_k, q_k, t_0, t)$, and for $i \in [k]$, let P_i be the path (s, s_i, t_i, t). Consider the feasible flow f such that $f_{P_0} = r_0$ and $f_{P_i} = r_1$ for $i \in [k]$. The latency induced by f is 0 on arcs of type A, C, D, E and 1 on arcs of type B. So $C(o) \le C(f) = k \cdot r_1 = (1+\alpha)/2 \le 1$. □

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

Lemma 3.6. For any Stackelberg strategy g inducing a Nash flow h, the following hold:

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

Proof. We prove statement (i); the proof for (ii) is similar. We have $C(g+h) \ge (g_a + h_a) \cdot \ell_a(g_a + h_a) = (g_a + h_a) \cdot u_{r_0,\delta/3k^3}(g_a + h_a) \ge (r_0 + \delta/3k^3) \cdot M/(\delta/3k^3) \ge M$. The proof follows from Lemma 3.5. □

For the remainder of the proof we assume that there is no arc satisfying the conditions of Lemma 3.6; otherwise the theorem follows immediately.

Lemma 3.7. For any Stackelberg strategy g inducing a Nash flow h, the following hold:

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

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

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

For $i = 1$, notice that by Lemma 3.6 the flow along each of $(s, s_1), \ldots, (s, s_k)$ is at most $r_1 + \delta/3k^3$, so the flow on (s, s_0) must be at least $1 - \sum_{i=1}^{k} \left(r_1 + \delta/3k^3\right) = 1 - kr_1 - \delta/3k^2 =$

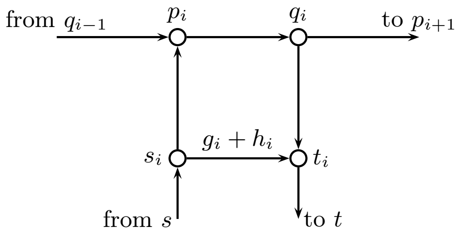


Figure 4. The ith block of the graph G_k.

$r_0 - \delta/3k^2$. But the flow on (s, s_0) is the same as that on arc $(s_0, p_1) = (q_0, p_1)$. Notice that a similar argument allows also to conclude that the flow on each (s, s_i) arc ($i \in [k]$) is at least $r_1 - \delta/3k^2$. This implies (ii) for all $i \in [k]$.

To prove (i) for $i > 1$, consider the ith block in the graph (Figure 4) and let $f = g + h$. By flow conservation, $f_{(q_i,p_{i+1})} = f_{(q_{i-1},p_i)} + f_{(s,s_i)} - f_{(t_i,t)}$. Using induction and Lemma 3.6,

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

□

We are now ready to conclude the proof of Theorem 3.4.

Proof. Theorem 3.4 For any $i \in [k]$, consider the ith block in the graph (Figure 4). Let g_i, h_i be the Stackelberg and selfish flow on the arc (s_i, t_i), respectively. We have two cases:

(1) $h_i = 0$: in this case, using Lemma 3.7, the flow on arc (p_i, q_i) is at least $r_0 - \delta/k + r_1 - \delta/k - g_i$. The latency on that same arc is thus at least $\ell_0(r_0 + r_1 - 2\delta/k - g_i)$.

(2) $h_i > 0$: in this case, the Nash flow on path $P'_i = (s, s_i, t_i, t)$ is strictly positive. Consider the path $P''_i = (s, s_i, p_i, q_i, t_i, t)$. By definition of Nash flow, $\ell_{P''_i}(g + h) \geq \ell_{P'_i}(g + h)$. Notice that the two paths P'_i, P''_i share all their nonzero-latency arcs except for (s_i, t_i) (only present in P'_i) and (p_i, q_i) (only present in P''_i). Thus $\ell_{P''_i}(g + h) \geq \ell_{P'_i}(g + h)$ implies $\ell_{(p_i,q_i)}(g + h) \geq \ell_{(s_i,t_i)}(g + h) = 1$. As a consequence, $\ell_{(p_i,q_i)}(g + h) \geq 1 = \ell_0(r_0 + r_1) \geq \ell_0(r_0 + r_1 - 2\delta/k - g_i)$ since g_i and δ/k are nonnegative.

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

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

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

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

Choosing $\delta < (1 - \alpha)/4$, we can conclude that $\ell_{P_0}(g + h) = \Omega(k)$. Together with Lemma 3.5 and Lemma 3.7, this gives

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

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

Remark 3.8. Suppose the Stackelberg leader is solely interested in minimizing the cost of the flow that he controls, i.e., $C_1(g + h) = \sum_{a \in A} g_a \ell_a(g_a + h_a)$. Our result also implies that even the ratio $C_1(g + h)/C(o)$ can be unbounded, independent of the Stackelberg strategy g.

4. Upper Bounds for LLF

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

We answer this question negatively by proving (for any fixed α) an upper bound on the price of anarchy of $O(b(n, m, k))$, where $b(n, m, k)$ is some function depending on the number of vertices, arcs and commodities of the network. The upper bound holds for a particular Stackelberg strategy, also known as *Largest-Latency-First (LLF)*; see Roughgarden [29] and Swamy [36]. Besides complementing the negative results of the previous section, these are also the first upper bounds for LLF in general networks that hold for *arbitrary* latency functions.

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

Consider the instance $\tilde{\mathcal{I}} = (G, \tilde{r}, \tilde{\ell})$ (as defined in Section 2). Recall that $\tilde{\ell}_a(x) := \ell_a(g_a + x)$ for all $a \in A$. The *maximum latency* of a flow f is defined as $L(f) := \max_{P\in\mathcal{P}: f_P>0} \tilde{\ell}_P(f)$. Let h be a Nash flow, and let $o^{\max}$ denote a flow that minimizes the maximum latency. Then $\rho_L := L(h)/L(o^{\max})$ denotes the worst-case ratio between the maximum latency of a Nash flow and the maximum latency of an arbitrary flow. To prove the upper bound, we bound the price of anarchy induced by LLF in terms of ρ_L. The upper bound will then follow from the previously known fact that ρ_L can be bounded in terms of the network size only [28, 24].

Theorem 4.1. Let $\mathcal{I} = (G, r, \ell, \alpha)$ be a multi-commodity instance with m arcs and let g be the LLF strategy. Then $C(g + h) \leq (m + \frac{1}{\alpha})\rho_L C(o)$.

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

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

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

We further observe that

$$\alpha\Delta_0\, L^g \leq \sum_{P\in\mathcal{P}} g_P \ell_P(o) = \sum_{a\in A} g_a \ell_a(o_a) \leq C(o).$$

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

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

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

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

Combining the bounds for C_1 and C_2 yields

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

As $\rho_L \ge 1$, the theorem is proved. □

Corollary 4.2. Let $\mathcal{I} = (G, r, \ell, \alpha)$ be a single-commodity instance with n vertices and m arcs, and let g be the LLF strategy. Then $C(g+h) \le (n-1)(m + \frac{1}{\alpha})\, C(o)$.

Proof. Roughgarden [28] proves that $\rho_L \le n-1$ for single-commodity instances with n vertices. □

Corollary 4.3. Let $\mathcal{I} = (G, r, \ell, \alpha)$ be a multi-commodity instance with n vertices, m arcs and k commodities, and let g be the LLF strategy. Then $C(g+h) \le b(n,m,k)(m + \frac{1}{\alpha})\, C(o)$, where $b(n,m,k) = 2^{O(\min\{kn, m\log n\})}$.

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

5. A Bicriteria Bound for General Latency Functions

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

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

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

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

Lemma 5.1. If g is the ASCALE strategy, $C(g+h) \leq \sum_{a\in A} \frac{1}{\beta} o_a^\beta \ell_a(g_a+h_a)$.

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

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

Adding $\sum_a g_a \ell_a(g_a+h_a)$ to both sides and using $g = \frac{\alpha}{\beta} o^\beta$ proves the lemma. □

Theorem 5.2. If g is the ASCALE strategy, $C(g+h) \leq \frac{1}{\beta-1} \cdot \left(1-\frac{\alpha}{\beta}\right) \cdot C(o^\beta)$. Furthermore, this bound is tight.

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

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

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

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

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

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

Invoking Lemma 5.1 we get

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

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

Corollary 5.3. Let $\beta = 1+\sqrt{1-\alpha}$. If g is the ASCALE strategy, then $C(g+h) \leq C(o^\beta)$.

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

Theorem 5.4. Let $g = \alpha o$ be the SCALE strategy for instance $\mathcal{I} = (G, r, \ell, \alpha)$. Define a modified instance $\hat{\mathcal{I}} = (G, r, \hat{\ell}, \alpha)$ with latency functions $\hat{\ell}_a(x) = \ell_a(x/\beta)/\beta$ for every arc a, where $\beta = 1+\sqrt{1-\alpha}$, and let $\hat{C}(\cdot)$ denote the cost of a flow with respect $\hat{\ell}$. Let $\hat{h}$ be the Nash flow induced by $\hat{g} = g$ in $\hat{\mathcal{I}}$. Then, $\hat{C}(\hat{g}+\hat{h}) \leq C(o)$.

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

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

for any flow y feasible for $(1-\alpha)r/\beta$. Since $\ell_a(x/\beta)/\beta = \hat{\ell}_a(x)$, we can rewrite (3) as

$$\sum_a (\beta h_a)\hat{\ell}_a(\hat{g}_a + \beta h_a) \le \sum_a (\beta y_a)\hat{\ell}_a(\hat{g}_a + \beta h_a).$$

This implies that βh is a Nash flow induced by $\hat{g}$ in $\hat{\mathcal{I}}$. Since the cost of Nash flows is unique, $\hat{C}(\hat{g}+\beta h) = \hat{C}(\hat{g}+\hat{h})$. Finally, since $\hat{C}(\beta x) = C(x)$ for any flow x, we can conclude $\hat{C}(\hat{g}+\hat{h}) = \hat{C}(\beta(g+h)) = C(g+h) \le C(o)$ where the inequality follows from Corollary 5.3. □

A class of latency functions that are of high practical relevance are so-called *M/M/1 latency functions* (see also [32]). These functions are of the form $\ell_a(x) = 1/(u_a - x)$, where u_a intuitively represents the capacity of arc a. Theorem 5.4 has a particularly nice interpretation in this case: The modified latency functions are $\hat{\ell}_a(x) = \ell_a(x/\beta)/\beta = 1/(\beta(u_a - x/\beta)) = 1/(\beta u_a - x)$. In a purely selfish scenario, Theorem 5.4 therefore implies that to beat optimal routing it is sufficient to double the capacity of every arc. This has been observed before by Roughgarden and Tardos [32]. In the Stackelberg scenario, Theorem 5.4 shows that it is sufficient to increase the capacities by a factor of $1+\sqrt{1-\alpha}$ if the SCALE strategy is used.

6. Bounds for Specific Classes of Latency Functions

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

6.1. λ-Approach

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

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

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

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

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

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

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

Now, the idea is to determine a λ that provides the tightest bound possible. Choosing $\lambda = 1$, the above approach resembles the one that was previously used by Correa, Schulz, and Stier-Moses [6] to bound the price of anarchy of network routing games; however, optimizing over the parameter λ provides an additional means to obtain better bounds. The idea of introducing the scaling parameter λ was first introduced in the context of bounding the price of anarchy in atomic congestion games (see Harks [14]).

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

Definition 6.2. Given an opt-restricted strategy g, the *feasible λ-region* is $\Lambda(g) := \{\lambda \in \mathbb{R}_+ \,|\, \omega(g,\lambda) < 1\}$.

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

Theorem 6.3. Let $\lambda \in \Lambda(g)$. Then $C(g+h) \leq \frac{\lambda}{1-\omega(g,\lambda)} C(o)$.

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

6.2. Bounds for SCALE

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

Definition 6.4. Let $\mathcal{L}_d$ be a class of continuous, nondecreasing, and standard latency functions satisfying

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

$\mathcal{L}_d$ contains, among others, polynomials with nonnegative coefficients and degree at most d. This characterization has been used before by Correa et al. [6].

6.2.1. SCALE: Latency Functions in $\mathcal{L}_1$

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

Lemma 6.5. Assume $\lambda \in [0,1]$ and latency functions in $\mathcal{L}_1$. Then,

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

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

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

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

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

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

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

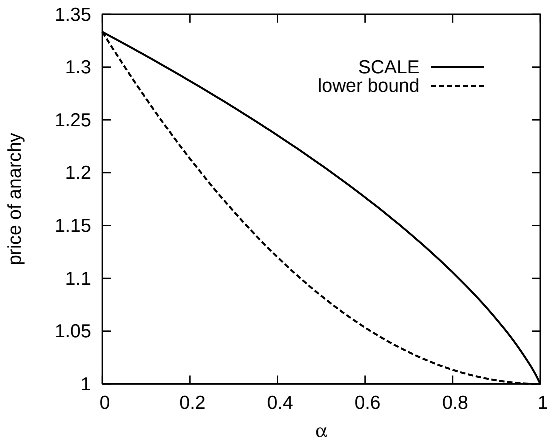
where the first inequality is valid since latencies are nondecreasing. ☐

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

Theorem 6.6. The price of anarchy of the SCALE strategy for latency functions in $\mathcal{L}_1$ is at most

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

Proof. Let $\lambda = \frac{1}{2}(1+\sqrt{1-\alpha})$. Then, by Lemma 6.5, $\omega(\alpha o, \lambda) \leq \frac{1}{2(1+\sqrt{1-\alpha})} < 1$ and thus $\lambda \in \Lambda(\alpha o)$. The proof now follows from Theorem 6.3. ☐



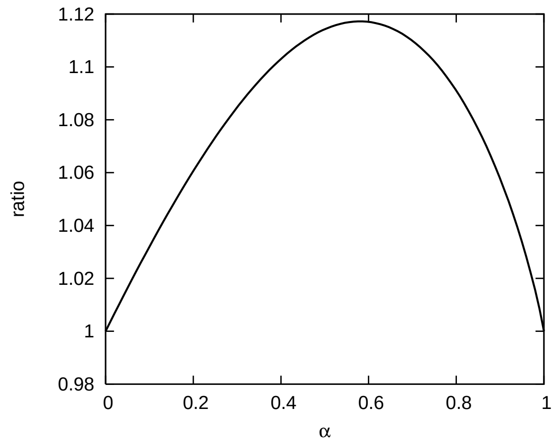


Figure 5. Lower bound for arbitrary Stackelberg strategies vs. upper bound of SCALE for linear latency functions (left) and the respective ratio (right).

Note that the same bound has been proven by Karakostas and Kolliopoulos [18] for the special case of affine latencies. We next present a family of instances that pointwise match the upper bound of Theorem 6.6 for infinitely many values of α. More precisely, the lower bound is matched for all values of α such that $1/\sqrt{1-\alpha}$ is an integer. To the best of our knowledge, this is the first tight bound for values of $\alpha \neq 0, 1$.

Theorem 6.7. Let $n \geq 2$ be an integer and let $c = 1 - (n-1)\alpha/n$. Then, the price of anarchy of the SCALE strategy for latency functions in $\mathcal{L}_1$ is at least

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

Moreover, for all $\alpha = 1 - 1/k^2$, with k a positive integer, there exists an n such that the corresponding bound matches the upper bound of Theorem 6.6.

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

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

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

and the bound follows.

To see that the bound is tight when $\alpha = 1 - 1/k^2$, pick $n = k + 1 = 1 + 1/\sqrt{1-\alpha}$. After substituting the expressions for n and c into the bound and appropriate rewriting we obtain the same expression as in Theorem 6.6. $\square$

We show that there exist instances such that no Stackelberg strategy can achieve a price of anarchy better than $(4 - 2\alpha + \alpha^2)/3$ for linear latency functions. That is, the upper bound on the price of anarchy of SCALE for latency functions in $\mathcal{L}_1$ (Theorem 6.6) is almost best possible (see Figure 5 for a comparison of the lower bound for arbitrary Stackelberg strategies and the upper bound of SCALE).

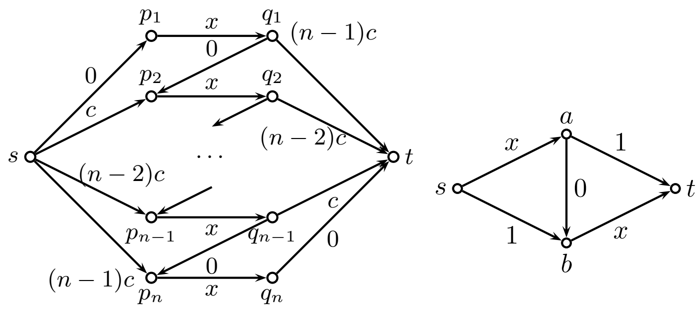


Figure 6. (a) Generalized Braess instance used in the proof of Theorem 6.7. (b) Braess instance. Arcs are labeled with their latency function.

Theorem 6.8. There is an instance $\mathcal{I} = (G, r, \ell, \alpha)$ with linear latency functions such that if g is an arbitrary Stackelberg strategy for $\mathcal{I}$ inducing a Nash flow h, and o is an optimal flow for the instance (G, r, ℓ), then $C(g+h) \geq (4 - 2\alpha + \alpha^2)/3 \cdot C(o)$.

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

We first prove the following lemma:

Lemma 6.9. Let g be an arbitrary Stackelberg strategy. The selfish flow h induced by g then satisfies $h_1 = h_3 = 0$.

Proof. The latency of the zig-zag path is $\ell_2 = g_1 + 2g_2 + g_3 + h_1 + 2h_2 + h_3 = 1 + g_2 + h_2$, where we exploit that $g_3 = \alpha - g_1 - g_2$ and $h_3 = (1-\alpha) - h_1 - h_2$. The latencies of the upper and lower paths are $\ell_1 = g_1 + g_2 + h_1 + h_2 + 1$ and $\ell_3 = 1 + g_2 + g_3 + h_2 + h_3$, respectively. Note that $\ell_1 \geq \ell_2$ and $\ell_3 \geq \ell_2$, independently of the choice of h_2. Since the selfish flow is routed on minimum latency paths, we must have $h_1 = h_3 = 0$ and $h_2 = (1-\alpha)$. $\square$

Proof. Theorem 6.8 The cost of an optimal flow o for the Braess instance is $C(o) = 3/2$. Consider the cost of the combined flow $g + h$. Using Lemma 6.9, we obtain

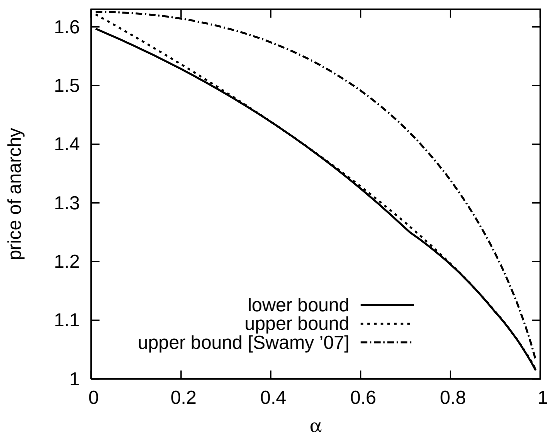
$$\begin{aligned} C(g+h) &= \big(g_1 + g_2 + (1-\alpha)\big)^2 + \big(g_2 + g_3 + (1-\alpha)\big)^2 + g_3 + g_1 \\ &= \big(g_1 + g_2 + (1-\alpha)\big)^2 + \big(1 - g_1\big)^2 + \alpha - g_2. \end{aligned}$$

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

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

$\square$

Since computing the best Stackelberg strategy is NP-hard [29], one may want to devise Stackelberg strategies that are efficiently computable and achieve a good approximation ratio. We say that a Stackelberg strategy g achieves an *approximation ratio of* $c \geq 1$ iff for every instance the cost of the (combined) flow induced by g is at most c times the cost of the (combined) flow induced by any other Stackelberg strategy. In this context, the following corollary follows immediately from Theorem 6.6 and Theorem 6.8.



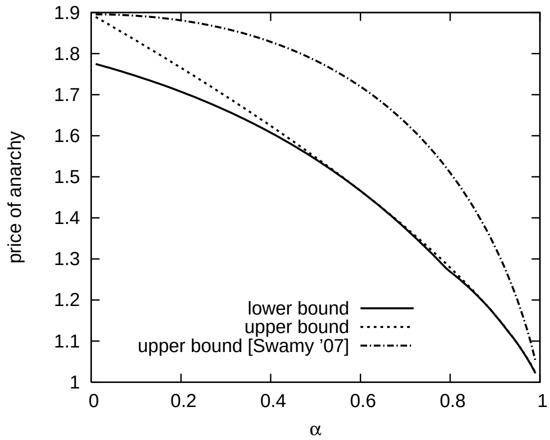


Figure 7. Upper vs. lower bounds for SCALE for polynomial latency functions of degree two (left) and three (right). The plots also show the previously best upper bound by Swamy [36].

Corollary 6.10. The approximation ratio that the SCALE strategy achieves for latency functions in $\mathcal{L}_1$ is at most

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

6.2.2. SCALE: Latency Functions in $\mathcal{L}_d$

Next, we consider the class $\mathcal{L}_d$ of continuous, nondecreasing, and standard latency functions with $d \geq 1$. The proof of the following lemma proceeds along the same lines as the proof of Lemma 6.5.

Lemma 6.11. Assume $\lambda \in [0,1]$ and latency functions in $\mathcal{L}_d$. Then,

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

Proof. The proof proceeds along the same lines as the proof of Lemma 6.5. The only difference is the first part: (i) $\alpha\, o_a + h_a \geq o_a$. As before, we define $\mu := \frac{o_a}{\alpha\, o_a + h_a} \in [0,1]$. We have

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

□

Lemma 6.12. There is a unique $\lambda \in (0,1)$, call it λ_d, such that

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

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

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

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

Theorem 6.13. The price of anarchy of the SCALE strategy for latency functions in the class $\mathcal{L}_d$ is at most

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

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

Proof. We will use Theorem 6.3 with $\lambda = \lambda_d$. However, in order to apply the theorem, we first need to upper bound $\omega(\alpha o, \lambda_d)$. Using Lemma 6.11 and Lemma 6.12, we know that

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

This implies $\lambda_d \in \Lambda(\alpha o)$ and we can invoke Theorem 6.3 to obtain a bound on the price of anarchy given by

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

□

A lower bound for polynomial latency functions of degree d can be obtained by generalizing the construction used in Theorem 6.7. We use again the network of Figure 6(a), except that we replace everywhere the latency function x by x^d and the constant c by $(1 - (n-1)\alpha/n)^d$. The optimal flow is still split evenly on the direct paths, so that with similar arguments we obtain the following lower bound.

Theorem 6.14. Let $n \geq 2$ be an integer and let $c = (1 - (n-1)\alpha/n)^d$. Then, the price of anarchy of the SCALE strategy for latency functions in the class $\mathcal{L}_d$ is at least

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

Notice that the theorem does not fix n, so it is possible to optimize n based on α as in Theorem 6.7. For polynomial latency functions of degree two and three, we compare in Figure 7 the lower bound thus obtained with the upper bound of Theorem 6.13 and also indicate the improvement with respect to the previously best bounds obtained by Swamy [36].

Acknowledgments. Research of the first author partially supported by the Future and Emerging Technologies Unit of EC (IST priority - 6th FP), under contract no. FP6-021235-2 (project ARRIVAL).

References

[1] M. Babaioff, R. Kleinberg, and C. H. Papadimitriou, *Congestion games with malicious players*, in Proc. of the 8th ACM Conf. on Electronic Commerce, 2007, pp. 103–112.

[2] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics and Transportation*, Yale University Press, 1956.

[3] D. Braess, *Über ein Paradoxon der Verkehrsplanung*, Unternehmenforschung **11** (1968), pp. 258–268.

[4] P.-A. Chen and D. Kempe, *Altruism, selfishness, and spite in traffic routing*, in Proc. of the 9th ACM Conf. on Electronic Commerce, 2008, pp. 140–149.

[5] R. Cole, Y. Dodis, and T. Roughgarden, *Pricing network edges for heterogeneous selfish users*, in Proc. of the 35th ACM Symp. on Theory of Computing, 2003, pp. 521–530.

[6] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *Selfish routing in capacitated networks*, Mathematics of Operations Research **29** (2004), pp. 961–976.

[7] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *On the inefficiency of equilibria in congestion games*, in Proc. of the 11th Int. Conf. on Integer Programming and Combinatorial Optimization, 2005, pp. 167–181.

[8] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *Fast, fair, and efficient flows in networks*, Operations Research **55**, no. 2 (2007), pp. 215–225.

[9] J. R. Correa and N. E. Stier-Moses, *Stackelberg routing in atomic network games*, Tech. Report DRO-2007-03, Columbia Business School, February 2007.

[10] P. Dubey, *Inefficiency of Nash equilibria*, Mathematics of Operations Research **11** (1986), pp. 1–8.

[11] L. K. Fleischer, K. Jain, and M. Mahdian, *Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games*, in Proc. of the 45th IEEE Symp. on Foundations of Computer Science, 2004, pp. 277–285.

[12] D. Fotakis, *Stackelberg strategies for atomic congestion games.*, in Proc. of the 15th European Symp. on Algorithms, 2007, pp. 299–310.

[13] G. Gallo, C. Sandi, and C. Sodini, *An algorithm for the min concave cost flow problem*, European Journal of Operational Research **4**, no. 4 (1980), pp. 248–255.

[14] T. Harks, *Stackelberg strategies and collusion in network games with splittable flow*, in Proc. of the 6th Workshop on Approximation and Online Algorithms, 2008, pp. 133–146.

[15] D. W. Hearn and M. V. Ramana, *Solving congestion toll pricing models*, in Equilibrium and Advanced Transportation Modelling, P. Marcotte and S. Nguyen, eds., Springer, 1998, pp. 109–124.

[16] A. C. Kaporis and P. G. Spirakis, *The price of optimum in Stackelberg games on arbitrary single commodity networks and latency functions*, in Proc. of the 18th ACM Symp. on Parallelism in Algorithms and Architectures, 2006, pp. 19–28.

[17] G. Karakostas and S. G. Kolliopoulos, *Edge pricing of multicommodity networks for heterogeneous selfish users*, in Proc. of the 45th IEEE Symp. on Foundations of Computer Science, 2004, pp. 268–276.

[18] G. Karakostas and S. G. Kolliopoulos, *Stackelberg strategies for selfish routing in general multicommodity networks*, Algorithmica **53**, no. 1 (2009), pp. 132–153.

[19] F. H. Knight, *Some fallacies in the interpretation of social cost*, The Quarterly Journal of Economics **38**, no. 4 (1924), pp. 582–606.

[20] D. E. Knuth, *The art of computer programming*, Addison-Wesley, 1997.

[21] Y. A. Korilis, A. A. Lazar, and A. Orda, *Achieving network optima using Stackelberg routing strategies*, IEEE/ACM Transactions on Networking **5**, no. 1 (1997), pp. 161–173.

[22] E. Koutsoupias and C. H. Papadimitriou, *Worst-case equilibria*, in Proc. of the 16th Symp. on Theoretical Aspects of Computer Science, Springer, 1999, pp. 404–413.

[23] V. S. A. Kumar and M. V. Marathe, *Improved results for Stackelberg scheduling strategies.*, in Proc. of the 33rd Int. Colloquium of Automata, Languages and Programming, Springer, 2002, pp. 776–787.

[24] H. Lin, T. Roughgarden, É. Tardos, and A. Walkover, *Braess's paradox, Fibonacci numbers, and exponential inapproximability*, in Proc. of the 32nd Int. Coll. on Automata, Languages and Programming, 2005, pp. 497–512.

[25] G. Perakis, *The "price of anarchy" under nonlinear and asymmetric costs*, Mathematics of Operations Research **32**, no. 3 (2007), pp. 614–628.

[26] T. Roughgarden, *Selfish Routing*, PhD thesis, Cornell University, 2002.

[27] T. Roughgarden, *The price of anarchy is independent of the network topology*, Journal of Computer and System Sciences **67**, no. 2 (2003), pp. 341–364.

[28] T. Roughgarden, *The maximum latency of selfish routing*, in Proc. of the 15th ACM-SIAM Symp. on Discrete Algorithms, 2004, pp. 980–981.

[29] T. Roughgarden, *Stackelberg scheduling strategies.*, SIAM Journal on Computing **33**, no. 2 (2004), pp. 332–350.

[30] T. Roughgarden, *Selfish Routing and the Price of Anarchy*, The MIT Press, 2005.

[31] T. Roughgarden, *On the severity of Braess's paradox: Designing networks for selfish users is hard*, Journal of Computer and System Sciences **72**, no. 5 (2006), pp. 922–953.

[32] T. ROUGHGARDEN AND É. TARDOS, *How bad is selfish routing?*, Journal of the ACM **49**, no. 2 (2002), pp. 236–259.
[33] Y. SHARMA AND D. WILLIAMSON, *Stackelberg thresholds in network routing games or the value of altruism*, in Proc. of the 8th ACM Conf. on Electronic Commerce, 2007, pp. 93–102.
[34] M. J. SMITH, *The existence, uniqueness and stability of traffic equilibria*, Transportation Research **13**, no. 4 (1979), pp. 295–304.
[35] M. J. SMITH, *The marginal cost taxation of a transportation network*, Transportation Research Part B: Methodological **13**, no. 3 (1979), pp. 237–242.
[36] C. SWAMY, *The effectiveness of Stackelberg strategies and tolls for network congestion games*, in Proc. of the 18th ACM-SIAM Symp. on Discrete Algorithms, 2007, pp. 1133–1142.
[37] J. G. WARDROP, *Some theoretical aspects of road traffic research*, Proceedings of the Institute of Civil Engineers **1**, no. Part II (1952), pp. 325–378.
[38] H. YANG AND H.-J. HUANG, *The multi-class, multi-criteria traffic network equilibrium and systems optimum problem*, Transportation Research Part B: Methodological **38**, no. 1 (2004), pp. 1–15.

Paper 2

Stackelberg Strategies and Collusion in Network Games with Splittable Flow

Tobias Harks
Stackelberg Strategies and Collusion in Network Games with Splittable Flow
Theory of Computing Systems **48** (2011), no. 4, pp. 781–802

Abstract. We study the impact of collusion in network games with splittable flow and focus on the well established price of anarchy as a measure of this impact. We first investigate symmetric load balancing games and show that the price of anarchy is at most m, where m denotes the number of coalitions. For general networks, we present an instance showing that the price of anarchy is unbounded, even in the case of two coalitions. If latencies are restricted to polynomials with nonnegative coefficients and bounded degree, we prove upper bounds on the price of anarchy for general networks, which improve upon the current best ones except for affine latencies.

In light of the negative results even for two coalitions, we analyze the effectiveness of Stackelberg strategies as a means to improve the quality of Nash equilibria. In this setting, an α fraction of the entire demand is first routed centrally by a *Stackelberg leader* according to a predefined *Stackelberg strategy* and the remaining demand is then routed selfishly by the coalitions (*followers*).

For a single coalitional follower and parallel arcs, we develop an efficiently computable Stackelberg strategy that reduces the price of anarchy to one. For general networks and a single coalitional follower, we show that a simple strategy, called SCALE, reduces the price of anarchy to $1+\alpha$. Finally, we investigate SCALE for multiple coalitional followers, general networks, and affine latencies. We present the first known upper bound on the price of anarchy in this case. Our bound smoothly varies between 1.5 for $\alpha = 0$ and full efficiency for $\alpha = 1$.

1. Introduction

Over the past years, the impact of the behavior of selfish, uncoordinated users in congested networks has been investigated intensively in the theoretical computer science and operations research literature. In this context, *network routing games* have proved to be a reasonable means of modeling selfish behavior in networks. The basic idea is to model the interaction of selfish network users as a *noncooperative game*. We are given a directed graph with latency functions on the arcs and a set of origin-destination pairs, called *commodities*. Every commodity is associated with a *demand*, which specifies the rate of flow that needs to be sent from the respective origin to the destination. In the nonatomic variant, every demand represents a continuum of agents, each controlling an infinitesimal amount of flow. The latency that an agent experiences to traverse an arc is given by a (non-decreasing) function of the total flow on that arc. Agents are assumed to act selfishly and route their flow along a minimum-latency path from their origin to the destination; a solution in which no agent can switch to a path with smaller travel time corresponds to a Wardrop equilibrium [37, 15].

Koutsoupias and Papadimitriou [24] initiated the investigation of the efficiency loss caused by selfish behavior. They introduced a measure to quantify the inefficiency of Nash equilibria which they termed the *price of anarchy*. The price of anarchy is defined as the worst-case ratio of the cost of a Nash equilibrium over the cost of a system optimum. In a seminal work, Roughgarden and Tardos [34] showed that the price of anarchy for network routing games with nonatomic players and linear latency functions is $4/3$; in particular, this bound holds independently of the underlying network topology. The case of more general families of latency functions has been studied by Roughgarden [29] and Correa, Schulz, and Stier-Moses [12]. (For an overview of these results, we refer to the book by Roughgarden [31].) Despite these bounds for specific classes of latency functions, it is known that the price of anarchy for general latency functions is unbounded even on simple parallel-arc networks [34].

In this paper, we study nonatomic network games in which the agents are partitioned into a (in)finite number of sets, which we interprete (and term) as coalitions of agents. We allow that agents of different commodities may belong to the same coalition and further assume that every coalition aims at minimizing the *average delay* experienced by this coalition. In this setting, we study the worst case efficiency (price of anarchy) of Nash equilibria: stable points, where no coalition can unilaterally improve its cost by rerouting its flow. While the model under consideration (also known as *atomic splittable flow games*) has been studied by many researchers, see among others Cominetti et al. [11], Hayrapetyan et al. [20], Korilis et al.[23], and Roughgarden and Tardos [34], several intriguing open questions still persist.

Cominetti et al. [11] discovered that the price of anarchy in these games may exceed that of corresponding nonatomic games without coalitions. More precisely, Cominetti et al. [11] presented an instance showing that for polynomial latency functions of degree d, the price of anarchy grows as $\Omega(d)$. On the positive side, they presented upper bounds of $1.5, 2.56$, and 7.83, for polynomial latency functions of degree $d = 1, 2, 3$, respectively. For polynomials of larger degree, the previously best known upper bound is $O(2^d\, d^{d+1})$, which is due to Hayrapetyan et al. [20].

1.1. Our Results

We investigate nonatomic network routing games with coalitions. Our contribution in this setting is the following:

(1) First, we consider symmetric load balancing games, that is, we are given parallel arcs that connect a common source and a common sink. For this setting, we show that the price of anarchy is at most m, where m denotes the number of coalitions. This result holds for *arbitrary* convex latencies and is related to a previous result of Cominetti et al. [11], who showed that for single-commodity network games with m coalitions each of which controlling the same amount of flow is at most m. Our result is a generalization in the sense that we do not require that the flow is evenly distributed among coalitions. On the other hand, our result is more restrictive as it only holds for parallel arcs.

(2) We then investigate the efficiency of Nash equilibria for general networks. We show that the price of anarchy in such games is unbounded, even for two coalitions. For semi-convex latency functions, we derive a generic upper bound on the price of anarchy

using a variational inequality approach. We further show that if the class of allowable latencies is restricted to polynomials with nonnegative coefficients and maximum degree d, the price of anarchy is at most $d^{\sqrt{d}}$ for $d \geq 4$. Our bounds improve upon all previous known bounds, except for affine latencies, i.e., $d = 1$. For an overview of our bounds, we refer to Table 1.

Due to the large efficiency loss of Nash equilibria, researchers have proposed different approaches to reduce the price of anarchy in network routing games. One of the most promising approaches is the use of *Stackelberg routing*, see [23, 30]. In this setting, it is assumed that a fraction $\alpha \in [0, 1]$ of the entire demand is controlled by a central authority, termed *Stackelberg leader*, while the remaining demand is controlled by the selfish players, also called the *followers*. In a *Stackelberg game*, the Stackelberg leader first routes the centrally controlled flow according to a predetermined policy, called the *Stackelberg strategy*, and then the remaining demand is routed by the selfish followers. The aim is to devise Stackelberg strategies so as to minimize the price of anarchy of the resulting combined flow.

(3) In light of the negative results that hold even for only two coalitions, we investigate Stackelberg strategies as a way to improve the quality of Nash equilibria. Recently, Bonifaci et al. [6] showed that for nonatomic followers and single commodity networks, no Stackelberg strategy can reduce the price of anarchy to a constant. This result, however, does not rule out the existence of a Stackelberg strategy inducing a constant price of anarchy, when the number of coalitional followers is small. For a single coalitional follower, parallel arcs and semi-convex latencies, we develop an efficiently computable Stackelberg strategy (called SFS) that reduces the price of anarchy to one. For general networks, semi-convex latencies and a single coalitional follower, we prove that the SCALE strategy (see Roughgarden [30]) reduces the price of anarchy to $1 + \alpha$. This result holds for convex latencies and general networks.

(4) Finally, we consider general networks and multiple coalitional followers. For affine linear latencies, we prove that the SCALE strategy yields an upper bound on the price of anarchy, which smoothly varies between the best known bound on the price of anarchy of 1.5 when $\alpha = 0$ and full efficiency when $\alpha = 1$.

1.2. Applications

There are numerous applications that can be interpreted as a network routing game with coalitions. Here, we focus on highlighting only a few (as we find) particularly interesting ones.

In recent years, the number of traffic participants that use a navigation device has increased significantly. Already nowadays, navigation systems feature bidirectional data communication which, among other services, opens the possibility to transmit the current location of a customer to a central server of the service provider (see, e.g., [26]). This way, the current traffic situation can be monitored accurately in real-time (given that a sufficient number of traffic participants are using this technology). Based on this data, the service provider can provide a better route guidance, e.g., in the case of traffic congestion, by centrally computing routes for their customers which are then communicated back to the respective navigation devices. A natural objective that the service provider might want to achieve in order to provide a good quality of service is to minimize the average travel time of their customers. This scenario can be modeled as nonatomic network game with coalitions, where the members of coalition are the customers of a specific service-provider.

One important application of Stackelberg routing is the routing of Internet traffic within the domain of an Internet service provider, see also Sharma and Williamson [35]. Here, the Internet service provider centrally controls a fraction of the overall traffic traversing its domain, while the remaining traffic is controlled by other service providers. In this setting, a natural goal for a service provider is to devise routes to the centrally controlled flow so as to minimize the overall delay in its domain. Our results for the Stackelberg strategies SCALE and SFS provides the Internet service provider with efficient algorithms to compute routes for the centrally controlled traffic.

1.3. Related Work

Awerbuch et al. [3], Christodoulou and Koutsoupias [10], and Aland et al. [1] derived tight bounds on the price of anarchy for weighted and unweighted congestion games with polynomial latency functions. These works, however, did not study the impact of coalitions on the price of anarchy.

Closer to our work are the papers by Hayrapetyan et al. [20] and Cominetti et al. [11]. The former presented a general framework for studying congestion games with colluding players. Their goal is to investigate the price of collusion: the factor by which the quality of Nash equilibria can deteriorate when coalitions form. Their results imply that for symmetric nonatomic load balancing games with coalitions the price of anarchy does not exceed that of the game without coalitions. For weighted congestion games with coalitions and polynomial latencies they proved upper bounds of $O(2^d\, d^{d+1})$, where d denotes the degree of the considered polynomials. They also presented examples showing that in discrete (atomic) network games, the price of collusion may be strictly larger than 1, i.e., coalitions may strictly increase the social cost.

Cominetti et al. [11] studied the atomic splittable selfish routing model in which the flow of every commodity forms a coalition (atomic player). Thus, this model can be incorporated as a special case of nonatomic network congestion games with arbitrary coalitions. They observed that the price of anarchy of this game may exceed that of the standard nonatomic selfish routing game without coalitions. Based on the work of Catoni and Pallotino [9], they presented an instance with affine latency functions in which the price of anarchy is 1.34. Using a variational inequality approach, they presented bounds on the price of anarchy for linear and polynomial latency functions of degree two and three of $1.5, 2.56$, and 7.83, respectively. As noted by Cominetti et al., these positive bounds directly carry over to the case of nonatomic network congestion games with arbitrary coalitions (considered in this paper), since the variational inequalities are still valid in this more general model. For polynomials of larger degree, their approach does not yield bounds. For single commodity networks with symmetric demands (every coalition controls the same amount of flow), Cominetti et al. [11] proved an upper bound of m on the price of anarchy.

Altman et al. [2] proved for monomial latency functions and single commodity networks that there is a Nash flow, which is optimal. They also derived conditions under which Nash equilibria are unique. Uniqueness of Nash equilibria has been further studied by Fleischer et al. [4] and Orda et al. [27].

Haurie and Marcotte [19] presented a general framework for studying atomic splittable network games with *elastic* demands. They characterized the relationship between nonatomic and atomic splittable network games. Haurie and Marcotte, however, do not study the efficiency of Nash equilibria with respect to an optimal solution.

Fotakis, Kontogiannis, and Spirakis [17] studied algorithmic issues in the setting of atomic congestion games with coalitions and unsplittable flows. They proved upper bounds on the price of anarchy, where the cost of a coalition is defined as the maximum latency, see also the KP-model [24].

The idea of using Stackelberg strategies to improve the performance of a system was first proposed by Korilis et al. [23]. The authors identified necessary and sufficient conditions for the existence of Stackelberg strategies that induce a system optimum; their model also considers atomic splittable followers. In particular, they showed that for a single atomic splittable follower, parallel arcs, and $M/M/1$ latencies, there exists an optimal Stackelberg strategy that reduces the price of anarchy to one.

Roughgarden [30] proposed some natural Stackelberg strategies, e.g., SCALE and Largest-Latency-First (LLF). For parallel-arc networks he showed that the price of anarchy for LLF is bounded by $4/(3+\alpha)$ and $1/\alpha$ for linear and arbitrary latency functions, respectively. Both bounds are best possible. Moreover, he also proved that it is NP-hard to compute the best Stackelberg strategy. Kumar and Marathe [25] gave a PTAS to compute the best Stackelberg strategy for the case of parallel-arc networks. Karakostas and Kolliopoulos [22] proved upper bounds on the price of anarchy for SCALE and LLF. Their bounds hold for arbitrary multi-commodity networks and linear latency functions. Swamy [36] obtained upper bounds on the price of anarchy for SCALE and LLF for polynomial latency functions. He also proved a bound of $1+1/\alpha$ for single-commodity, series-parallel networks with arbitrary latency functions. Bonifaci et al. [6] proved that even for single-commodity networks no Stackelberg strategy can induce a bounded price of anarchy for any $\alpha \in (0,1)$. On the positive side, they proved that LLF induces an upper bound on the price of anarchy, which only depends on the size of the network (number of vertices, arcs and commodities). They also derived almost tight bounds for SCALE and polynomial latencies. Correa and Stier-Moses [14] proved, besides some other results, that strategies in which the Stackelberg leader sends no more flow on every edge than the system optimum, does not increase the price of anarchy. Sharma and Williamson [35] considered the problem of determining the smallest value of α such that the price of anarchy can be improved. They obtained results for parallel-arc networks and linear latency functions. Kaporis and Spirakis [21] studied a related question of finding the minimum demand that the Stackelberg leader needs to control in order to enforce an optimal flow. Given that the Stackelberg leader controls a sufficiently large fraction of the overall demand, they also showed that one can efficiently compute the optimal Stackelberg strategy. Finally, Fotakis [16] studied Stackelberg routing with unsplittable flows and proved (among other results) that the $1/\alpha$ bound for parallel links still holds.

2. The Model

In a network routing game, we are given a directed network $G=(V,A)$ and k origin-destination pairs $(s_1,t_1),\dots,(s_k,t_k)$ called *commodities*. We will use the shorthand $[k] := \{1,2,\dots,k\}$. For every commodity $i \in [k]$, a demand $r_i > 0$ is given that specifies the amount of flow with origin s_i and destination t_i. Let $\mathcal{P}_i$ be the set of all paths from s_i to t_i in G and let $\mathcal{P} = \cup_i \mathcal{P}_i$. A *flow* is a function $f : \mathcal{P} \to \mathbb{R}_+$, and we denote by $f_P = f(P)$ the amount of flow that is send along path P. The flow f is *feasible* (with respect to r) if for all i, $\sum_{P\in\mathcal{P}_i} f_P = r_i$.

For a given flow f, we define the flow on an arc $a \in A$ as $f_a = \sum_{P \ni a} f_P$. Moreover, each arc $a \in A$ has an associated load-dependent *latency* denoted by $\ell_a(\cdot)$. For each $a \in A$, the latency function ℓ_a is assumed to be nonnegative, nondecreasing and differentiable. We also assume that ℓ_a is defined on $[0,\infty)$ and that $x\,\ell_a(x)$ is a convex function of x. Such functions are called *semi-convex* or *standard* [29]. The latency of a path P with respect to a flow f is defined as the sum of the latencies of the arcs in the path, denoted by $\ell_P(f) = \sum_{a\in P} \ell_a(f_a)$. The *total cost* of a flow f is $C(f) = \sum_{a\in A} f_a \ell_a(f_a)$. The feasible flow of minimum total cost is called *optimal*. We will denote the optimal flow by o.

In a nonatomic network game, infinitely many agents are carrying the flow rate and each agent controls only an infinitesimal fraction of the demand. The continuum of agents of type j (traveling from s_j to t_j) is represented by the interval $[0,r_j]$. It is well known that for this setting Nash flows exist and their total cost is unique, see [31]. Furthermore, the price of anarchy, which measures the worst case ratio of the total cost of any Nash flow and that of an optimal flow is well understood, see Correa et al. [12, 13], Perakis [28], Roughgarden [31], and Roughgarden and Tardos [34].

In this paper, we study nonatomic network games in which the agents are partitioned into a (in)finite set of coalitions. In our model, we allow that agents of different commodities, i.e., agents traveling from different sources to different destinations, may belong to the same coalition. We assume that the partition of agents into coalitions is *fixed* and given a priori.

Let $[m] = \{1, \ldots, m\}$ denote a set of coalitions. To this end, we represent every agent of commodity i as a real number in $[0, r_i]$. Then, the distribution of agents among the coalitions is modeled by a collection of Lebesgue-measurable functions $c^i : [0, r_i] \to [m]$, $i \in [k]$, which map an agent of type $i \in [k]$ to coalition $j \in [m]$. The continuum of agents of type i belonging to coalition j is defined as the Lebesgue-measure of $\{\xi \in [0, r_i] : c^i(\xi) = j\}$ and denoted by $c^{i,j}$. Using this notation, we define by f^j the flow for coalition j and say that f^j is *feasible for coalition* j if f^j satisfies the demands $c^{i,j}, i \in [k]$ in the usual sense. The amount of flow of coalition j on arc a is defined as $f_a^j = \sum_{P \in \mathcal{P} : P \ni a} f_P^j$, where f_P^j denotes the flow of coalition j along path P.

We assume that every coalition aims at minimizing the *average delay* or *total travel time* experienced by this coalition, see also [11]. Thus, the cost for coalition j is defined as $C^j(f^j; f^{-j}) := \sum_{a \in A} \ell_a(f_a)\, f_a^j$, where f^{-j} denotes the flow of all other coalitions.

The tuple $I = (G, r, \ell, c, m)$ is called an *instance* of the nonatomic network game with coalitions. Our model is similar to the one proposed by Hayrapetyan et al. [20] and it includes the special case, where we have exactly k coalitions each of which controlling the flow for commodity k.

Definition 2.1. A feasible flow f is a Nash equilibrium if and only if for all $j \in [m]$ it holds $C^j(f^j; f^{-j}) \le C^j(x^j; f^{-j})$ for all feasible flows x^j for coalition $j \in [m]$.

In a Nash equilibrium, every coalition routes its flow so as to minimize $C^j(f^j; f^{-j})$ with the understanding that coalition j optimizes over f^j while the flow f^{-j} of all other coalitions is fixed.

Definition 2.2. Let $\mathcal{L}$ be a class of latency functions. Let $\mathcal{I}_m(\mathcal{L})$ be the set of all instances with at most m coalitions and latency functions in $\mathcal{L}$. For $I \in \mathcal{I}_m(\mathcal{L})$, let o_I be an optimal profile and let Θ_I be the set of Nash equilibria, respectively. Then, the *price of anarchy* is defined by

$$\sup_{I \in \mathcal{I}_m(\mathcal{L})} \sup_{f \in \Theta_I} \frac{C(f)}{C(o_I)} .$$

Note that this definition of the price of anarchy is slightly different from the standard nonatomic selfish routing model ([31]), since there may be qualitatively different equilibria, see [4].

If latencies are restricted to be standard, minimizing $C^j(f^j; f^{-j})$ is a convex optimization problem. The following necessary and sufficient optimality conditions characterize Nash flows for a nonatomic network game with coalitions. This characterization can also be found in Haurie and Marcotte [19] (Theorem 2.3) and in Cominetti et al. [11].

Lemma 2.3. A feasible flow $(f^1, \ldots, f^m)$ is a Nash equilibrium for a nonatomic network game with m coalitions if and only if for every $j \in [m]$ the following inequality is satisfied:

$$\sum_{a \in A} \big(\ell_a(f_a) + \ell'_a(f_a)\, f_a^j\big)(f_a^j - x_a^j) \le 0 \text{ for all feasible flows } x^j. \tag{1}$$

Proof. A flow f is a Nash equilibrium if and only if every $f^j, j \in [m]$, is a global minimizer of $C^j(f^j; f^{-j})$. Since the feasible region of all feasible flows for coalition j forms a convex and compact set, and the objective $C^j(f^j; f^{-j})$ is nondecreasing, differentiable and convex, the variational inequality (1) constitutes a first order necessary and sufficient optimality condition for the global minimum of $C^j(\cdot; f^{-j})$ at f^j, see the book by Boyd and Vandenberghe [7]. This condition expresses that at the optimum f^j, there is no feasible gradient descent direction. $\square$

3. Nonatomic Network Games with Coalitions

In the subsequent sections, we will investigate the price of anarchy for specific network topologies and classes of latency functions.

3.1. Symmetric Load Balancing Games

A symmetric load balancing game is a network game, where the underlying digraph simply connects two distinguished nodes with parallel links.

Theorem 3.1. *For symmetric load balancing games with m coalitions and nondecreasing, differentiable, and standard latency functions, the price of anarchy is at most m.*

Proof. As usual, let f denote a Nash flow and o an optimal flow. We bound the cost of each coalition individually. Assume the flow for coalition j carries α_j units of flow. We claim that there exists a feasible flow g^j such that $g^j_a + f^{-j}_a \leq o_a$ for all $a \in A$ with $g^j_a > 0$. To see this, we define the flow $\bar{g} = [o - f^{-j}]^+$, where the positive projection is applied component wise, that is, for arc a we have $[\bar{g}_a]^+ = \bar{g}_a$, if $\bar{g}_a \geq 0$, and 0 otherwise. It is straight-forward to verify that $\bar{g}$ is a feasible flow for $\beta \geq \alpha_j$ units of flow. Hence, the flow $g = \frac{\alpha_j}{\beta}\bar{g}$ is feasible for coalition j. The cost of coalition j when applying strategy g can be bounded by $C^j(g; f^{-j}) = \sum_{a\in A} \ell_a(g_a + f^{-j}_a)\, g_a \leq \sum_{a\in A} \ell_a(o_a)\, g_a \leq \sum_{a\in A} \ell_a(o_a)\, o_a$. The first inequality is valid since for arcs a with $g_a > 0$, we have $\frac{\alpha_j}{\beta}[o_a - f^{-j}_a]^+ + f^{-j}_a \leq o_a$, because $o_a \geq f^{-j}_a$ and $\frac{\alpha_j}{\beta} \leq 1$. The second inequality follows since g is by definition opt-restricted, that is, $g_a \leq o_a$ for all $a \in A$. Using that coalition j plays a best response in equilibrium, we have $C^j(f^j; f^{-j}) \leq C^j(g; f^{-j}) \leq C(o)$. We apply the same argument for every coalition, thus, $C(f) = \sum_{j\in[m]} C^j(f^j; f^{-j}) \leq m\, C(o)$. □

3.2. Multi-commodity Networks

We present the following negative result.

Proposition 3.2. *Let $M > 0$. There is a multi-commodity instance $I = (G, r, \ell, c, m)$ with $m = 2$ such that for a Nash flow f, and an optimal flow o, $C(f) \geq \Omega(M) \cdot C(o)$.*

Proof. Consider the construction in Fig. 1. We have two players, where one player has a demand of size M from s_0 to t_0. The second player has a demand of size 1 from s_1 from t_1. All latencies are constant (1 or 0 as indicated in Fig. 1) except for the latency function $\ell(x)$, which is defined as $\ell(x) = \max\{0, x-M\}$. In a Nash equilibrium, the second player will route 1/2 of its flow along the upper path. Indeed, in this case the marginal latency evaluates to $\ell(1/2+M)+\ell'(1/2+M)\,1/2 = 1$. The total cost of the combined flow f evaluates to $C(f) = 1/2\,(M + 1/2) + 1/2 = \Omega(M)$. A feasible flow can always be constructed by routing the flow of the two commodities along the direct path. Thus, we obtain $C(o) \leq 1$, proving the proposition. □

Note that the function $\ell(x)$ used in the above proposition is not differentiable in $x = M$. But this can be removed by defining a different function $\bar{\ell}(x)$, which smoothly interpolates between $\ell(M) = 0$ and $\ell(1/2 + M)$ and satisfies $\ell(1/2 + M) + \ell'(1/2 + M)\,1/2 = 1$.

3.3. Bounding the Price of Anarchy via the λ-Approach

The previous example showed that for multi-commodity networks, the price of anarchy is unbounded even for two coalitions. In the following, we will therefore restrict the class of allowable latency functions in order to obtain upper bounds on the price of anarchy.

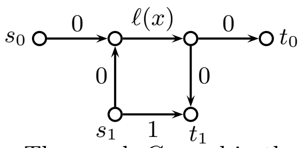


Figure 1. The graph G, used in the proof of Proposition 3.2.

For a latency function ℓ and nonnegative parameter λ we define the following nonnegative value:

$$\omega(\ell; m, \lambda) := \sup_{f,x \geq 0} \frac{(\ell(f) - \lambda\,\ell(x))\,x + \ell'(f)\,\Big(\sum_{j\in[m]} (f^j\,x^j - (f^j)^2)\Big)}{\ell(f)\,f}. \tag{2}$$

Here, we slightly abuse notation and denote by f (under the supremum) the vector $f = (f^1, \dots, f^m)$ and also the sum $f = \sum_{j=1}^m f^j$.

We assume $0/0 = 0$ by convention. For a given class of latency functions $\mathcal{L}$, we define $\omega_m(\mathcal{L}; \lambda) := \sup_{\ell\in\mathcal{L}} \omega(\ell; m, \lambda)$ and $\Lambda_m(\mathcal{L}) := \{\lambda \in \mathbb{R}^+ | \big(1 - \omega_m(\mathcal{L}; \lambda)\big) > 0\}$.

Theorem 3.3. Consider a family of instances $\mathcal{I}_m(\mathcal{L})$, where $\mathcal{L}$ is a class of nondecreasing, differentiable, and standard latency functions. Then, the price of anarchy is at most

$$\inf_{\lambda\in\Lambda_m(\mathcal{L})} \Big[\lambda\,(1 - \omega_m(\mathcal{L}; \lambda))^{-1}\Big].$$

Proof. Let f be a Nash flow, and x be any feasible flow. Then,

$$\begin{aligned} C(f) &\leq \sum_{a\in A} \Big(\ell_a(f_a)\,f_a + \sum_{j\in[m]} \big(\ell_a(f_a) + \ell'_a(f_a)\,f_a^j\big)(x_a^j - f_a^j)\Big) && (3)\\ &= \sum_{a\in A} \Big(\lambda\,\ell_a(x_a)\,x_a + \big(\ell_a(f_a) - \lambda\,\ell_a(x_a)\big)\,x_a + \sum_{j\in[m]} \ell'_a(f_a)\,f_a^j\,(x_a^j - f_a^j)\Big) \\ &\leq \lambda\,C(x) + \omega_m(\mathcal{L}; \lambda)\,C(f). && (4) \end{aligned}$$

Here, (3) follows from the variational inequality stated in Lemma 2.3. The last inequality (4) follows from the definition of $\omega_m(\mathcal{L}; \lambda)$. Taking x as the optimal flow the claim is proven. □

Note that whenever $\Lambda_m(\mathcal{L}) = \emptyset$ or $\Lambda_m(\mathcal{L}) = \{\infty\}$, the approach does not yield a finite price of anarchy. Our definition of $\omega_m(\mathcal{L}; \lambda)$ is closely related to the parameter $\beta^m(\mathcal{L})$ in Cominetti et al. [11] and $\alpha^m(\mathcal{L})$ in Roughgarden [32] for the atomic splittable selfish routing model. For $\lambda = 1$, we have $\beta^m(\mathcal{L}) = \omega_m(\mathcal{L}; 1)$ and $\alpha^m(\mathcal{L}) = (1 - \omega_m(\mathcal{L}; 1))^{-1}$. As we show in the next section, the generalized value $\omega_m(\mathcal{L}; \lambda)$ implies improved bounds for a large class of latency functions, e.g., polynomial latency functions. The previous approaches with $\beta^m(\mathcal{L})$ (or $\alpha^m(\mathcal{L})$) failed for instance to generate upper bounds for polynomials of degree $d \geq 4$ because this value exceeds 1 (or is infinite). The advantage of Theorem 3.3 is that we can tune the parameter λ and, hence, $\omega_m(\mathcal{L}; \lambda)$ so as to minimize the price of anarchy given by $\lambda/(1 - \omega_m(\mathcal{L}; \lambda))$.

We make use of a result of Cominetti et al. [11].

Theorem 3.4 (Cominetti et al. [11])**.** The value $\beta^m(\ell) = \omega(\ell; m, 1)$ is at most

$$\sup_{x,f\geq 0} \frac{\big(\ell(f) - \ell(x)\big)\,x + \ell'(f)\,\big[(x)^2/4 - \big(f - x/2\big)^2/m\big]}{\ell(f)\,f}.$$

Since the necessary calculations to prove the above claim only affect the last term in (2), which is the same for $\omega(\ell; m, \lambda)$ and $\beta^m(\ell)$, this bound carries over for arbitrary nonnegative values of λ.

Corollary 3.5. If $\lambda \geq 0$, the value $\omega(\ell; m, \lambda)$ is at most

$$\sup_{x,f\geq 0} \frac{\big(\ell(f) - \lambda\, \ell(x)\big)\, x + \ell'(f)\, \big[x^2/4 - \big(f - x/2\big)^2/m\big]}{\ell(f)\, f}.$$

3.4. Linear and Affine Linear Latency Functions

Cominetti et al. [11] proved an upper bound of 1.5 for affine latencies. In the following, we present a stronger result for linear latencies. We also show that for affine latencies the best bound can be achieved by setting $\lambda = 1$. In this case, we have $\beta^m(\mathcal{L}) = \omega_m(\mathcal{L}; 1)$.

Theorem 3.6. Consider linear latency functions in $\mathcal{L}_1^* = \{a_1\, z : a_1 \geq 0\}$ and $m \geq 2$ coalitions. Then, the price of anarchy is at most

$$P(m) = \frac{\big(2\, m + \sqrt{2}\, \sqrt{m\, (m+1)}\big)\, \big(m + 1 + \sqrt{2}\, \sqrt{m\, (m+1)}\big)\, \sqrt{2}}{8\, \sqrt{m\, (m+1)}\, (m+1)}.$$

Furthermore, $\lim\limits_{m\to\infty} P(m) = \frac{3}{4} + \frac{1}{2}\sqrt{2} \approx 1.46$.

Proof. For proving the first claim, we start with the bound on $\omega(\ell; m, \lambda)$ given in Corollary 3.5. We define $\mu := \frac{x}{f}$ for $f > 0$ and 0, otherwise, and replace $x = \mu\, f$. This yields $\omega(\ell; m, \lambda) \leq \max_{\mu\geq 0} \big(\mu^2 \big(\frac{m-1-\lambda\, 4\, m}{4\, m}\big) + \mu \big(\frac{m+1}{m}\big) - \frac{1}{m}\big)$. For $\lambda > \frac{m-1}{4\, m}$ this is a strictly convex program with a unique solution given by $\mu^* = \frac{-2\,(m+1)}{m-1-\lambda\, 4\, m}$. Inserting the solution, yields $\omega(\ell; m, \lambda) \leq \frac{m+3-4\,\lambda}{4\,\lambda\, m+1-m}$. The condition $\lambda \in \Lambda_m(\mathcal{L}_1^*)$ is equivalent to $\lambda > \max\Big\{\frac{m-1}{4\, m}, \frac{m-2}{2\, m-2}\Big\}$. We define the value $\lambda = \frac{1}{2} + \frac{1}{4}\sqrt{2\,(m+1)/m}$, which is contained in $\Lambda_m(\mathcal{L}_1^*)$. Applying Theorem 3.3 with this value proves the claim. □

The proof for affine latencies is similar and leads to $C(f) \leq \min_{\lambda\geq 1} \frac{4\,\lambda^2 - \lambda}{4\,\lambda - 2}\, C(x)$ showing that the best bound can be achieved by setting $\lambda = 1$.

3.5. Polynomial Latency Functions

To facilitate the result of Theorem 3.3 for polynomial latency functions, one needs to bound $\omega_m(\mathcal{L}_d; \lambda)$ for the class $\mathcal{L}_d$ of polynomials with nonnegative coefficients and degree at most $d \in \mathbb{N}$:

$$\mathcal{L}_d = \{c_d\, x^d + \cdots + c_1\, x + c_0 \;:\; c_s \geq 0,\; s = 0, \ldots, d\}.$$

Note that polynomials in $\mathcal{L}_d$ are nonnegative for nonnegative arguments, continuous, nondecreasing, and convex.

We focus in the following on the general case $m \in \mathbb{N} \cup \{\infty\}$. Therefore, we define

$$\omega(\ell; \infty, \lambda) := \sup_{x,f\geq 0} \frac{\big(\ell(f) - \lambda\, \ell(x)\big)\, x + \ell'(f)\, (x)^2/4}{\ell(f)\, f}. \tag{5}$$

Then, it follows from Theorem 3.4 that $\omega(\ell; m, \lambda) \leq \omega(\ell; \infty, \lambda)$, since the square is nonnegative and $\lim_{m\to\infty} \big(f - x/2\big)^2/m = 0$.

We now observe that the total cost function $C(f)$ is linear in each of the latency functions $\ell(\cdot)$. We can therefore reduce the analysis to monomial latency functions. For this we subdivide each arc a into d arcs $a_1, \ldots, a_d$ with monomial latency functions $\ell_{a_s}(x) = c_s\, x^s$ for $s = 1, \ldots, d$.

Lemma 3.7. Consider the class $\mathcal{M}_s := \{c_s\, x^s \;:\; c_s \geq 0\}$ for $s \in \mathbb{N}$. Then, $\omega_\infty(\mathcal{M}_s; \lambda) \leq \max_{0\leq\mu} \mu\,\big(1 - \lambda\,\mu^s + s\,\mu/4\big)$.

Proof. Let $\ell \in \mathcal{M}_s$. Then, by (5) we get

$$\omega(\ell; \infty, \lambda) \leq \sup_{x,f\geq 0} \frac{\big(f^s - \lambda\,x^s\big)\,x + s\,f^{s-1}\,x^2/4}{f^{s+1}}.$$

Substituting $x = \mu\,f, \mu \geq 0$, we obtain

$$\omega(\ell; \infty, \lambda) \leq \max_{0\leq\mu} \mu\,\big(1 - \lambda\,\mu^s + s\,\mu/4\big).$$

□

The next Lemma states that $\omega_\infty(\mathcal{M}_s; \lambda)$ is monotonically increasing in s for $\lambda \geq 1$.

Lemma 3.8. For $\lambda \geq 1$, $\omega_\infty(\mathcal{M}_s; \lambda) \leq \omega_\infty(\mathcal{M}_d; \lambda)$ for all $s \leq d, s \in \mathbb{N}, d \in \mathbb{N}$.

Proof. Let $\lambda \geq 1$. By Lemma 3.7, we have for $\ell \in \mathcal{M}_s$

$$\omega(\ell; \infty, \lambda) \leq \max_{0\leq\mu} \mu\,\big(1 - \lambda\,\mu^s + s\,\mu/4\big).$$

It is enough to prove that the argument maximum satisfies $\mu^* \leq 1$. We define $T_s(\mu) := \mu\,\big(1 - \lambda\,\mu^s + s\,\mu/4\big)$ and show that $T_s'(\mu) \leq 0$ for all $\mu \geq 1$. To this end, we obtain

$$\begin{aligned} T_s'(\mu) &= 1 - (s+1)\,\lambda\,\mu^s + (s\,\mu)/2 \\ &= 1 - \mu\,\big((s+1)\,\lambda\,\mu^{s-1} - s/2\big) \\ &\leq 1 - \mu\,\big((s+1)\,\lambda\, - s/2\big) \\ &\leq 1 - \mu\,(s/2 + 1) \\ &\leq 0, \end{aligned}$$

where the first inequality follows from $\mu \geq 1$, while the second inequality follows from $\lambda \geq 1$. □

The next theorem presents an upper bound on the price of anarchy for latencies in $\mathcal{L}_d$.

Theorem 3.9. Consider latency functions in $\mathcal{L}_d$, $d \geq 2$. Then, the price of anarchy is at most $\big(\frac{1}{2}\sqrt{d} + \frac{1}{2}\big)^d\,\frac{\left(d^2+1-\sqrt{d}-d^{\frac{3}{2}}\right)}{(\sqrt{d}-1)\,(d-1)}$.

Proof. We define $\lambda(d) := \big(\frac{1}{2}\sqrt{d} + \frac{1}{2}\big)^d\,\frac{\left(d^2+1-\sqrt{d}-d^{\frac{3}{2}}\right)}{(\sqrt{d}-1)\,(d^2-1)}$.

The proof proceeds by proving a claim, which yields a bound on $\omega(\mathcal{L}_d; \infty, \lambda(d))$.

Claim. $\max_{0\leq\mu\leq 1} \big[T(\mu) := \mu\,\big(1 - \lambda(d)\,\mu^d + d\,\frac{\mu}{4}\big)\big] = d/(d+1),\;$ for all $d \geq 2$.

Proof. To prove the claim it is convenient to write $\lambda(d)$ as

$$\lambda(d) = \frac{d^2 + 1 - \sqrt{d} - d^{\frac{3}{2}}}{\mu_1(d)^d\,(\sqrt{d} - 1)\,(d^2 - 1)},$$

where $\mu_1(d) := 2/(\sqrt{d} + 1)$.

Then, the claim is proven by verifying the following facts:

(1) $T'(\mu_1(d)) = 0$, $T''(\mu_1(d)) < 0$ and $T''(\mu)$ has at most one zero in $(0, 1)$
(2) $T(0) = 0$, $T(1) \leq d/(d+1)$ and $T(\mu_1(d)) = d/(d+1)$.

Before we prove these facts, we show how they imply the claim. The first fact implies that $\mu_1(d)$ is the *only* local maximum of $T(\mu)$ in the open interval $(0,1)$. Then, by comparing $T(\mu_1(d))$ to the boundary values $T(0)$ and $T(1)$ it follows that $T(\mu_1(d)) = d/(d+1)$ is the global maximum.

We start by proving the first fact. The expression $T(\mu_1(d))$ evaluates to:

$$\begin{aligned} T'(\mu_1(d)) &= 1 - (d+1)\,\lambda(d)\,\mu_1(d)^d + d\,\mu_1(d)/2 \\ &= 1 - (d+1)\,\frac{d^2+1-\sqrt{d}-d^{\frac{3}{2}}}{(\sqrt{d}-1)\,(d^2-1)} + d\,\mu_1(d)/2 \\ &= 1 - \frac{d^2+1-\sqrt{d}-d^{\frac{3}{2}}}{(\sqrt{d}-1)\,(d-1)} + \frac{d}{\sqrt{d}+1} \\ &= 0. \end{aligned}$$

We now prove $T''(\mu_1(d)) < 0$. First, we simplify as follows

$$\begin{aligned} T''(\mu_1(d)) &= -d\,(d+1)\,\lambda(d)\,\mu_1(d)^{d-1} + d/2 \\ &= -\frac{d\left(d^2+1-\sqrt{d}-d^{\frac{3}{2}}\right)}{2\,(\sqrt{d}-1)^2} + d/2. \end{aligned}$$

Then, $T''(\mu_1(d)) < 0$ if and only if

$$\frac{d\left(d^2+1-\sqrt{d}-d^{\frac{3}{2}}\right)}{2\,(\sqrt{d}-1)^2} > 1/2 \Leftrightarrow d^2+\sqrt{d}-d^{\frac{3}{2}}-d > 0.$$

The last inequality is fulfilled for all $d \geq 1$.

To verify that $T''(\mu)$ has at most one zero in $(0,1)$, use for example Descartes' rule of signs. The second fact follows by simple calculations. □

The claim implies $\omega_\infty(\mathcal{L}_d; \lambda(d)) \leq d/(d+1)$, hence, $\lambda(d) \in \Lambda_\infty(\mathcal{L}_d)$ so we can use Theorem 3.3 to obtain the claimed bound of $(d+1)\,\lambda(d)$. □

In the following we analyze the growth of the derived upper bound for large d, $(d \geq 4)$. The proof consists of standard calculus and is omitted.

Corollary 3.10. $\left(\frac{1}{2}\sqrt{d}+\frac{1}{2}\right)^d \frac{\left(d^2+1-\sqrt{d}-d^{\frac{3}{2}}\right)}{(\sqrt{d}-1)\,(d-1)} \leq \sqrt{d}^d$ for $d \geq 4$.

In Table 1, we present an overview about achievable upper bounds on the price of anarchy when numerically optimizing over $\lambda \in \Lambda_m(\mathcal{L}_d)$ so as to calculate the minimum in Theorem 3.3.

Table 1. Overview of upper bounds on the price of anarchy for polynomials with nonnegative coefficients and maximum degree d. The result in the first column marked with (*) is with respect to linear latencies $\{a_1\,x \,:\, a_1 \geq 0\}$. The result of the second column (affine latencies) is due to [11].

$d=1^*$	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$	$d=10$
1.46	**1.5**	**2.55**	**5.06**	**11.09**	**26.32**	**66.89**	**180.27**	**512**	**1,524**	**4,734**

4. Stackelberg Strategies with Coalitional Followers

Since the price of anarchy in network games with only two coalitions is already unbounded (Proposition 3.2), we investigate coordination mechanisms as a means to improve the quality of Nash equilibria. One of the most prominent coordination mechanisms in the context of network routing games is the use of *Stackelberg routing*, see Korilis et al. [23] and Roughgarden [30].

In this setting, it is assumed that a fraction $\alpha \in [0,1]$ of the entire demand is controlled by a central authority, termed *Stackelberg leader*, while the remaining demand is controlled by selfish *followers* which in our case are the selfish coalitions. In a *Stackelberg game*, the Stackelberg leader first routes the centrally controlled flow according to a predetermined policy, called the *Stackelberg strategy*, and then the remaining demand is routed by the selfish followers. The aim is to devise Stackelberg strategies so as to minimize the price of anarchy of the resulting combined flow with respect to an optimal solution for the entire demand.

An instance of a Stackelberg routing game with coalitional followers is characterized by a tuple $I(\alpha) = (G, r, \ell, c, m, \alpha)$, where in addition to G, r, ℓ, c and m, a parameter $\alpha \in (0,1)$ is given that specifies the fraction of the demand controlled by the Stackelberg leader.

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

Given a Stackelberg strategy g, let $\tilde{\ell}_a(x) = \ell_a(g_a + x)$ for all $a \in A$ and let $\tilde{r} = r - r'$. We assume that the Stackelberg leader may choose arbitrarily which amount of flow (up to $\alpha\, r$) of a commodity *and* coalition it controls. Thus, the remaining set and demands of the coalitional followers denoted by $\tilde{m}$ and $\tilde{c}$, respectively, is obtained by reducing every $c^{i,j}$ by the amount of demand that the Stackelberg leader wishes to control from coalition j and commodity i.

We say that a flow h is *induced by* g if it is a Nash flow for the instance $(G, \tilde{r}, \tilde{\ell}, \tilde{c}, \tilde{m})$.

A Nash flow h can be characterized by the following *variational inequality* (see Lemma 2.3): h is a Nash flow induced by g if and only if for all flows x feasible with respect to $\tilde{r}$,

$$\sum_{j \in [m]} \sum_{a \in A} \left(\ell_a(g_a + h_a) + \ell'_a(g_a + h_a)\, h_a^j\right) (x_a^j - h_a^j) \geq 0. \tag{6}$$

We will mainly be concerned with the total cost of the combined induced flow $g + h$, given by $C(g+h) = \sum_{a \in A} (g_a + h_a)\ell_a(g_a + h_a)$. In particular, we are interested in bounding the the *price of anarchy*, that is, the worst case ratio of $C(g+h)/C(o)$. It will be convenient to separate the total cost $C(g+h)$ in $C_1(g;h) := \sum_{a \in A} \ell_a(g_a + h_a)\, g_a$ and $C_2(h;g) := \sum_{a \in A} \ell_a(g_a + h_a)\, h_a$.

4.1. Symmetric Load Balancing Games

We consider symmetric load balancing games in which the underlying digraph simply connects two distinguished nodes with parallel links. Let g be a flow according to the Largest-Latency-First (LLF) strategy introduced by Roughgarden [30]. LLF simply calculates an optimal flow o and saturates the arcs with largest latencies first. On the one hand, Roughgarden showed that for Stackelberg routing games with nonatomic followers (without coalitions), LLF reduces the price of anarchy to $1/\alpha$. On the other hand, Hayrapetyan et al. [20] showed that for symmetric load balancing games colluding nonatomic players only decrease the total cost. Combining these two results (Hayrapetyan et al. [20] (Theorem 2.3) and Roughgarden [30] (Theorem 4.2)), it follows that the LLF strategy induces a flow of total cost of at most $1/\alpha\, C(o)$. Thus, the LLF strategy reduces the price of anarchy to $1/\alpha$ even in Stackelberg routing games with coalitional followers.

4.2. Symmetric Load Balancing Games with a Single Follower

We now consider the case of a single follower. This setting has been previously studied by Korilis et al. [23]. The authors showed that for a single coalitional follower, parallel arcs, and $M/M/1$ latencies, there exists an efficiently computable Stackelberg strategy that reduces the price of anarchy to one. Our main result in this section is a generalization of their result to arbitrary semi-convex latencies. We are given an instance $I(\alpha)$ of a Stackelberg game on parallel arcs and a single coalitional follower.

We define a Stackelberg strategy g that we call *single-follower-support* (SFS) strategy as in Algorithm 1.

Algorithm 1 SINGLE-FOLLOWER-SUPPORT

Input: $I(\alpha)$
Output: Stackelberg strategy g

compute a system optimal flow o
$A_1 := \{a \in A : o_a > 0 \text{ and } \ell'_a(o_a) = 0\}$, $A_2 := \{a \in A : o_a > 0 \text{ and } \ell'_a(o_a) > 0\}$

$$x^* := arg \max_{0 \le x_a \le o_a,\, a \in A_1} \sum_{a \in A_1} x_a, \quad \text{s.t.} \quad : \sum_{a \in A_1} x_a \le \alpha$$

$g^1 := x^*$, $g^2 = (g^2_a)_{a \in A_2} := 0$
if $\sum_{a \in A_1} g^1_a < \alpha$ **then**

$$g^2_a := \frac{\alpha - \sum_{a \in A_1} g^1_a}{\ell'_a(o_a) \left(\sum_{\bar{a} \in A_2} \frac{1}{\ell'_{\bar{a}}(o_{\bar{a}})} \right)} \text{ for all } a \in A_2.$$

end if
Return $g = g^1 + g^2$

We prove that this algorithm computes an optimal Stackelberg strategy.

Theorem 4.1. Consider an instance $I(\alpha)$ of a Stackelberg game on parallel arcs with a single coalitional follower. Let g be according to the SFS strategy and let h be an induced Nash flow. Then, the combined flow $g + h$ is optimal.

Proof. First, we consider the case $\sum_{a \in A_1} g^1_a = \alpha$, which implies $g = g^1$.

Since g is opt-restricted, it suffices to prove that the flow $h_a = o_a - g_a$ is a feasible Nash flow for $1 - \alpha$. More precisely, we have to verify that

$$\ell_a(o_a) + \ell'_a(o_a)\,(o_a - g_a) \le \ell_{\hat{a}}(o_{\hat{a}}) + \ell'_{\hat{a}}(o_{\hat{a}})\,(o_{\hat{a}} - g_{\hat{a}}),$$

for all $a, \hat{a} \in A$ with $o_a - g_a > 0$. These inequalities are satisfied since $\ell'_a(o_a) = 0$ for all $a \in A_1$.

Now we consider the case $\sum_{a \in A_1} g^1_a < \alpha$. Notice that in this case $o_a - g_a = 0$ for all $a \in A_1$. Thus, we have to show that

$$\ell_a(o_a) + \ell'_a(o_a)\,(o_a - g_a) = C \text{ for some } C \ge 0 \text{ and all } a \in A_2 \tag{7}$$

$$C \le \ell_{\hat{a}}(o_{\hat{a}}) + \ell'_{\hat{a}}(o_{\hat{a}})\,(o_{\hat{a}} - g_{\hat{a}}) \text{ for all } \hat{a} \in A. \tag{8}$$

We now use that the system optimal flow o satisfies

$$\begin{aligned} \ell_a(o_a) + \ell'_a(o_a)\,(o_a) = \bar{C} \text{ for some } \bar{C} \geq 0 \text{ and all } a \in A_2 \\ \bar{C} \leq \ell_{\hat{a}}(o_{\hat{a}}) + \ell'_{\hat{a}}(o_{\hat{a}})\,(o_{\hat{a}}) \text{ for all } \hat{a} \in A. \end{aligned}$$

Hence, the conditions (7) and (8) are equivalent to

$$\ell'_a(o_a)\, g_a = D \text{ for some } D \geq 0 \text{ and all } a \in A_2.$$

Defining $D = \Big(\alpha - \sum_{a\in A_1} g_a^1\Big)/\Big(\sum_{\bar{a}\in A_2} \frac{1}{\ell'_{\bar{a}}(o_{\bar{a}})}\Big)$ together with $g_a = D/\ell'_a(o_a)$ proves the result. □

4.3. General Networks with a Single Follower

In the following section, we will analyze a simple and easy-to-implement Stackelberg strategy termed *SCALE*. According to the SCALE strategy, a flow g is obtained by computing an optimal flow o and scaling this flow by α, i.e., $g = \alpha\, o$.

We show that SCALE achieves a bound of $(1 + \alpha)$ on the price of anarchy that even holds for general networks and latency functions.

Theorem 4.2. Consider a family of instances $\mathcal{I}(\alpha)$ of Stackelberg games with a single coalitional follower and let g be according to the SCALE strategy. Then, the price of anarchy of the equilibrium flow $g + h$ is at most $1 + \alpha$.

Proof. We bound the cost $C_1(g; h)$ and $C_2(h; g)$ separately. For the follower, we know that $\bar{h} = (1 - \alpha)\, o_a$ is a feasible flow. Since the follower plays a best response in equilibrium, we have $C_2(h; \alpha\, o) \leq C_2((1 - \alpha)\, o; \alpha\, o) = \sum_{a\in A} \ell_a(o_a)\,(1 - \alpha)\, o_a \leq (1 - \alpha)\, C(o)$. Now we bound the cost of the leader. Let h denote the best response of the follower. We consider the following cases. (i) $0 \leq h_a \leq (1-\alpha) o_a$. In this case it follows that $\ell_a(\alpha\, o_a + h_a)\alpha\; o_a \leq \alpha\, \ell_a(o_a)\, o_a$. (ii) $h_a > (1-\alpha)\, o_a$. This case implies $o_a < \frac{1}{1-\alpha}\, h_a$ and we get $\ell_a(\alpha\, o_a + h_a)\alpha\, o_a \leq \frac{\alpha}{1-\alpha}\, \ell_a(\alpha\, o_a + h_a)\, h_a$. Using both cases, we have $C_1(\alpha\, o; h) \leq \alpha\, C(o) + \frac{\alpha}{1-\alpha}\, C_2(h; \alpha\, o) \leq 2\, \alpha\, C(o)$, where the last inequality follows because $C_2(h; \alpha\, o) \leq (1-\alpha)\, C(o)$. Summing both inequalities for C_1 and C_2 proves the claim. □

Based on a simple single-commodity Braess instance [8], one can show that no Stackelberg strategy can induce a price of anarchy of one, even if there is only a single coalitional follower.

4.4. General Networks with Multiple Followers

In this section, we study SCALE for general networks and multiple coalitional followers.

Lemma 4.3. Consider an instance $I(\alpha)$ of a Stackelberg game and let g be according to the SCALE strategy. Then, the following inequality holds:

$$\sum_{k\in[m]} \sum_{a\in A} \Big(\ell_a(\alpha o_a + h_a) + \ell'_a(\alpha o_a + h_a)\, h_a^k\Big)\, (x_a^k - h_a^k) \geq 0,$$

where h is the flow of the followers and x is any feasible flow for the demand $(1 - \alpha)\, r$.

Proof. The lemma follows directly from (6). Taking $x_a := (1 - \alpha)\, o_a$, which is a feasible flow for the remaining $(1 - \alpha)\, r$ demand, we get

$$\sum_{k\in[m]} \sum_{a\in A} \Big(\ell_a(\alpha o_a + h_a) + \ell'_a(\alpha o_a + h_a)\, h_a^k\Big) \Big((1 - \alpha)\, o_a^k - h_a^k\Big) \geq 0.$$

□

For a latency function ℓ and a nonnegative number λ_1, we define the nonnegative value:

$$\omega_1(\ell;\alpha,\lambda_1) := \sup_{o,x\geq 0} \frac{\ell(\alpha\, o+h)\,\alpha\, o - \lambda_1\,\ell(o)\, o}{\ell(\alpha\, o+h)\,(\alpha\, o+h)}. \tag{9}$$

We assume by convention $0/0 = 0$.

For a given class $\mathcal{L}$, we further define $\omega_1(\mathcal{L};\alpha,\lambda_1) := \sup_{\ell\in\mathcal{L}} \omega_1(\ell;\alpha,\lambda_1)$. Similarly,

$$\omega_2(\ell;\alpha,m,\lambda_2) := \sup_{o,h\geq 0} \frac{\big((1-\alpha)\,\ell(\alpha\, o+h) - \lambda_2\,\ell(o)\big)\, o + z(f,h)}{\ell(\alpha\, o+h)\,(\alpha\, o+h)},$$

with $z(f,h) := \ell'(\alpha\, o+h)\,\big(\sum_{k\in[m]} [(1-\alpha)\, h^k\, o^k - (h^k)^2]\big)$. Note that the value $z(f,h)$ is at most $\ell'(\alpha\, o+h)\,\frac{(1-\alpha)\,(o)^2}{4}$. We define $\omega_2(\mathcal{L};\alpha,m,\lambda_2) := \sup_{\ell\in\mathcal{L}} \omega_2(\ell;\alpha,m,\lambda_2)$.

Proposition 4.4. Consider an instance $I(\alpha)$ of a Stackelberg game and let g be according to the SCALE strategy. Then,

$$C_1(g;h) \leq \lambda_1\, C(o) + \omega_1(\mathcal{L};\alpha,\lambda_1)\, C(g+h)$$
$$C_2(h;g) \leq \lambda_2\, C(o) + \omega_2(\mathcal{L};\alpha,m,\lambda_2)\, C(g+h).$$

The proof simply uses Lemma 4.3 and the definitions of ω_1 and ω_2.

Before we state the main theorem, we define

$$\Lambda_m(\mathcal{L};\alpha) := \{(\lambda_1,\lambda_2)\in\mathbb{R}_+^2 \,|\, \big(1-\big(\omega_1(\mathcal{L};\alpha,\lambda_1)+\omega_2(\mathcal{L};\alpha,m,\lambda_2)\big)\big) > 0\}.$$

Note that the set $\Lambda(\mathcal{L};\alpha)$ may be empty.

Theorem 4.5. Consider a family of instances $\mathcal{I}(\alpha)$ of Stackelberg games, where g is defined according to the SCALE strategy. Then, the price of anarchy is at most

$$\inf_{(\lambda_1,\lambda_2)\in\Lambda_m(\mathcal{L};\alpha)} \left[\frac{\lambda_1+\lambda_2}{1-\big(\omega_1(\mathcal{L};\alpha,\lambda_1)+\omega_2(\mathcal{L};\alpha,m,\lambda_2)\big)}\right].$$

The proof uses the previous proposition.

Affine Latency Functions. We will use Theorem 4.5 to prove upper bounds on the price of anarchy for affine latencies. First, we need two technical lemmas.

Lemma 4.6. For $\lambda_1\in\mathbb{R}_+$, $\omega_1(\mathcal{L}_1;\alpha,\lambda_1) \leq \max\left\{\frac{\alpha-\lambda_1}{\alpha}, \frac{\alpha^2}{4\,\lambda_1}\right\}$.

Proof. We start with constant latency functions $\ell(z) = c_0$. By definition of $\omega_1(\mathcal{L};\alpha,\lambda_1)$ we get

$$\omega_1(\mathcal{L};\alpha,\lambda_1) = \sup_{o,h\geq 0} \frac{\alpha\, o\, c_0 - \lambda_1\, o\, c_0}{(\alpha\, o+h)\, c_0} \leq \max\left\{\frac{\alpha-\lambda_1}{\alpha}, 0\right\}.$$

For linear latency functions $\ell(z) = c_1\, z$, we get

$$\begin{aligned}\omega_1(\mathcal{L};\alpha,\lambda_1) &= \sup_{o,h\geq 0} \frac{c_1\,(\alpha\, o+h)\,\alpha\, o - \lambda_1\, c_1\, o^2}{c_1\,(\alpha\, o+h)^2}\\ &= \sup_{o,h\geq 0} \frac{(\alpha\, o+h)\,\alpha\, o - \lambda_1\, o^2}{(\alpha\, o+h)^2}.\end{aligned}$$

We define $\mu := \frac{h}{o}$ if $o > 0$ and zero otherwise. This yields

$$\omega_1(\mathcal{L};\alpha,\lambda_1) \leq \max_{\mu\geq 0} \frac{\alpha^2+\alpha\,\mu-\lambda_1}{(\alpha+\mu)^2} \leq \frac{\alpha^2}{4\,\lambda_1}.$$

Since $\frac{\alpha^2}{4\lambda_1} \geq 0$, we get the claim. □

Lemma 4.7. For $\lambda_2 \geq \frac{1+\alpha-2\,\alpha^2}{4}$, $\omega_2(\mathcal{L}_1;\alpha,m,\lambda_2) \leq \max\left\{\frac{1-\alpha-\lambda_2}{\alpha}, \frac{(1-\alpha)^2}{4\,\lambda_2+\alpha-1}\right\}$.

Proof. We start with constant latency functions $\ell(z) = c_0$. By definition of $\omega_2(\mathcal{L};\alpha,\lambda_2)$ and since $h \geq 0$ we get

$$\omega_2(\mathcal{L};\alpha,m,\lambda_2) = \sup_{o,h\geq 0} \frac{(1-\alpha)\,o\,c_0 - \lambda_2\,o\,c_0}{(\alpha\,o+h)\,c_0} \leq \frac{1-\alpha-\lambda_2}{\alpha}.$$

For linear latency functions $\ell(z) = c_1\,z$, we get

$$\begin{aligned}\omega_2(\mathcal{L};\alpha,m,\lambda_2) &\leq \sup_{o,h\geq 0} \frac{c_1\left(\alpha\,o+h\right)(1-\alpha)\,o - \lambda_2\,c_1\,o^2 + c_1\,\frac{1-\alpha}{4}\,o^2}{c_1\,(\alpha\,o+h)^2}\\ &= \sup_{o,h\geq 0} \frac{\left(\alpha\,o+h\right)(1-\alpha)\,o - \lambda_2\,o^2 + \frac{1-\alpha}{4}\,o^2}{(\alpha\,o+h)^2}.\end{aligned}$$

We define $\mu := \frac{h}{o}$ if $o > 0$ and zero otherwise. This yields

$$\omega_2(\mathcal{L};\alpha,m,\lambda_2) \leq \max_{\mu\geq 0} \frac{(1-\alpha)\,(\alpha+\mu) - \lambda_2 + \frac{1-\alpha}{4}}{(\alpha+\mu)^2} \leq \frac{(\alpha-1)^2}{\alpha+4\,\lambda_2-1},$$

where $\mu^* = \frac{2\,\alpha^2+4\,\lambda_2-1-\alpha}{2\,(1-\alpha)}$ is the optimal solution to the above convex program. Using $\lambda_2 \geq \frac{1+\alpha-2\,\alpha^2}{4}$ we have $\mu^* \geq 0$, which proves the claim. $\square$

Theorem 4.8. Consider a family of instances $\mathcal{I}(\alpha)$ of Stackelberg games such that latency functions are affine. Then, the price of anarchy for the SCALE strategy and m coalitional followers is at most

$$\frac{(1+2\sqrt{1-\alpha})\,(1+\sqrt{1-\alpha})^2}{4+4\sqrt{1-\alpha}-3\,\alpha} \quad \text{for } \alpha \in [0, \frac{1}{2}\sqrt{3}].$$

and

$$\frac{(-3\,\alpha-2\,\alpha\sqrt{1-\alpha}-1+2\,\alpha^2)\,(1+\sqrt{1-\alpha})\,\alpha}{2\,(-3\,\alpha-3\,\alpha\sqrt{1-\alpha}+1+\sqrt{1-\alpha}+\alpha^2)} \quad \text{for } \alpha \in [\frac{1}{2}\sqrt{3}, 1].$$

Proof. We define for $\alpha \in [0, \frac{1}{2}\sqrt{3}]$

$$\lambda_1 = \frac{1}{2}\,(1+\sqrt{1-\alpha})\,\alpha,\ \lambda_2 = \frac{1}{2}\,(1+\sqrt{1-\alpha})\,(1-\alpha).$$

This choice satisfies the conditions:

$$\frac{\alpha-\lambda_1}{\alpha} = \frac{\alpha^2}{4\,\lambda_1},\ \frac{1-\alpha-\lambda_2}{\alpha} \leq \frac{(1-\alpha)^2}{4\,\lambda_2}.$$

Note that for $\alpha \in [0, \frac{1}{2}\sqrt{3}]$ we have $\lambda_2 \geq \frac{1+\alpha-2\,\alpha^2}{4}$ as required in Lemma 4.7. From Lemma 4.6 and Lemma 4.7, we thus obtain

$$\begin{aligned}\omega_1(\mathcal{L}_1;\alpha,\lambda_1) + \omega_2(\mathcal{L}_1;\alpha,m,\lambda_2) &= \frac{1-\lambda_1}{\alpha} + \frac{(1-\alpha)^2}{4\,\lambda_2+\alpha-1}\\ &= \frac{2+2\sqrt{1-\alpha}-\alpha}{2(1+\sqrt{1-\alpha})\,(1+2\sqrt{1-\alpha})}\\ &= \frac{1}{(1+2\sqrt{1-\alpha})} - \frac{\alpha}{2(1+\sqrt{1-\alpha})\,(1+2\sqrt{1-\alpha})}\\ &< 1.\end{aligned}$$

Thus $(\lambda_1, \lambda_2) \in \Lambda_m(\mathcal{L}_1;\alpha)$ and applying Theorem 4.5 proves the first claim.

For $\alpha \in [\frac{1}{2}\sqrt{3}, 1]$ we define

$$\lambda_2 = \frac{1 + \alpha - 2\,\alpha^2}{4}.$$

It is easy to prove that for $\alpha \in [\frac{1}{2}\sqrt{3}, 1]$ we have

$$\frac{1 - \alpha - \lambda_2}{\alpha} \leq \frac{(1-\alpha)^2}{4\,\lambda_2}.$$

Then, it also straightforward to check that $(\lambda_1, \lambda_2) \in \Lambda_m(\mathcal{L}_1; \alpha)$. Hence, applying Theorem 4.5 proves the second claim. □

5. Conclusions and Final Remarks

In the first part of this paper, we investigated the price of anarchy in nonatomic network games with coalitions. On the positive side, we derived an upper bound on the price of anarchy for restricted topologies (load balancing games). For general topologies and (semi-convex) latency functions, we developed a generic upper bound on the price of anarchy which depends on the specific class of allowable latency functions. We note that this bound actually holds for the larger class of congestion games with fractional demand assignments, because the proof technique does not use the network structure, but only uses variational inequalities which remain valid in this more general setting.

After the publication of a preliminary version of this article [18], there has been some work extending our results. Bhaskar et al. [5] showed that the upper bound of m on the price of anarchy for load balancing games (see Theorem 3.1) continues to hold for series-parallel networks. Roughgarden and Schoppmann [33] proved that the generic upper bound of Theorem 3.3 is in fact tight. They also give an exact closed-form expression for the price of anarchy for polynomial latency functions with nonnegative coefficients and bounded degree.

In the second part of this paper, we investigated Stackelberg routing as a means to improve the quality of Nash equilibria. In this setting, we investigated and designed Stackelberg strategies and derived bounds on the price of anarchy for restricted network topologies, number of followers, and classes of latency functions, respectively. Perhaps, the most intriguing open question in this setting is whether there exists a Stackelberg strategy that induces a constant price of anarchy (depending on α) for a finite number of following coalitions. So far, we only understand the extreme cases: for one follower, the answer is yes (Theorem 4.2), while for infinitely many followers, the answer is no, see [6].

References

[1] S. Aland, D. Dumrauf, M. Gairing, B. Monien, and F. Schoppmann, *Exact price of anarchy for polynomial congestion games.*, in Proc. of the 23th Annual Symposium on Theoretical Aspects of Computer Science (STACS), Lecture Notes in Computer Science 3884, Springer, 2006, pp. 218–229.

[2] E. Altman, T. Basar, T. Jimnez, and N. Shimkin, *Competitive routing in networks with polynomial costs*, IEEE Transactions on Automatic Control **47** (2002), pp. 92–96.

[3] B. Awerbuch, Y. Azar, and A. Epstein, *The price of routing unsplittable flow*, in Proc. of the thirty-seventh annual ACM Symposium on Theory of Computing, ACM Press, 2005, pp. 57–66.

[4] U. Bhaskar, L. Fleischer, D. Hoy, and C. Huang, *Equilibria of atomic flow games are not unique*, in Proc. of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2009, pp. 748–757.

[5] U. Bhaskar, L. Fleischer, and C. Huang, *The price of collusion in series-parallel networks*, in Proc. of the 14th International IPCO Conference Integer Programming and Combinatorial Optimization, Springer, 2010. to appear.

[6] V. Bonifaci, T. Harks, and G. Schäfer, *Stackelberg Routing in Arbitrary Networks*, Math. of Oper. Res. **35**, no. 2 (2010), pp. 330–346.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[8] D. Braess, *Über ein Paradoxon aus der Verkehrsplanung*, Unternehmensforschung **12** (1969), pp. 258–268.

[9] S. Catoni and S. Pallotino, *Traffic equilibrium paradoxes*, Transportation Science **25** (1991), pp. 240–244.

[10] G. Christodoulou and E. Koutsoupias, *The price of anarchy of finite congestion games*, in Proc. of the thirty-seventh annual ACM Symposium on Theory of Computing (STOC), ACM Press, 2005, pp. 67–73.

[11] R. Cominetti, J. R. Correa, and N. E. Stier-Moses, *The impact of oligopolistic competition in networks*, Oper. Res. **57**, no. 6 (2009), pp. 1421–1437.

[12] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *Selfish routing in capacitated networks*, Math. Oper. Res. **29** (2004), pp. 961–976.

[13] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *A geometric approach to the price of anarchy in nonatomic congestion games*, Games and Economic Behavior **64**, no. 2 (2008), pp. 457–469.

[14] J. R. Correa and N. E. Stier-Moses, *Stackelberg routing in atomic network games*, tech. report, Columbia Business School, February 2007.

[15] S. Dafermos and F. Sparrow, *The traffic assignment problem for a general network*, J. Res. Natl. Bur. Stand., Sect. B **73** (1969), pp. 91–118.

[16] D. Fotakis, *Stackelberg strategies for atomic congestion games.*, in Proc. of the 15th Annual European Symposium on Algorithms (ESA), Lecture Notes in Computer Science 4698, Springer, 2007, pp. 299–310.

[17] D. Fotakis, S. Kontogiannis, and P. G. Spirakis, *Atomic congestion games among coalitions*, ACM Trans. Algorithms **4**, no. 4 (2008), pp. 1–27.

[18] T. Harks, *Stackelberg strategies and collusion in network games with splittable flow*, in Proc. of the 6th Workshop on Approximation and Online Algorithms (WAOA), Lecture Notes in Computer Science 5426, Springer, 2008.

[19] A. Haurie and P. Marcotte, *On the relationship between nash-cournot and wardrop equilibria*, Networks **15** (1985), pp. 295–308.

[20] A. Hayrapetyan, E. Tardos, and T. Wexler, *The effect of collusion in congestion games*, in Proc. of the thirty-eighth annual ACM Symposium on Theory of Computing (STOC), ACM Press, 2006, pp. 89–98.

[21] A. C. Kaporis and P. G. Spirakis, *The price of optimum in stackelberg games on arbitrary single commodity networks and latency functions*, Theor. Comput. Sci. **410**, no. 8-10 (2009), pp. 745–755.

[22] G. Karakostas and S. G. Kolliopoulos, *Stackelberg strategies for selfish routing in general multicommodity networks*, Algorithmica **53**, no. 1 (2009), pp. 132–153.

[23] Y. A. Korilis, A. A. Lazar, and A. Orda, *Achieving network optima using Stackelberg routing strategies*, IEEE/ACM Transactions on Networking **5**, no. 1 (1997), pp. 161–173.

[24] E. Koutsoupias and C. H. Papadimitriou, *Worst-case equilibria*, in Proc. of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS), Lecture Notes in Computer Science 1563, Springer, 1999, pp. 404–413.

[25] V. S. A. Kumar and M. V. Marathe, *Improved results for Stackelberg scheduling strategies.*, in Proc. of the 33rd International Colloquium of Automata, Languages and Programming (ICALP), Springer, 2002, pp. 776–787.

[26] W. Mossberg, *Dash's car navigation gives smart directions, if others participate.* The Wall Street Journal, March 27, 2008. http://ptech.allthingsd.com/20080327/dashs-car-navigator-gives-smart-directions-if-others-participate.

[27] A. Orda, R. Rom, and N. Shimkin, *Competitive routing in multiuser communication networks*, IEEE/ACM Transactions on Networking **1**, no. 5 (1993), pp. 510–521.

[28] G. Perakis, *The price of anarchy when costs are non-separable and asymmetric*, in Proc. of 10th International Conference on Integer Programming and Combinatorial Optimization, (IPCO), New

York, Lecture Notes in Computer Science 3064, Springer, Berlin Heidelberg, 2004, pp. 46–58.

[29] T. ROUGHGARDEN, *The price of anarchy is independent of the network topology*, Journal of Computer and System Science **67** (2002), pp. 341–364.

[30] T. ROUGHGARDEN, *Stackelberg scheduling strategies.*, SIAM Journal on Computing **33**, no. 2 (2004), pp. 332–350.

[31] T. ROUGHGARDEN, *Selfish Routing and the Price of Anarchy*, The MIT Press, 2005.

[32] T. ROUGHGARDEN, *Selfish routing with atomic players*, in Proc. of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2005, pp. 973–974.

[33] T. ROUGHGARDEN AND F. SCHOPPMANN, *The price of anarchy in atomic splittable congestion games.* unpublished manuscript, 2010.

[34] T. ROUGHGARDEN AND E. TARDOS, *How bad is selfish routing?*, Journal of the ACM **49**, no. 2 (2002), pp. 236–259.

[35] Y. SHARMA AND D. WILLIAMSON, *Stackelberg thresholds in network routing games or the value of altruism*, in Proc. of the 8th ACM conference on Electronic Commerce (EC), ACM Press, 2007, pp. 93–102.

[36] C. SWAMY, *The effectiveness of Stackelberg strategies and tolls for network congestion games*, in Proc. of the 18th annual ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2007, pp. 1133–1142.

[37] J. G. WARDROP, *Some theoretical aspects of road traffic research*, Proc. of the Institute of Civil Engineers **1**, no. Part II (1952), pp. 325–378.

Paper 3

The Worst-Case Efficiency of Cost Sharing Methods in Resource Allocation Games

Tobias Harks and Konstantin Miller
The Worst-Case Efficiency of Cost Sharing Methods in Resource Allocation Games
Operations Research 56(6): 1491–1503 (2011)

Abstract. Resource allocation problems play a key role in many applications, including traffic networks, telecommunication networks and economics. In most applications, the allocation of resources is determined by a finite number of independent players, each optimizing an individual objective function. An important question in all these applications is the degree of suboptimality caused by selfish resource allocation.

We consider the worst-case efficiency of cost sharing methods in resource allocation games in terms of the ratio of the minimum guaranteed surplus of a Nash equilibrium and the maximal surplus. Resource allocation games are closely related to congestion games and model the strategic interaction of players competing over a finite set of congestible resources. Our main technical result is an upper bound on the efficiency loss that depends on the class of allowable cost functions and the class of allowable cost sharing methods. We demonstrate the power of this bound by evaluating the worst-case efficiency loss for three well known cost sharing methods: incremental cost sharing, marginal cost pricing, and average cost sharing.

1. Introduction

Resource allocation problems play a key role in many applications. Whenever a set of resources needs to be matched to a set of demands, the goal is to find the most profitable or least costly allocation of the resources to the demands. Examples of such applications come from a wide range of areas, including traffic networks ([4, 29, 41, 46, 49]), telecommunication networks ([23, 28, 47]), and economics ([30, 31, 32]). In most of the above applications, the allocation of resources is determined by a finite number of independent players, each optimizing an individual objective function. A natural framework for analyzing such non-cooperative games are *congestion games* as introduced by [39]. Congestion games model the interaction of a finite set of strategic players that compete over a finite set of resources. A pure strategy of a player consists of a subset of resources, and the payoff of a player depends only on the number of players choosing the same or overlapping strategies.

An important variant of congestion games are known as *resource allocation games* in which each player assigns a non-negative demand to each of its subsets available. The payoff for a player is defined as the difference between the utility associated with the sum of the demands and the costs associated with the resources used. A prominent example of such a game is the traffic routing game of [17], which builds upon the classical model of [49]: The arcs in a given network represent the resources, the different origin-destination pairs correspond to the players, and the subsets of resources are the paths available in the network for each origin-destination pair. A strategy of a player is a distribution of traffic flow over its avaliable paths. The latency that a player experiences traversing an arc is given by a (non-decreasing) function of the total flow on that arc. The cost for a player on an arc is given by the product of the latency and the player's flow contribution on the arc.

Resource allocation games also play a key role in telecommunication networks, where users want to route packets from their source node to some sink node in the network. In this type of application it is frequently assumed that each user receives a non-negative utility from transmitting at a certain packet rate and that each link (resource) determines a congestion-dependent price per unit flow that is charged to its users, see [28] and [47]. In [28] it is assumed that every link has a total cost function (modeling total delay or packet loss) and the price per unit flow is defined by the marginal cost function.

The above two examples can be cast in the light of cost sharing methods: every resource incurs a cost that is passed on to its users by charging every user a cost share. In the terminology of the cost sharing literature, the prevailing cost sharing method in transportation networks is *average cost sharing*, because the cost of a resource is the total delay, while every user pays the product of the current latency and its flow contribution. In telecommunication networks (see [28]), every user is charged the *marginal cost* per unit of resource which corresponds to *marginal cost pricing*. Note that in both cases the cost sharing method charges a *single price per unit of resource*. This property is considered desirable and indispensable for large scale networks, because every resource only needs to pass a one-dimensional information to its users, see also the motivation given in [25], [28] and [47].

An important question in all these areas is the degree of suboptimality caused by selfish resource allocation. Since this suboptimality crucially depends on the specific cost sharing method used we first have to define the design space of cost sharing methods. To this end, we define the following five properties listed below which are defined more formally in Section 3:

(1) *Separability*: The cost sharing method of a resource is a function only of the consumption of the considered resource.
(2) *Cost-covering*: The cost of a resource is covered by the cost shares collected from the users.
(3) *No charge for zero demand*: The cost share for every player is zero on resources not used by her.
(4) *Nash-inducing*: The cost sharing method is a non-negative, non-decreasing, differentiable and convex function in the resource consumption of every player.
(5) *Scalability*: The cost sharing method charges a single price per unit of resource.

We briefly discuss the above five requirements. The first assumption requires that the cost share of a resource only depends on the vector of its consumption by the players. This implies that the cost shares of a resource are independent of the usage of other resources and, thus, precludes any coordination between different resources. While this property seems restrictive, it is crucial for practical applications in which cost sharing methods have only local information about their own usage (see for instance the TCP/IP protocol design, where routers drop packets based on some function of the number of packets in the queue, see [47]). Assumptions (2) and (3) are standard in the economics literature and the least controversial. The fourth assumption gives a sufficient condition on the existence of a pure Nash equilibrium of the induced resource allocation game and is frequently used in the economics literature, see [31]. The last requirement, certainly the most restrictive one, is motivated by focusing on cost sharing methods that are applicable in the context of large scale networks, see the discussion above. In the following, we will call a cost sharing method *basic*, if it satisfies assumptions (1)-(4). We will call a cost sharing method *scalable*, if it satisfies assumptions (1)-(5).

1.1. Our Results

We study the efficiency loss of Nash equilibria in the context of resource allocation games with basic and scalable cost sharing methods. Given a class of cost functions $\mathcal{C}$ and a class of basic cost sharing methods $\mathcal{D}_n$ for n players, we develop a general lower bound on the worst-case efficiency of Nash equilibria that only depends on $\mathcal{C}$ and $\mathcal{D}_n$ but not on the player's private utilities. We show that among all basic cost sharing mechanisms, there is an optimal mechanism (incremental cost sharing) that achieves full efficiency. Because the incremental cost sharing method is not scalable, we analyze the worst-case efficiency of two well known scalable cost sharing methods: *marginal cost pricing* and *average cost sharing*. By applying our generic lower bound to marginal cost pricing and average cost sharing we obtain the following results that are summarized below.

Results for Marginal Cost Pricing. For differentiable, non-decreasing and convex marginal cost functions, we prove a lower bound of $\frac{4}{3+\sqrt{5+4\,n}}$ on the worst-case efficiency. In particular, this bound carries over to practically relevant M/M/1 functions that model queuing delays with arc-capacities. We complement this bound by presenting an asymptotically matching upper bound of $\frac{2(n-\sqrt{n})}{\sqrt{n}(n-1)}$ leaving only a gap for small n. We completely characterize the worst-case efficiency for polynomial cost functions with non-negative coefficients (previous results only covered affine marginal costs). For symmetric games (players have equal utility functions and equal strategy space), we present a series of results showing that the worst-case efficiency of Nash equilibria significantly improves. In particular, we prove a lower bound of $2\,n/(2\,n+1)$ for differentiable, non-decreasing and convex marginal cost functions. For polynomial cost functions with non-negative coefficients we prove a tight bound of $3/4$.

Results for Average Cost Sharing. For differentiable, non-decreasing and convex cost functions, we prove a lower bound of $1/n$ on the worst-case efficiency. If we further assume that the average cost functions are convex (e.g., polynomials with non-negative coefficients) we present a tight bound of $4/(n+3)$. For symmetric games this bound improves to $4\,n/(n+1)^2$.

1.2. Significance and Techniques Used

Our main technical contribution is a general template to derive an upper bound on the efficiency loss of basic cost sharing methods in resource allocation games. This generality stems from two aspects: On the one hand side the restriction to *basic* cost sharing methods requires only mild assumptions on the feasible design space, see also the discussion in [31]. On the other hand, our template works for general resource allocation games including the single resource case as in [31] as well as multi-commodity network variants considered in [23]. We see this as a non-trivial generalization of previous works, as, for instance, in [23] the network structure is explicitely used to prove bounds on the price of anarchy (essentially through max-flow computations).

Our proof technique is quite simple and different from [23] and [31]. In [23], the authors consider marginal cost pricing and explicitely identify the worst possible game by analytically solving a sequence of quadratic optimization problems (assuming linear marginal cost functions).

The resulting optimization problem explicitly involves the coefficients a and b of an affine marginal cost function $c(x) = a\,x + b$. Hence, this approach becomes increasingly technical if this optimization problem involves, e.g., polynomial cost functions of higher degree. For general convex marginal cost functions it is not clear whether the approach of [23] gives an optimization problem that is structured enough to be solved.

[31] derives lower bounds on the worst-case efficiency (using a different measure of efficiency) of three cost sharing methods: *average cost sharing, serial cost sharing and incremental cost sharing*. His bounds are valid for resource allocation games with a single resource. Clearly, this assumption simplifies the subsequent analysis. From a technical point of view, Moulin proves an upper bound on the efficiency loss for each of the three cost sharing methods separately. Our approach gives a unified bound on the efficiency loss for an entire class of cost sharing methods including those considered in [31] and [23].

Key to our approach is the use of variational inequalities, which allow to relate the surplus of a Nash equilibrium to that of an optimal profile. Because variational inequalities do not rely on the specific combinatorial structure of the strategy spaces, this approach is applicable to general resource allocation games, which contain games with network structure as a special case. We note here that variational inequalities have been used before for bounding the efficiency loss of Nash equilibria, see [11], [12], [40] and [50].

1.3. Outline

The remainder of this paper is structured as follows. After reviewing the related work in Section 2 we introduce in Section 3 the fundamentals of a resource allocation game consisting of a congestion model and a cost sharing method. For the class of basic cost sharing methods, we develop in Section 4 a general lower bound on the worst-case efficiency of Nash equilibria that only depends on the used cost functions and cost sharing methods but not on the player's private utilities. We use this general bound to show that the incremental cost sharing method is optimal. Since the incremental cost sharing method is basic but not scalable, we focus in the rest of the paper on two scalable cost sharing methods: marginal cost pricing and average cost sharing. In Section 5, we apply our general lower bound to marginal cost pricing and derive several lower and upper bounds on the worst-case efficiency of Nash equilibria depending on the used cost functions. In Section 6, we subsequently apply our generic bound to average cost sharing. We conclude the paper in Section 7 with a brief summary of our results and a discussion of open problems. Appendix 8 provides a table of notation. All missing proofs can be found in the e-companion to this paper.

2. Related Work

Network Resource Allocation Games. [28] and [27] studied network resource allocation games and proposed a pricing mechanism termed proportionally fair pricing in which every resource charges a price per unit resource equal to *marginal cost*. Despite the simplicity and scalability of this mechanism, Kelly et al. showed that an optimal solution can be achieved as an equilibrium if players are *price takers*, that is, if they do not anticipate the consequence of price change in response to a change of their flow.

[21] and [24] studied network resource allocation games, where players submit a bid to each resource in the network and resources are allocated to the players according to Kelly's proportionally fair allocation mechanism. For this mechanism they established a bounded efficiency loss of the marginal pricing scheme with fixed and elastic resource capacities. However, the proposed mechanism is not scalable since each player has to submit an individual bid to each resource. If,

instead, players can only submit a single bid per path, it was proven that the efficiency can be arbitrary low for the case of hard capacities by [54] and for the case of elastic capacities by [20].

[22] and [23] studied network resource allocation games with marginal cost pricing. On the negative side, they showed that for non-differentiable marginal cost functions, the price of anarchy is unbounded even for games with two players. For the special case of linear marginal cost functions, [23] showed that the efficiency loss is bounded by $2/3$. Remarkably, this result holds for an arbitrary collection of concave utility functions and arbitrary networks. For a game with one resource and n players having equal utility functions, [22] proved a bound of $2\,n/(2\,n{+}1)$ for convex marginal cost functions.

[9] recently presented a class of pricing mechanisms for network resource allocation games satisfying four axioms that are considered desirable. In particular, their mechanisms are characterized by the axioms *rescaling, additivity, positivity, and weak consistency*, which have been proposed by [44]. This family of price mechanisms includes marginal cost pricing, Aumann-Shapley pricing, and average cost pricing. The main objective of Chen and Zhang is to find among all mechanisms that satisfy the four axioms an *optimal mechanism*, i.e., one that minimizes the induced price of anarchy. Their main result states that for affine cost functions, the optimal mechanism is obtained by an affine transformation of marginal cost prices, and that marginal cost pricing itself is nearly optimal (achieving a slightly better efficiency guarantee (0.686) than the bound (2/3)).

Cost Sharing in Cournot Games. Cournot's oligopoly model clearly is one of the cornerstones of economic theory, see [30, 34] for an overview of classical work in this area. [22] proved that Cournot oligopoly games are basically equivalent (in terms of the worst-case efficiency of Nash equilibria) to resource allocation games with a single resource (which are termed Cournot oligopsonies in [22]). [31] studied the price of anarchy for resource allocation games on a single resource with three different pricing mechanisms: average cost sharing, incremental cost sharing, and serial cost sharing. An important difference between our approach and that of Moulin is the definition of the efficiency loss of a cost-sharing method. The total surplus of a Nash equilibrium in [31] is defined as the sum of the player's payoffs, which inevitably involve the cost shares collected. In our model (and that of [22, 23]), we assume that the collected cost shares are internalized, so that we only count the player's utilities for using the resources minus the *actual* cost of using the resources. Only for exactly balanced cost sharing methods (such as average cost sharing) this difference vanishes as the actual costs and the collected cost shares coincide. In fact, it turns out that two of our results for average cost sharing (the bound $1/n$ in Theorem 6.1 and $4/(3+n)$ in Theorem 6.2) coincide with Moulins bounds.

[14] studied Cournot oligopoly models and derived bounds on the price of anarchy for marginal cost pricing. In a Cournot oligopoly game, there is a set of players each producing quantities so as to satisfy an elastic demand. The production cost for every player is modeled by a convex cost function and the market price is modeled by a decreasing function in the total supplied quantity. The goal of every player is to maximize revenue. [14] derived among other results a lower bound of the worst-case efficiency of $4/(\sqrt{4\,n+5}+3)$ for concave marginal price functions. Using the equivalence result of [22], this bound translates to the case of resource allocations games with a single resource, marginal cost pricing and convex marginal cost functions. We show in Theorem 5.3 that the same bound even holds for general resource allocation games.

Nonatomic Network Routing. In nonatomic network routing games, [43] showed that the price of anarchy for network routing games with nonatomic players and linear latency functions is $4/3$. The case of more general families of latency functions has been studied by [40] and [12]. (For an overview of related results, we refer to the book by [41] and the survey by [3].) Despite these bounds for specific classes of latency functions, it is known that the price of anarchy in routing games with general latency functions is unbounded even on simple parallel-arc networks

([43]). [8] studied the price of anarchy for nonatomic network games with elastic demands and general cost functions. They obtain bounds for the more general case of separable cost functions and elastic demands. The case of asymmetric cost functions has been studied by [37]. [52] presented a detailed study on the worst-case efficiency loss of different variants of marginal cost pricing for the case of non-atomic users with fixed and elastic demands, respectively.

Atomic Splittable Network Routing. In atomic splittable network routing games there is a finite number of players who can split the flow along available paths, see [2], [11], [16], [17], [18], [50]. Haurie and Marcotte presented a general framework for studying atomic splittable network games with elastic demands. Haurie and Marcotte, however, do not study the efficiency of Nash equilibria with respect to an optimal solution. Along similar lines as [17], [15] considers games with atomic players and nonatomic players at the same time. Harker referred to the equilibria of those games as mixed behavior equilibria, and gave a characterization of these equilibria by means of variational inequalities.

[18] studied congestion games with colluding players. Their goal is to investigate the price of collusion: the factor by which the quality of Nash equilibria can deteriorate when coalitions form. [2] and [11] studied the atomic splittable selfish routing model. Altman et al. bounded the price of anarchy for monomial latency functions (plus a constant). They also derived conditions under which a Nash equilibrium is unique. Uniqueness of Nash equilibria has been further studied by [5], [33] and [53]. Cominetti et al. observed that the price of anarchy of the atomic splittable game may exceed that of the standard nonatomic selfish routing game. Based on the work of [7], they presented an instance with affine latency functions, where the price of anarchy is 1.34. For affine latencies, they presented an upper bound of 1.5 on the price of anarchy. In [16] a general upper bound on the price of anarchy is derived that depends on the class of latency functions. This bound is tight as shown in [42].

An important difference between our model and that of [18] and [11] is that our model involves *elastic* demands that are varied by players. As a result, in our model the payoff of players is a linear combination of utility (derived from sending flow) and associated costs.

Tolls in Network Games. A large body of work in the area of transportation networks is concerned with congestion toll pricing, see for example [29], [4], [46], and [19]. This mechanism assigns tolls to certain arcs of the network which are charged to those users that decide to take routes through them. The toll mechanism has the desirable property that every user is charged a single price per unit resource.

[4] showed that for the Wardrop model with homogeneous users charging the difference between the marginal cost and the real cost in the socially optimal solution (marginal cost pricing) leads to an equilibrium flow which is optimal. [10] considered the case of heterogeneous users, that is, users value latency relative to monetary cost differently. For single-commodity networks, the authors showed the existence of tolls that induce an optimal flow as Nash flow. [13], [26], and [51] proved that there are tolls inducing an optimal flow for heterogenous users even in general networks. [48] and [53] proved the existence of optimal tolls for the atomic splittable model with fixed demands. Note that for computing the corresponding tolls, the works by [10, 13, 26, 48, 51, 53] use a mathematical programming approach which requires central knowledge about the users including their locations, private utility functions and demands. In this sense, the toll mechanisms are not scalable, because the underlying cost sharing method is a function of these private values.

Finally, [1] and [35] study a model of parallel arc networks in which the arcs are owned by service providers that compete for the available traffic by setting prices. For this model they prove a tight worst-case bound for the efficiency loss of equilibria.

3. The Model

In this section, we introduce resource allocation games as natural generalizations of variants of congestion games. As the two building-blocks of a resource allocation game, we first define a congestion model and then introduce the notion of a cost sharing method.

3.1. Congestion Model

Definition 3.1 (Congestion Model)**.** A tuple $\mathcal{M} = (N, R, \{X_i\}_{i \in N}, \{C_r\}_{r \in R})$ is called a *congestion model* if $N = \{1, \ldots, n\}$ is a non-empty, finite set of players, $R = \{1, \ldots, m\}$ is a non-empty, finite set of resources, and for each player $i \in N$, her collection of accessible sets $X_i = \{R_{i1}, \ldots, R_{i\,m_i}\}$, $m_i \in \mathbb{N}$, is a non-empty, finite set of subsets of R. We will use the shorthand notation $M_i = \{1, \ldots, m_i\}$. Every resource $r \in R$ has a cost function $C_r : \mathbb{R}_+ \to \mathbb{R}_+$.

Assumption 1. Cost functions $C_r : \mathbb{R}_+ \to \mathbb{R}_+$, $r \in R$, are differentiable, convex, non-decreasing functions, with $\lim_{x \to \infty} \frac{C_r(x)}{x} = \infty$.

Given a congestion model $\mathcal{M} = (N, R, \{X_i\}_{i \in N}, \{C_r\}_{r \in R})$, we derive a corresponding *resource allocation model* $\mathcal{RM} = (N, R, \{X_i\}_{i \in N}, \Phi, \{C_r\}_{r \in R})$, where $\Phi = \times_{i \in N} \Phi_i$, and $\Phi_i = \mathbb{R}_+^{m_i}$ defines the strategy space for player i. A strategy profile $\varphi_i = (\varphi_{i1}, \ldots, \varphi_{i\,m_i})$ of player i can be interpreted as a distribution of non-negative *demands* over the elements in X_i. The total demand of player i is defined by $d_i(\varphi) = \sum_{j=1}^{m_i} \varphi_{ij}$. For $i \in N$, $\Phi_{-i} = \Phi_1 \times \ldots \times \Phi_{i-1} \times \Phi_{i+1} \times \ldots \times \Phi_n$ denotes the strategy space of all players except for player i. With a slight abuse of notation we will sometimes write a strategy profile as $\varphi = (\varphi_i, \varphi_{-i})$ meaning that $\varphi_i \in \Phi_i$ and $\varphi_{-i} \in \Phi_{-i}$. For a given profile φ, the load generated by player $i \in N$ on resource $r \in R$ is defined by $\varphi_i^r = \sum_{j \in M_i : r \in R_{ij}} \varphi_{ij}$. We denote by $\varphi^r = \big(\varphi_i^r,\, i \in N\big)$ the load vector of resource $r \in R$. The total load on resource $r \in R$ is defined by $\ell_r(\varphi) = \sum_{i=1}^n \varphi_i^r$. We will give two examples of a resource allocation model.

Example 3.2 (Network Resource Allocation)**.** A resource allocation model $\mathcal{RM}$ is called a network resource allocation model if the set of resources correspond to the set of arcs of a directed or undirected graph G, every player i corresponds to a commodity having two distinguished vertices (s_i, t_i) (s_i is the source and t_i the terminal vertex in G, respectively), and the collection of player i's accessible sets (X_i) is the set of corresponding (s_i, t_i)-paths. Thus, a strategy for player i corresponds to sending a non-negative demand along the available (s_i, t_i)-paths.

Example 3.3 (Matroid Resource Allocation)**.** A resource allocation model $\mathcal{RM}$ is called *matroid* resource allocation model if for every $i \in N$, there is a matroid $M_i = (R, \mathcal{I}_i)$ (note that $\mathcal{I}_i$ refers to an independence system in R, see [45] for an introduction to matroids) such that X_i equals the set of bases of M_i. A prominent example of a *matroid* resource allocation models arises if the resources form a graph and the set of bases correspond to the set of spanning trees in G. In this case, a strategy for player i corresponds to sending a non-negative demand along the available spanning trees of G.

3.2. Cost Sharing Methods

We define a cost sharing method as a collection of functions, one for each resource that takes as input the vector of the players' loads on the resource and outputs a vector of cost shares for each player. We restrict the set of feasible cost sharing methods as defined below.

Definition 3.4. Given a resource allocation model $\mathcal{RM} = (N, R, \{X_i\}_{i \in N}, \Phi, \{C_r\}_{r \in R})$, a cost sharing method for a resource $r \in R$ is a mapping $\xi^r : \mathbb{R}_+^n \to \mathbb{R}_+^n$. We define the following conditions

(1) **Cost-covering:** $\sum_{i=1}^{n} \xi_i^r(\varphi^r) \geq C\big(\ell_r(\varphi)\big)$ for all $\varphi \in \Phi$;
(2) **Nash-inducing:** $\xi_i^r(\varphi^r)$ is non-decreasing, differentiable and convex in φ_i^r for all $i \in N$;
(3) **No charge for zero demand:** $\xi_i^r(\varphi^r) = 0$ for all $\varphi \in \Phi$ with $\varphi_i^r = 0$, for all $i \in N$;
(4) **Scalability:** $\xi_i^r(\varphi^r) \cdot \varphi_j^r = \xi_j^r(\varphi^r) \cdot \varphi_i^r$ for all $i, j \in N$, and all $\varphi \in \Phi$.

A cost sharing method is called *basic*, if it satisfies the assumptions (1)-(3) and it is called *scalable*, if it satisfies the assumptions (1)-(4). Note that a basic cost sharing method is automatically *separable* in the sense of condition (1) in Section 1, because every ξ^r has only φ^r as argument.

We next discuss the above assumptions in detail. The first assumption is standard in the economics literature and the least critical: the cost of using a resource is passed to its users. The second assumption ensures the existence of a pure Nash equilibrium of the induced resource allocation game. Moreover, a positive charge for zero resource consumption prevents users from participation and is thus considered undesirable, see [31]. Assumption (4) stating that the price per unit resource consumption must be equal for all players is perhaps the most restrictive and controversial one. In the context of large scale networks (e.g., the TCP/IP protocol suite used in the Internet) this property is considered desirable and indispensable, because every resource only needs to pass a one-dimensional information to its users. For a detailed discussion on this subject, we refer the reader to [25], [28] and [47]. We give in the following three examples of cost-sharing methods that we will analyze throughout this paper.

Example 3.5 (Average Cost Sharing)**.** In average cost sharing, the cost share for player i on resource r under profile φ is defined as $\xi_i^r(\varphi^r) = \varphi_i^r \cdot \frac{C_r(\ell_r(\varphi))}{\ell_r(\varphi)}$. This cost sharing method is widely in the transportation literature (cf. [4, 17]) for modeling the experienced travel time, where the term $c_r(\ell_r(\varphi)) := \frac{C_r(\ell_r(\varphi))}{\ell_r(\varphi)}$ models the load-dependent latency function on r. Note that average cost sharing is a scalable cost sharing method.

Proposition 3.6. The only cost sharing method that is exactly budget balanced and fulfills Assumption (4) (single price per unit) in Definition 3.4 is average cost sharing.

Example 3.7 (Marginal Cost Pricing)**.** In marginal cost pricing, the cost share for player i on resource r under profile φ is defined as $\xi_i^r(\varphi^r) = \varphi_i^r \cdot C_r'(\ell_r(\varphi))$. Note that marginal cost pricing is a scalable cost sharing method.

Example 3.8 (Incremental Cost Sharing)**.** In incremental cost sharing, the cost share for player i on resource r under profile φ is defined as $\xi_i^r(\varphi^r) = C_r(\ell_r(\varphi)) - C_r(\ell_r(0, \varphi_{-i}))$. One can easily show that incremental cost sharing is not scalable.

Remark 3.9. While the incremental cost sharing method is not scalable, it still satisfies the symmetry condition: $\xi_i^r(\varphi^r) = \xi_j^r(\varphi^r)$ for all $i, j \in N$ and $\varphi \in \Phi$ with $\varphi_i^r = \varphi_j^r$. The above property is considered desirable in the economics literature and refers to the notion of fairness between resource consumers: if two players have an equal resource consumption, their cost share must be equal.

3.3. Resource Allocation Games

We are now ready to formally define a resource allocation game. By choosing a strategy φ_i, player i receives a certain benefit measured by a utility function $U_i\big(d_i(\varphi)\big)$. We assume that utility functions satisfy the following conditions.

Assumption 2. Each utility function $U_i : \mathbb{R}_+ \to \mathbb{R}_+$, is differentiable, strictly increasing, and concave.

Definition 3.10 (Resource Allocation Game)**.** Given a resource allocation model $\mathcal{RM}$, the corresponding *resource allocation game* is the strategic game $G(\mathcal{RM}) = (N, \Phi, \pi)$, where the payoff $\pi = (\pi_1, \dots, \pi_n)$ is defined as $\pi_i(\varphi) := U_i\big(d_i(\varphi)\big) - \sum_{r\in R} \xi_i^r\big(\varphi^r\big)$, where $\xi_i^r : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is the cost sharing method of resource $r \in R$.

For the remainder of this paper, we will write G instead of $G(\mathcal{RM})$.

Remark 3.11. Assumptions 1, 2, and Definition 3.4 imply $\lim_{\|\varphi_i\|\to\infty} \pi_i\left(\varphi_i;\, \varphi_{-i}\right) = -\infty$, hence, we can effectively restrict the strategy space for every player to a compact set. As the payoff functions are concave, a pure Nash equilibrium exists, see the result of [38].

The *total surplus* of a profile φ is defined as $\mathcal{U}(\varphi) := \sum_{i=1}^n U_i\big(d_i(\varphi)\big) - C(\varphi)$, where $C(\varphi) = \sum_{r\in R} C_r\big(\ell_r(\varphi)\big)$ is the total cost function for the profile φ. A profile of maximum total surplus is called *optimal*. We define the following functions: $\hat{\xi}_i^r\left(\varphi^r\right) := \frac{\partial \xi_i^r(\varphi^r)}{\partial \varphi_i^r}$ and $\hat{\xi}_{ij}(\varphi) := \sum_{r\in R_{ij}} \hat{\xi}_i^r(\varphi^r)$. The next lemma establishes necessary and sufficient conditions for a profile to be optimal and a Nash equilibrium, respectively.

Lemma 3.12. Consider a resource allocation game G with basic cost sharing methods $\big(\xi_r,\, r \in R\big)$. The profiles ϑ and ψ are a Nash equilibrium and an optimal profile, respectively, if and only if for all players i the following conditions hold:

$$\nabla\pi_i\big(\vartheta_i;\, \vartheta_{-i}\big)\cdot\big(\varphi_i - \vartheta_i\big) \le 0, \quad \text{for all } \varphi_i \in \Phi_i\,, \tag{1}$$

$$\begin{aligned} U_i'\big(d_i(\vartheta)\big) &= \hat{\xi}_{ij}(\vartheta), \quad \text{for all } j \in M_i \text{ with } \vartheta_{ij} > 0,\\ U_i'\big(d_i(\vartheta)\big) &\le \hat{\xi}_{ij}(\vartheta), \quad \text{for all } j \in M_i \text{ with } \vartheta_{ij} = 0, \end{aligned} \tag{2}$$

$$\begin{aligned} U_i'\big(d_i(\psi)\big) &= \sum_{r\in R_{ij}} C_r'\big(\ell_r(\psi)\big), \quad \text{for all } j \in M_i \text{ with } \psi_{ij} > 0,\\ U_i'\big(d_i(\psi)\big) &\le \sum_{r\in R_{ij}} C_r'\big(\ell_r(\psi)\big), \quad \text{for all } j \in M_i \text{ with } \psi_{ij} = 0. \end{aligned} \tag{3}$$

In the following sections, we will analyze the worst-case efficiency of Nash equilibria for several cost sharing methods. In Section 4, we first develop a general lower bound for basic cost sharing methods. We proceed by studying marginal cost pricing and average cost sharing in Section 5 and Section 6, respectively.

4. Worst-Case Efficiency of Basic Cost Sharing Methods

In the following, we will study the worst-case efficiency loss for a class of basic cost sharing methods. Throughout the analysis we assume that cost functions satisfy Assumption 1 and utility functions satisfy Assumption 2. Before we give a formal definition of the worst-case efficiency loss, we prove an auxiliary lemma, showing that basic cost sharing methods always guarantee a non-negative total surplus for every Nash equilibrium.

Lemma 4.1. Let G be a resource allocation games with n players, cost functions in $\mathcal{C}$, and basic cost sharing methods $\xi_r \in \mathcal{D}_n$ for all $r \in R$. Let Θ_G be the set of Nash equilibria. Then, $\mathcal{U}(\vartheta) \ge 0$ for all $\vartheta \in \Theta_G$.

Next, we provide a formal definition of the worst-case efficiency loss.

Definition 4.2. Let $\mathcal{C}$ be a class of cost functions. Let $\mathcal{G}_n\big(\mathcal{C}, \mathcal{D}_n\big)$ be the set of all resource allocation games with n players, cost functions in $\mathcal{C}$, and basic cost sharing methods $\xi_r \in \mathcal{D}_n$ for all $r \in R$. For $G \in \mathcal{G}_n(\mathcal{C}, \mathcal{D}_n)$, let ψ_G be an optimal profile and let Θ_G be the set of pure Nash equilibria, respectively. Then, the *worst case efficiency* is defined by

$$\rho_n(\mathcal{C}, \mathcal{D}_n) = \begin{cases} \inf_{G \in \mathcal{G}_n(\mathcal{C}, \mathcal{D}_n)} \inf_{\vartheta \in \Theta_G} \frac{\mathcal{U}_G(\vartheta)}{\mathcal{U}_G(\psi_G)}, & \text{if } \mathcal{U}_G(\psi_G) > 0, \\ 1, \text{ otherwise.} \end{cases}$$

Here, $\mathcal{U}_G$ denotes the total surplus function for game G. Conversely, $1 - \rho_n(\mathcal{C}, \mathcal{D}_n)$ defines the *worst-case efficiency loss* or *price of anarchy.*

Remark 4.3. Note that by Lemma 4.1 the case $\mathcal{U}_G(\psi_G) = 0$ implies that for basic cost sharing methods, every Nash equilibrium is optimal. Therefore, we can assume without loss of generality that every optimal profile recovers a strictly positive total surplus, i.e., $\mathcal{U}_G(\psi_G) > 0$.

We show next that for bounding the worst-case efficiency of basic cost sharing methods it is sufficient to consider games with only linear utility functions. The next lemma can be proved using ideas of [23], [31], and [9].

Lemma 4.4. Let $\mathcal{G}_n(\mathcal{C}, \mathcal{D}_n)$ be the set of all resource allocation games with n players, cost functions in $\mathcal{C}$, and basic cost sharing methods $\xi_r \in \mathcal{D}_n$ for all $r \in R$. Then, for bounding the worst-case efficiency it is enough to consider resource allocation games in which all utility functions are linear.

We proceed by calculating the surplus of an optimal solution and that of a Nash equilibrium in terms of the cost functions and cost sharing methods involved, respectively.

Lemma 4.5. Consider a game G with basic cost sharing methods and linear utility functions, that is, $U_i(x) = u_i \cdot x$, $u_i \geq 0$, $i \in N$. Let ψ be an optimal profile and ϑ be a Nash equilibrium. Then, ψ and ϑ generate a total surplus of $\mathcal{U}(\psi) = \sum_{r \in R} \Big(\ell_r(\psi) \cdot C_r'\big(\ell_r(\psi)\big) - C_r\big(\ell_r(\psi)\big) \Big)$ and $\mathcal{U}(\vartheta) = \sum_{i=1}^n \sum_{j=1}^{m_i} \hat{\xi}_{ij}(\vartheta)\, \vartheta_{ij} - C(\vartheta)$.

We provide in this section a general proof template that enables us to derive a bound on the worst-case efficiency for a resource allocation game with basic cost sharing methods. The main idea for proving such bounds is an application of the variational inequality. Let ψ and ϑ be an optimal and a Nash profile, respectively. Observe that for any λ, the following inequality holds: $\mathcal{U}(\psi) \leq \lambda \mathcal{U}(\vartheta) + \mathcal{U}(\psi) + \sum_{i=1}^n \nabla \pi_i(\vartheta_i;\, \vartheta_{-i}) \cdot (\vartheta_i - \psi_i) - \lambda \mathcal{U}(\vartheta)$. If we can derive an inequality of the form $\Big(\mathcal{U}(\psi) + \sum_{i=1}^n \nabla \pi_i(\vartheta_i;\, \vartheta_{-i}) \cdot (\vartheta_i - \psi_i) - \lambda \mathcal{U}(\vartheta) \Big) / \mathcal{U}(\psi) \leq \omega(\lambda)$ for some $\omega(\lambda) < 1$, we would obtain the inequality $\mathcal{U}(\psi) \leq \lambda \mathcal{U}(\vartheta) + \omega(\lambda) \mathcal{U}(\psi)$, which yields a bound on the worst case efficiency of $\frac{1 - \omega(\lambda)}{\lambda}$. As a consequence, we could then optimize over λ (which of course involves $\omega(\lambda)$) so as to derive the best possible bound. This technique (λ-technique) has been previously applied to bound the price of anarchy in atomic splittable congestion games, see [16].

In the following, we denote by $\mathcal{D}_n$ a class of basic cost sharing methods for n players. For a cost function C, a cost sharing method $\xi \in \mathcal{D}_n$, and a parameter $\lambda > 0$, we define the following value

$$\omega_n(C, \xi, \lambda) := \sup_{x, y \in \mathbb{R}_+^n} \frac{\sum_{i=1}^n \hat{\xi}_i(x)\, (y_i - \lambda\, x_i) + \lambda\, C\big(\ell(x)\big) - C\big(\ell(y)\big)}{C'\big(\ell(y)\big) \cdot \ell(y) - C\big(\ell(y)\big)}, \tag{4}$$

where $\ell(x) = \sum_{i=1}^n x_i$. For a class of cost functions $\mathcal{C}$, and a class of basic cost sharing methods $\mathcal{D}_n$, we define $\omega_n(\mathcal{C}, \mathcal{D}_n, \lambda) := \sup_{\xi \in \mathcal{D}_n} \sup_{C \in \mathcal{C}} \omega_n(C, \xi, \lambda)$. We define the feasible λ-region as $\Lambda(\mathcal{C}, \mathcal{D}_n) := \{\lambda > 0 \,|\, \omega_n(\mathcal{C}, \mathcal{D}_n, \lambda) < 1\}$.

Theorem 4.6. Consider the set $\mathcal{G}_n\big(\mathcal{C}, \mathcal{D}_n\big)$ of resource allocation games with basic cost sharing methods $\xi^r \in \mathcal{D}_n$, $r \in R$, and cost functions in $\mathcal{C}$. Then, the worst case efficiency is at least

$$\rho(\mathcal{C}, \mathcal{D}_n) \geq \sup_{\lambda \in \Lambda(\mathcal{C}, \mathcal{D}_n)} \left[\frac{1 - \omega_n(\mathcal{C}, \mathcal{D}_n, \lambda)}{\lambda}\right].$$

Proof. Let $G \in \mathcal{G}_n\big(\mathcal{C}, \mathcal{D}_n\big)$. Using Lemma 4.4 we may assume that utility functions are linear. Let ψ and ϑ be an optimal and a Nash profile, respectively. Observe that for any λ, the following inequalities hold

$$\mathcal{U}(\psi) \leq \lambda \mathcal{U}(\vartheta) + \mathcal{U}(\psi) + \sum_{i=1}^{n} \nabla \pi_i\big(\vartheta_i;\, \vartheta_{-i}\big) \cdot \big(\vartheta_i - \psi_i\big) - \lambda \mathcal{U}(\vartheta) \tag{5}$$

$$= \lambda \mathcal{U}(\vartheta) - C(\psi) + \sum_{i=1}^{n} u_i \cdot d_i(\vartheta) + \sum_{r \in R} \sum_{i=1}^{n} \hat{\xi}_i^r\big(\vartheta^r\big)\,\big(\psi_i^r - \vartheta_i^r\big) - \lambda \mathcal{U}(\vartheta)$$

$$= \lambda \mathcal{U}(\vartheta) - C(\psi) + \sum_{r \in R} \sum_{i=1}^{n} \hat{\xi}_i^r\big(\vartheta^r\big)\,\psi_i^r - \lambda \mathcal{U}(\vartheta) \tag{6}$$

$$= \lambda \mathcal{U}(\vartheta) - C(\psi) + \sum_{r \in R} \sum_{i=1}^{n} \hat{\xi}_i^r\big(\vartheta^r\big)\,\big(\psi_i^r - \lambda\,\vartheta_i^r\big) + \lambda\, C(\vartheta)\,. \tag{7}$$

Here, (5) and (6) follow from Lemma 3.12, while (7) follows from Lemma 4.5. In order to complete the proof, we need to show that

$$\sum_{r \in R} \sum_{i=1}^{n} \hat{\xi}_i^r\big(\vartheta^r\big)\,\big(\psi_i^r - \lambda\,\vartheta_i^r\big) + \lambda\, C(\vartheta) - C(\psi) \leq \omega_n\big(\mathcal{C}, \mathcal{D}_n, \lambda\big) \cdot \mathcal{U}(\psi)\,. \tag{8}$$

By definition of $\omega_n\big(\mathcal{C}, \mathcal{D}_n, \lambda\big)$, we have

$$\frac{\sum_{i=1}^{n} \hat{\xi}_i^r\big(\vartheta^r\big)\,\big(\psi_i^r - \lambda\,\vartheta_i^r\big) + \lambda\, C_r\big(\ell_r(\vartheta)\big) - C_r\big(\ell_r(\psi)\big)}{\ell_r(\psi) \cdot C_r'\big(\ell_r(\psi)\big) - C_r\big(\ell_r(\psi)\big)} \leq \omega_n\big(\mathcal{C}, \mathcal{D}_n, \lambda\big)$$

for all $r \in R$. Multiplying this inequality by $\ell_r(\psi) \cdot C_r'\big(\ell_r(\psi)\big) - C_r\big(\ell_r(\psi)\big)$, summing up over $r \in R$, and using Lemma 4.5, we obtain (8). □

We briefly pause here to discuss implications of the above result. Theorem 4.6 provides a lower bound on the worst-case efficiency of Nash equilibria that only depends on $\mathcal{C}$ and $\mathcal{D}_n$ but neither on the player's private utilities nor on the strategy space. If the sets $\mathcal{C}$ and $\mathcal{D}_n$ have a specific form (e.g., convex cost functions and marginal cost pricing), then, evaluating the concrete bound in Theorem 4.6 amounts to solving a highly structured optimization problem. In the remainder of the paper we will actually solve this optimization problem for three specific cost sharing methods (incremental cost sharing, marginal cost pricing, and average cost sharing) and different classes of cost functions. We will first apply Theorem 4.6 to prove that the incremental cost sharing method is actually an optimal mechanism among all basic mechanisms.

Proposition 4.7. For incremental cost sharing, every Nash equilibrium is optimal.

[31] showed that incremental cost sharing is optimal for resource allocation games with a single resource. Proposition 4.7 generalizes Moulin's result to hold for general resource allocation games.

5. The Worst-Case Efficiency of Marginal Cost Pricing

In the previous section, we showed that among all basic cost sharing mechanisms, there is an optimal mechanism (incremental cost sharing) that achieves full efficiency. Because the incremental cost sharing method is not scalable, we will focus in this section on *marginal cost pricing* which is a well-known scalable cost sharing method. More precisely, we will study the price of anarchy in games where all resources use marginal cost pricing as cost sharing method. We thus have $\xi_i^r\big(\phi^r\big) = \phi_i^r \cdot C_r'\big(\ell_r(\phi)\big)$ for all $r \in R$, $i \in N$, and $\phi \in \Phi$. We will call $C_r'(\cdot)$ the *marginal cost function* of resource $r \in R$. For the rest of this section, we assume that all cost functions $C_r(\cdot)$ are twice differentiable for all $r \in R$. Instead of $\mathcal{G}_n\big(\mathcal{C}, \mathcal{D}_n\big)$, we will use the shorthand $\mathcal{G}_n\big(\mathcal{C}\big)$ assuming that $\mathcal{D}_n$ corresponds to marginal cost pricing. In Lemma 4.5, we represented the total surplus of a Nash equilibrium and that of an optimal profile in terms of the involved cost functions for a general cost sharing method. The following lemma is a special case of this result for marginal cost pricing.

Lemma 5.1. Consider a game G with marginal cost pricing and linear utility functions, that is, $U_i(x) = u_i \cdot x$, $u_i \geq 0$, $i \in N$. Let ϑ be a Nash equilibrium. Then, ϑ generates total surplus of

$$\mathcal{U}(\vartheta) = \sum_{r\in R} \Big(\ell_r(\vartheta)\cdot C_r'\big(\ell_r(\vartheta)\big) + \sum_{i=1}^{n} (\vartheta_i^r)^2 \cdot C_r''\big(\ell_r(\vartheta)\big) - C_r\big(\ell_r(\vartheta)\big)\Big)\,. \tag{9}$$

The proof follows from Lemma 4.5. We proceed by deriving an upper bound for $\omega_n\big(C, \xi, \lambda\big)$ using that ξ corresponds to marginal cost pricing.

Lemma 5.2. Let ξ be a marginal cost pricing method for n players. Then, $\omega_n\big(C, \xi, \lambda\big) \leq \omega_n^{mcp}(C, \lambda)$, where

$$\begin{aligned}\omega_n^{mcp}(C,\lambda) := \sup_{\substack{x,y\in\mathbb{R}_+\\ \mu\in\{0\}\cup[\frac{1}{n},1]}} \Big[\Big(&C'(x)y + C''(x)\mu x y + \lambda C(x) - C(y) \\ &- \lambda\big(C'(x)x + \big(\mu^2 + \frac{(1-\mu)^2}{n-1}\big)C''(x)x^2\big)\Big)/\Big(C'(y)\cdot y - C(y)\Big)\Big].\end{aligned} \tag{10}$$

An essential element of the definition of $\omega_n^{mcp}(C, \lambda)$ is the parameter μ defined as the largest ratio of the load of a single player and the overall load on a resource. We note that this ratio has been used before in the context of bounding the price of anarchy in atomic splittable network games with fixed demands, see [11], [16] and [50].

5.1. Cost Functions with a Convex Derivative

We start by applying Lemma 5.2 to convex marginal cost functions, that is, we consider cost functions with a convex derivative.

Theorem 5.3. Let $\mathcal{C}^{convD}$ be a class of cost function that have a convex derivative. Consider the set $\mathcal{G}_n(\mathcal{C}^{convD})$ of games with at most n players. Then, $\rho_n(\mathcal{C}^{convD}) \geq \frac{4}{3+\sqrt{5+4\,n}}$.

Proof. We define $\lambda = \frac{3+\sqrt{5+4\,n}}{4}$ and prove the claim by showing that $\omega_n^{mcp}(C; \lambda) \leq 0$ for all $C \in \mathcal{C}^{convD}$. We bound the nominator of (10) by a case distinction. First, we assume $x \geq y$. We

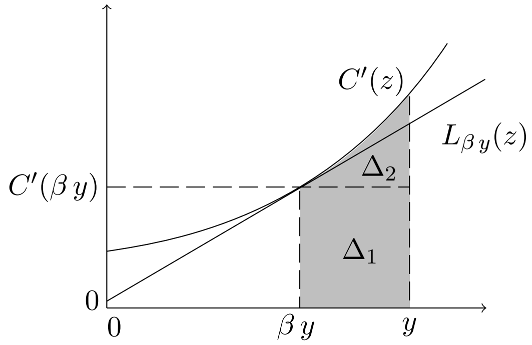


Figure 1. Illustration of the inequality (11) in the proof of Theorem (5.3). The shaded area illustrates the term $C(y) - C(\beta y) = \Delta_1 + \Delta_2$. The linear approximation $L_{\beta y}(\cdot)$ of the convex function $C'(\cdot)$ bounds $C'(z)$ from below, i.e., $L_{\beta y}(z) \le C'(z)$. Then, we have $\Delta_1 = (y - \beta y)\, C'(\beta y)$ and $\Delta_2 \ge \frac{(y-\beta y)^2}{2}\, C''(\beta y)$.

get

$$
\begin{aligned}
&C'(x)\, y + C''(x)\, \mu\, x\, y - \lambda\, (C'(x)\, x + \Big(\mu^2 + \frac{(1-\mu)^2}{n-1}\Big)\, C''(x)\, x^2) + \lambda\, C(x) - C(y) \\
&\le C''(x)\, \Big(\mu\, x\, y - \lambda\, \big(\mu^2 + \frac{(1-\mu)^2}{n-1}\big)\, x^2\Big) \\
&\le\; C''(x)\, x^2 \Big(\mu - \lambda\, \big(\mu^2 + \frac{(1-\mu)^2}{n-1}\big)\Big)\,.
\end{aligned}
$$

For the first inequality, we used that $C'(x)\, y - \lambda\, C'(x)\, x + \lambda\, C(x) - C(y) \le 0$, because $y \le x$, $\lambda \ge 1$, and $C'(\cdot)$ is convex. The second inequality follows from $y \le x$ and $C''(x) \ge 0$ (since $C(\cdot)$ convex). Then, $\lambda = \frac{3+\sqrt{5+4n}}{4}$ yields $\omega_n^{mcp}(C;\lambda) \le 0$, because

$$
\max_{\mu \in \{0\} \cup [\frac{1}{n},1]} \mu - \Big(\frac{3+\sqrt{5+4\,n}}{4}\Big)\, \Big(\mu^2 + \frac{(1-\mu)^2}{n-1}\Big) \le 0.
$$

Now, we consider the case $x < y$. We define $\beta := \frac{x}{y} \in [0,1)$. Observe that $C(y) - \lambda\, C(\beta\, y) = C(y) - C(\beta\, y) - (\lambda - 1)\, C(\beta\, y)$. Then, we use the following inequality, which is illustrated in Fig. 1.

$$
C(y) - C(\beta\, y) \ge (y - \beta\, y)\, C'(\beta\, y) + \frac{(y-\beta\, y)^2}{2}\, C''(\beta\, y)\,. \tag{11}
$$

Together with $(\lambda - 1)\, C(\beta\, y) \le (\lambda - 1)\, C'(\beta\, y) \beta\, y$, we obtain

$$
\omega_n^{mcp}(C;\lambda) \le \sup_{\substack{\beta \in [0,1),\, y \in \mathbb{R}_+ \\ \mu \in \{0\} \cup [\frac{1}{n},1]}} \frac{C''(\beta\, y)\, y^2 \big(\beta\, \mu - \lambda\, \big(\mu^2 + \frac{(1-\mu)^2}{n-1}\big)\, \beta^2 - \frac{(1-\beta)^2}{2}\big)}{C'(y)\, y - C(y)}\,.
$$

We then use $\max_{\beta\in[0,1),\mu\in[0,1]} \Big(\beta\, \mu - \lambda\, \Big(\mu^2 + \frac{(1-\mu)^2}{n-1}\Big)\, \beta^2 - \frac{(1-\beta)^2}{2}\Big) \le \frac{n-4\,\lambda^2+6\,\lambda-1}{4\,\lambda\,(n+2\,\lambda)}$, where $\beta^* = \frac{n+1}{n+2\,\lambda}$ and $\mu^* = \frac{n-1+4\,\lambda}{2\,\lambda\,(n+1)}$ are the unique maximizer. The value of λ solves $n - 4\,\lambda^2 + 6\,\lambda - 1 = 0$, thus, we obtain $\omega_n^{mcp}(C;\lambda) \le 0$. Applying Lemma 5.2 for both cases proves the claim. □

Remark 5.4. The bound of Theorem 5.3 has been established before by [14] for the case of a single resource.

The above result gives a bound on the efficiency loss for differentiable and convex marginal cost functions scaling with the number of players. This result complements a negative result of [22] for two-player games with *non-differentiable* convex marginal cost functions, where the efficiency loss may be arbitrarily high. For non-differentiable marginal cost functions, a Nash equilibrium can be characterized by optimality conditions expressed by the left and right directional derivatives of the marginal cost function. The key ingredient of the instance in [22] is to increase the difference between two such values (for a point of non-differentiability) giving rise to a Nash equilibrium with low total surplus. In contrast, if marginal cost functions are differentiable, then by Lemma 5.1 the total surplus of an arbitrary Nash equilibrium can be expressed in terms of the involved cost functions and their *well-defined* derivatives ruling out the instance constructed in [22]. Note that in many applications the considered marginal cost functions are differentiable, e.g., polynomial delay functions considered in transportation networks ([6]) and $M/M/1$ functions modeling queuing delays in telecommunication networks ([47]).

Remark 5.5. In the next section (Proposition 5.10), we present an asymptotically matching upper bound for the efficiency loss of $\rho_n(\mathcal{C}^{convD}) \leq \frac{2(n-\sqrt{n})}{\sqrt{n}(n-1)}$ which grows as $O(1/\sqrt{n})$.

5.2. Polynomial Cost Functions

In practice, the most frequently used functions modeling delay are polynomials whose degrees and coefficients are determined from real-world data through statistical evaluation methods, see [36] and [6]. Thus, we will explicitly calculate the price of anarchy for the class $\mathcal{C}_d := \Big\{C(z) = \sum_{j=0}^d a_j\, z^j,\ a_j \geq 0,\ j = 2,\dots,d\Big\}$, $d \in \{2,3,\dots\}$. Note that we have to demand $d \geq 2$ since otherwise Assumption 1 would be violated and a Nash equilibrium might not exist.

To simplify the analysis, we focus on the general case $n \in \mathbb{N}^* \cup \{\infty\}$. Let us define $\omega_\infty^{mcp}(C;\lambda) := \lim_{n\to\infty} \omega_n^{mcp}(C;\lambda)$. It is easy to see that $\omega_\infty^{mcp}(C;\lambda) \geq \omega_n^{mcp}(C;\lambda)$ for any $n \in \mathbb{N}^*$, implying $\rho_\infty(C) \leq \rho_n(C)$.

Remark 5.6. We observe that for marginal cost pricing, the payoff functions $\pi_i(\cdot)$ are affine linear in each of the cost functions $C_r(\cdot)$. We can reduce the analysis to monomial cost functions subdividing each resource r into $d+1$ resources $r_0,\dots,r_d$ with monomial cost functions $C_{r_s}\big(\ell_r(\varphi)\big) = C_{r_s} \cdot \big(\ell_r(\varphi)\big)^s$ for $s \in \{0,1,\dots,d\}$. By extending the accessible sets of every player accordingly, we obtain a transformed game in which the set of Nash equilibria, optimal profiles and corresponding surplus values coincide.

We present in the next lemma an upper bound for the value $\omega_\infty^{mcp}(\mathcal{M}_d;\lambda)$.

Lemma 5.7. Consider the class $\mathcal{M}_d := \Big\{C(z) = a_d\, z^d,\ a_d \geq 0,\ d \in \{2,3,\dots\}\Big\}$. Then, it holds that

$$\omega_\infty^{mcp}(\mathcal{M}_d;\lambda) \leq \left(\frac{1+\mu(d-1)}{\lambda(1+\mu^2 d)}\right)^{d-1}\left(\frac{d}{d-1}+\mu-1\right) - \frac{1}{d-1}, \text{ where } \mu(d) = \frac{1}{\sqrt{d-1}+1}.$$

Given the above upper bound on $\omega_\infty^{mcp}(\mathcal{M}_d;\lambda)$ we now give a precise bound for $\rho(\mathcal{C}_d)$.

Theorem 5.8. Let $\mathcal{C}_d$ be the class of polynomial cost functions with non-negative coefficients and maximum degree $d \in \{2,3,\dots\}$. Then, $\rho(\mathcal{C}_d) = \frac{1+\mu(d)^2\, d}{\big(1+\mu(d)\,(d-1)\big)^{1+\frac{1}{d-1}}}$, where $\mu(d) = \frac{1}{\sqrt{d-1}+1}$.

Proof. We define $\lambda = \frac{\big(1+\mu(d)\,(d-1)\big)^{1+\frac{1}{d-1}}}{1+\mu(d)^2\, d}$. Then, Lemma 5.7 implies $\omega_\infty(\mathcal{M}_{d'};\lambda) \leq 0$ for all $d' < d$ and $\omega_\infty(\mathcal{M}_d;\lambda) = 0$. Thus, using Lemma 5.2 and Theorem 4.6, we have $\rho(\mathcal{C}_d) \geq \frac{1+\mu(d)^2\, d}{\big(1+\mu(d)\,(d-1)\big)^{1+\frac{1}{d-1}}}$.

Now we prove the upper bound. Consider a game with one resource having the cost function $C(x) = \frac{1}{d}\,x^d$ for some $d \in \{2, 3, \dots\}$. Assume we have n players, where player 1 has the utility function $U_1(\varphi_1) = \varphi_1$, while the remaining $n-1$ players have utility functions $U_k(\varphi_k) = b\,\varphi_k$ for some $b \in [0,1]$ specified later. Consider a Nash equilibrium $\vartheta(n)$ in this game. W.l.o.g., we can assume $\ell(\vartheta(n)) > 0$. Using Lemma 3.12, we obtain $\vartheta_1(n) = \frac{1-\ell(\vartheta(n))^{d-1}}{(d-1)\,\ell(\vartheta(n))^{d-1}}$ for player 1 and $\vartheta_k(n) = \frac{b-\ell(\vartheta(n))^{d-1}}{(d-1)\,\ell(\vartheta(n))^{d-1}}$ for players $k = 2, \dots, n$. Summing all demands we get

$$\ell(\vartheta(n)) = \frac{1-\ell(\vartheta(n))^{d-1}}{(d-1)\,\ell(\vartheta(n))^{d-1}} + (n-1)\,\frac{b-\ell(\vartheta(n))^{d-1}}{(d-1)\,\ell(\vartheta(n))^{d-1}} \quad \Leftrightarrow \quad \ell(\vartheta(n)) = \left(\frac{1+b(n-1)}{d-1+n}\right)^{\frac{1}{d-1}}.$$

In the limit, we get $\lim_{n\to\infty} \ell(\vartheta(n)) = b^{\frac{1}{d-1}}$, $\lim_{n\to\infty} \vartheta_1(n) = \frac{b^{\frac{1}{d-1}}(1-b)}{b\,(d-1)}$, and $\lim_{n\to\infty} b\,(n-1)\,\vartheta_k(n) = \frac{b^{\frac{1}{d-1}}(b\,d-1)}{d-1}$. Thus, we get as limit for the total surplus of the Nash equilibrium $\vartheta(n)$

$$\lim_{n\to\infty} \mathcal{U}\big(\vartheta(n)\big) = \frac{b^{\frac{1}{d-1}}\,(1-b)}{b\,(d-1)} + \frac{b^{\frac{1}{d-1}}\,(b\,d-1)}{d-1} - \frac{b^{\frac{1}{d-1}}\,b}{d}.$$

An optimal solution is given by $\psi = (1, 0, \dots, 0)$ with total surplus of $\mathcal{U}(\psi) = 1 - \frac{1}{d}$. Now choosing $b = \frac{1+(d-1)^{\frac{3}{2}}}{d^2-d+2}$ the ratio $\frac{\mathcal{U}(\vartheta)}{\mathcal{U}(\psi)}$ coincides with the lower bound of the theorem. $\square$

Remark 5.9. The worst case efficiency for cost functions in $\mathcal{C}_d$ is asymptotically bounded from below by $\Omega\left(\frac{1}{\sqrt{d-1}}\right)$.

Note that the example used in the previous proof can also be used to construct an upper bound for $\rho_n(\mathcal{C}^{convD})$ complementing Theorem 5.3.

Proposition 5.10. Let $\mathcal{C}^{convD}$ be a class of cost functions with a convex derivative. Consider the set $\mathcal{G}_n(\mathcal{C}^{convD})$ of games with at most $n \in \mathbb{N}^*$ players. Then, $\rho_n(\mathcal{C}^{convD}) \leq \frac{2(n-\sqrt{n})}{\sqrt{n}(n-1)}$ which grows as $O(1/\sqrt{n})$.

5.3. Symmetric Games

In this section, we consider *symmetric games* in which all players have the same utility function $U_i = U_j$ for all $i, j \in N$ and the same strategy space, that is, $\Phi_i = \Phi_j$ for all $i, j \in N$. Symmetric resource allocation games have been considered before in the context of single-commodity network games with atomic players, unit demands and splittable flows, see [11] and [2]. Another example of a symmetric resource allocation game arises in *scheduling games* in which there are m machines used by n players having the same utility function. The strategy of every player is simply a distribution of her workload over the machines. We prove that the symmetry assumption implies improved bounds on the worst-case efficiency loss. Consider a symmetric game with n players. Then, there exists a symmetric optimal profile ψ in the sense that $\psi_i^r = \frac{1}{n}\psi^r$ for all $i \in N$. We obtain the following bound for $\omega_n(C, \xi, \lambda)$.

Lemma 5.11. Consider a symmetric game with ξ being marginal cost pricing. Then, it holds that $\omega_n\big(C, \xi, \lambda\big) \leq \omega_n^{mcp,\,sym}(C, \lambda)$, where

$$\omega_n^{mcp,\,sym}(C;\lambda) := \sup_{x,y\in\mathbb{R}_+} \frac{C'(x)\,y + C''(x)\,\frac{x\,y}{n} - \lambda\big(C'(x)\,x + C''(x)\,\frac{x^2}{n} - C(x)\big) - C(y)}{C'(y)\cdot y - C(y)}.$$

Note that the above Lemma only uses the existence of a symmetric optimal profile but does not rely on symmetry of every Nash equilibrium. The following result for cost functions with a convex derivative has been previously obtained by [22] for the special case of games with a single resource. We present here a more general result (arbitrary symmetric strategy space) with a simpler proof.

Proposition 5.12. Let $\mathcal{C}^{convD}$ be the class of cost functions with a convex derivative. Consider the set $\mathcal{G}_n(\mathcal{C}^{convD})$ of symmetric games with at most $n \in \mathbb{N}^*$ players. Then, $\rho_n(\mathcal{C}^{convD}) \geq \frac{2n}{2n+1}$.

For polynomials with non-negative coefficients and arbitrary degree $d \in \{2, 3, \dots\}$, we prove the following.

Theorem 5.13. Let $\mathcal{C}_d$ be the class of polynomial cost function with non-negative coefficients and arbitrary degree $d \in \{2, 3, \dots\}$. Consider the set $\mathcal{G}_n(\mathcal{C}_d)$ of symmetric games with n players and cost functions in $\mathcal{C}_d$. Then, $\rho_n(\mathcal{C}_d) = \frac{3}{4}$.

Note that the above bound on the worst-case efficiency does neither depend on the maximum degree d of the polynomial nor on the number of players n.

6. The Worst-Case Efficiency of Average Cost Sharing

In this section, we derive lower bounds on the worst-case efficiency of average cost sharing which is the prevailing cost sharing method in transportation networks (cf. [4, 17]). In the context of transportation networks, there is a load-dependent latency function $c_r(\ell_r(\varphi))$ on every resource and the cost of resource r under profile φ is defined as $C_r\big(\ell_r(\varphi)\big) = c_r(\ell_r(\varphi))\,\ell_r(\varphi)$, while the cost share for user i on resource r is determined as $\xi_i^r(\varphi) = c_r(\ell_r(\varphi))\,\varphi_r^i = \frac{C_r\big(\ell_r(\varphi)\big)}{\ell_r(\varphi)}\,\varphi_r^i$. Note that average cost sharing is a scalable cost sharing method. Given a cost function C_r, we define the *per-unit cost function* by $c_r(\ell_r(\varphi)) = \frac{C_r\big(\ell_r(\varphi)\big)}{\ell_r(\varphi)}$. Instead of $\mathcal{G}_n\big(\mathcal{C}, \mathcal{D}_n\big)$, we will use the shorthand $\mathcal{G}_n\big(\mathcal{C}\big)$ assuming that $\mathcal{D}_n$ corresponds to average cost sharing.

Similar to the analysis of the marginal cost sharing method, we study the worst-case efficiency of average cost sharing for three types of cost functions. First, we consider general convex cost functions and derive lower bound for the worst-case efficiency of $1/n$. Second, we consider cost functions with convex per-unit costs and characterize the price of anarchy for this case. Finally, we conclude the section by analyzing average cost sharing in symmetric games.

Theorem 6.1. Let $\mathcal{C}^{conv}$ be a class of convex cost functions. Consider the set $\mathcal{G}_n(\mathcal{C}^{conv})$ of games with at most $n \in \mathbb{N}$ players. Then, $\rho_n(\mathcal{C}^{conv}) \geq \frac{1}{n}$.

We proceed by considering average cost functions, where the per-unit cost function is convex, that is, the functions $c_r, r \in R$ are convex.

Theorem 6.2. Let $\mathcal{C}^{convU}$ be a class of cost functions with convex unit costs. Then, $\rho(\mathcal{C}^{convU}) = \frac{4}{n+3}$.

Remark 6.3. The bounds of Theorem 6.1 and Theorem 6.2 have been established before by [31] for the case of a single resource.

We close this section by analyzing the efficiency loss of average cost sharing in symmetric games.

Theorem 6.4. Let $\mathcal{C}^{convU}$ be a class of cost functions with convex unit costs. Consider the set $\mathcal{G}_n(\mathcal{C}^{convD})$ of symmetric games with at most $n \in \mathbb{N}^*$ players. Then, $\rho(\mathcal{C}^{convD}) = \frac{4n}{(n+1)^2}$.

7. Conclusions and Future Work

In this work, we studied the worst-case efficiency of Nash equilibria in resource allocation games for different cost-sharing methods. We derived various new results about the efficiency loss for marginal cost pricing and average cost sharing depending on the structure of allowable cost functions. In particular, we were able to prove tight bounds for the worst-case efficiency loss for average cost sharing and marginal cost pricing involving polynomial costs with non-negative coefficients. As this class of functions is quite rich and widely used for modeling for instance queuing delays at resources, we see our results as an important contribution towards the applicability of these cost sharing methods in practice. While we proved that the incremental cost sharing method is optimal among all basic cost sharing methods, such a strong result is not known for the class of scalable cost sharing methods. In light of the high practical relevance of the scalability property of cost sharing methods, we see the design of an *optimal* cost sharing method among all scalable mechanisms for differentiable and convex cost functions as the most important open problem.

Acknowledgment

We are grateful to the two anonymous reviewers for their detailed suggestions for extending the scope of this paper. We also thank Felix G. König, Max Klimm and Rolf H. Möhring for their thorough proof reading. Some of the results of Section 5 appeared as an extended abstract in Proc. GAMENETS 2009.

8. Table of Notation

$N = \{1, \dots, n\}$, $R = \{1, \dots, m\}$	Set of players, set of resources
In the following:	$i = 1, \dots, n$ and $r = 1, \dots, m$
$R_{ij} \subset R$	j-th accessible set of player i
$X_i = \{R_{i1}, \dots, R_{im_i}\}$, $m_i \in \mathbb{N}$	Set of accessible sets of player i
$M_i = \{1, \dots, m_i\}$	Set of indices of accessible sets of player i
$\mathcal{M} = \big(N, R, \{X_i\}_{i\in N}, \{C_r\}_{r\in R}\big)$	Congestion model
$\Phi_i = \mathbb{R}_+^{m_i}$, Φ_{-i}	Strategy space of player i and all players except i
$\phi_i = (\phi_{i1}, \dots, \phi_{im_i}) \in \Phi_i$, $\phi = (\phi_i,\, i \in N)$	Strategy of player i, strategy profile
$\Phi = \times_{i\in N} \Phi_i$	Space of strategy profiles
$d_i(\phi) = \sum_{j=1}^{m_i} \phi_{ij}$	Total demand of player i
$\mathcal{RM} = \big(N, R, \{X_i\}_{i\in N}, \Phi, \{C_r\}_{r\in R}\big)$	Resource allocation model
$C_r : \mathbb{R}_+ \to \mathbb{R}_+$	Cost function of resource r
$C(\phi) = \sum_{r\in R} C_r\big(\ell_r(\phi)\big)$, $\mathcal{C}$	Total cost for ϕ, class of cost functions
$\phi_i^r = \sum_{j\in M_i : r\in R_{ij}} \phi_{ij}$	Load of player i on resource r
$\phi^r = \big(\phi_i^r,\, i \in N\big)$, $\ell_r(\phi) = \sum_{i=1}^n \phi_i^r$	Load vector of resource r, total load of r
$\xi^r : \mathbb{R}_+^n \to \mathbb{R}_+^n$	Cost sharing method for a resource r
$\hat{\xi}_i^r(\phi^r) = \frac{\partial \xi_i^r(\phi^r)}{\partial \phi_i^r}$, $\hat{\xi}_{ij}(\phi) = \sum_{r\in R_{ij}} \hat{\xi}_i^r(\phi^r)$	Short notation
$\mathcal{D}_n$	Class of cost sharing methods for n players
$U_i : \mathbb{R}_+ \to \mathbb{R}_+$	Utility function of player i
$\mathcal{U}(\phi) = \sum_{i=1}^n U_i\big(d_i(\phi)\big) - C(\phi)$	Total surplus of a profile ϕ
$\pi_i(\phi) = U_i\big(d_i(\phi)\big) - \sum_{r\in R} \xi_i^r\big(\phi^r\big)$	Payoff of player i
$\pi = (\pi_1, \dots, \pi_n)$	Payoff vector of all players
$G(\mathcal{RM}) = (N, \Phi, \pi)$ or G	Resource allocation game
ϑ, Θ_G, ψ_G	Nash equilibrium, set of Nash equilibria, optimal profile
$\mathcal{G}_n(\mathcal{C}, \mathcal{D}_n)$, $\rho_n(\mathcal{C}, \mathcal{D}_n)$	Set of games with n players, worst-case efficiency

References

[1] D. Acemoglu and A. E. Ozdaglar, *Competition and efficiency in congested markets*, Math. Oper. Res. **32**, no. 1 (2007), pp. 1–31.

[2] E. Altman, T. Basar, T. Jimnez, and N. Shimkin, *Competitive routing in networks with polynomial costs*, IEEE Trans. Autom. Control **47** (2002), pp. 92–96.

[3] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, *A survey on networking games in telecommunications*, Comput. Oper. Res. **33**, no. 2 (2006), pp. 286–311.

[4] M. Beckmann, C. McGuire, and C. Winsten, *Studies in the Economics and Transportation*, Yale University Press, 1956.

[5] U. Bhaskar, L. Fleischer, D. Hoy, and C. Huang, *Equilibria of atomic flow games are not unique*, in Proc. 20th Annual ACM-SIAM Sympos. on Discrete Algorithms, 2009, pp. 748–757.

[6] D. Branston, *Link capacity functions*, Transp. Res. **10** (1976), pp. 223–236.

[7] S. Catoni and S. Pallotino, *Traffic equilibrium paradoxes*, Transp. Sc. **25** (1991), pp. 240–244.

[8] C. K. Chau and K. M. Sim, *The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands*, Oper. Res. Lett. **31**, no. 3 (2003), pp. 327–334.

[9] Y.-J. Chen and J. Zhang, *Design of price mechanisms for network resource allocation via price of anarchy*, Math. Programming (2010). Forthcoming.

[10] R. Cole, Y. Dodis, and T. Roughgarden, *Pricing network edges for heterogeneous selfish users*, in Proc. of the 35th Annual ACM Symp. on Theory of Computing, 2003, pp. 521–530.

[11] R. Cominetti, J. R. Correa, and N. E. Stier-Moses, *The impact of oligopolistic competition in networks*, Oper. Res. **57**, no. 6 (2009), pp. 1421–1437.

[12] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *Selfish routing in capacitated networks*, Math. Oper. Res. **29** (2004), pp. 961–976.

[13] L. Fleischer, K. Jain, and M. Mahdian, *Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games*, in Proc. 45th Annual IEEE Sympos. Found. Comput. Sci., 2004, pp. 277–285.

[14] X. Guo and H. Yang, *The price of anarchy of cournot oligopoly*, in Proc. of the 1st international Workshop on Internet and Network Economics (WINE), 2005, pp. 246–257.

[15] P. T. Harker, *Multiple equilibrium behaviors on networks*, Tran. Sc. **22**, no. 1 (1988), pp. 39–46.

[16] T. Harks, *Stackelberg strategies and collusion in network games with splittable flow*, Theory of Computing Systems **48**, no. 4 (2011), pp. 781–802.

[17] A. Haurie and P. Marcotte, *On the relationship between Nash-Cournot and Wardrop equilibria*, Networks **15** (1985), pp. 295–308.

[18] A. Hayrapetyan, E. Tardos, and T. Wexler, *The effect of collusion in congestion games*, in Proc. 38th Annual ACM Sympos. Theory Comput., ACM Press, 2006, pp. 89–98.

[19] D. Hearn and M. Ramana, *Solving congestion toll pricing models*, in Equilibrium and Advanced Transportation Modeling, P. Marcotte and S. Nguyen, eds., Kluwer Academic Publishers, 1998, pp. 109–124.

[20] R. Johari, *Efficiency loss in market mechanisms for resource allocation*, PhD thesis, MIT, 2004.

[21] R. Johari, S. Mannor, and J. N. Tsitsiklis, *Efficiency loss in a network resource allocation game: The case of elastic supply*, IEEE Trans. Autom. Control **50**, no. 11 (2005), pp. 1712–1724.

[22] R. Johari and J. Tsitsiklis, *Efficiency loss in cournot games*, tech. report, LIDS-P-2639, Laboratory for Information and Decision Systems, MIT, 2005.

[23] R. Johari and J. Tsitsiklis, *A scalable network resource allocation mechanism with bounded efficiency loss.*, IEEE J. Sel. Areas Commun. **24**, no. 5 (2006), pp. 992–999.

[24] R. Johari and J. N. Tsitsiklis, *Efficiency loss in a network resource allocation game*, Math. Oper. Res. **29**, no. 3 (2004), pp. 407–435.

[25] R. Johari and J. N. Tsitsiklis, *Efficiency of scalar-parameterized mechanisms*, Oper. Res. **57**, no. 4 (2009), pp. 823–839.

[26] G. Karakostas and S. G. Kolliopoulos, *Edge pricing of multicommodity networks for heterogeneous selfish users*, in Proc. 45th Annual IEEE Sympos. Found. Comput. Sci., 2004, pp. 268–276.

[27] F. Kelly, *Charging and rate control for elastic traffic*, Europ. Trans. on Telecom. **8** (1997), pp. 33–37.

[28] F. Kelly, A. Maulloo, and D. Tan, *Rate Control in Communication Networks: Shadow Prices, Proportional Fairness, and Stability*, J. Oper. Res. Soc. **49** (1998), pp. 237–52.

[29] F. H. Knight, *Some fallacies in the interpretation of social cost*, The Quarterly J. of Econ. **38**, no. 4 (1924), pp. 582–606.

[30] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*, Oxford University Press, 1995.

[31] H. Moulin, *The price of anarchy of serial, average and incremental cost sharing*, Economic Theory **36**, no. 3 (2008), pp. 379–405.

[32] W. Novshek, *On the existence of Cournot equilibrium*, Rev. Econom. Stud. **52**, no. 1 (1985), pp. 85–98.

[33] A. Orda, R. Rom, and N. Shimkin, *Competitive routing in multi-user communication networks*, IEEE/ACM Trans. Networking **1** (1993), pp. 510–521.

[34] G. Owen, *Game Theory*, Academic Press, 2nd ed., 1982.

[35] A. E. Ozdaglar, *Price competition with elastic traffic*, Networks **52**, no. 3 (2008), pp. 141–155.

[36] M. Patriksson, *The Traffic Assignment Problem—Models and Methods*, Topics in Transportation, VSP, 1994.

[37] G. PERAKIS, *The "price of anarchy" under nonlinear and asymmetric costs*, Math. Oper. Res. **32**, no. 3 (2007), pp. 614–628.

[38] J. ROSEN, *Existence and uniqueness of equilibrium points in concave n-player games*, Econometrica **33**, no. 3 (1965), pp. 520–534.

[39] R. ROSENTHAL, *A class of games possessing pure-strategy Nash equilibria*, Internat. J. Game Theory **2**, no. 1 (1973), pp. 65–67.

[40] T. ROUGHGARDEN, *The price of anarchy is independent of the network topology*, Journal of Computer and System Sciences **67** (2002), pp. 341–364.

[41] T. ROUGHGARDEN, *Selfish Routing and the Price of Anarchy*, The MIT Press, 2005.

[42] T. ROUGHGARDEN AND F. SCHOPPMANN, *Local smoothness and the price of anarchy in atomic splittable congestion games*, in Proc. 22nd Annual ACM-SIAM Sympos. on Discrete Algorithms, 2011, pp. 255–267.

[43] T. ROUGHGARDEN AND É. TARDOS, *How bad is selfish routing?*, J. ACM **49**, no. 2 (2002), pp. 236–259.

[44] D. SAMET AND Y. TAUMAN, *The determination of marginal cost prices under a set of axioms*, Econometrica **50**, no. 4 (1982), pp. 895–909.

[45] A. SCHRIJVER, *Combinatorial Optimization. Polyhedra and Efficiency*, Springer, 2003.

[46] M. SMITH, *The marginal cost taxation of a transportation network*, Transp. Res. Part B: Methodological **13**, no. 3 (1979), pp. 237–242.

[47] R. SRIKANT, *The Mathematics of Internet Congestion Control*, Birkhäuser Boston, 2003.

[48] C. SWAMY, *The effectiveness of Stackelberg strategies and tolls for network congestion games*, in Proc. of the 18th ACM-SIAM Symp. on Discrete Algorithms, 2007, pp. 1133–1142.

[49] J. WARDROP, *Some theoretical aspects of road traffic research*, Proc. of the Institution of Civil Engineers **1**, no. Part II (1952), pp. 325–378.

[50] H. YANG, D. HAN, AND H. LO, *Efficiency of atomic splittable selfish routing with polynomial cost functions*, Networks and Spatial Economics **8**, no. 4 (2008), pp. 443–451.

[51] H. YANG AND H. HUANG, *The multi-class, multi-criteria traffic network equilibrium and systems optimum problem*, Transp. Res. Part B: Methodological **38**, no. 1 (2004), pp. 1–15.

[52] H. YANG, W. XU, AND B. HEYDECKER, *Bounding the efficiency of road pricing*, Transp. Res. E **46**, no. 1 (2010), pp. 90 – 108.

[53] H. YANG AND X. ZHANG, *Existence of anonymous link tolls for system optimum on networks with mixed equilibrium behaviors*, Transp. Res. B **42**, no. 2 (2008), pp. 99–112.

[54] S. YANG AND B. HAJEK, *Revenue and stability of a mechanism for efficient allocation of a divisible good.* Submitted, Oct. 2006.

Electronic Companion

Proof of Proposition 3.6.

The only cost sharing method that is exactly budget balanced and fulfills Assumption (4) (single price per unit) in Definition 3.4 is average cost sharing.

Proof. Observe that $C_r(\ell_r(\varphi)) \overset{*}{=} \sum_{i=1}^n \xi_i^r(\varphi^r) = \sum_{i=1}^n \varphi_i^r \cdot \frac{\xi_i^r(\varphi^r)}{\varphi_i^r} \overset{**}{=} \frac{\xi_{i_0}^r(\varphi^r)}{\varphi_{i_0}^r} \cdot \ell_r(\varphi)$, $\forall i_0 \in N$, where we use the definition of a budget balanced cost sharing method in $(*)$ and Assumption (4) from Definition 3.4 in $(**)$. We obtain $\xi_{i_0}^r(\varphi^r) = \varphi_{i_0}^r \cdot \frac{C_r(\ell_r(\varphi))}{\ell_r(\varphi)}$, $\forall i_0 \in N$, proving the claim. □

Proof of Lemma 3.12.

Consider a resource allocation game G with basic cost sharing methods $\big(\xi_r,\, r \in R\big)$. The profiles ϑ and ψ are a Nash equilibrium and an optimal profile, respectively, if and only if for all players i the following conditions hold:

$$\nabla \pi_i\big(\vartheta_i;\, \vartheta_{-i}\big) \cdot \big(\varphi_i - \vartheta_i\big) \le 0, \quad \text{for all } \varphi_i \in \Phi_i\,,$$

$$\begin{aligned} U_i'\big(d_i(\vartheta)\big) &= \hat{\xi}_{ij}(\vartheta), \quad \text{for all } j \in M_i \text{ with } \vartheta_{ij} > 0,\\ U_i'\big(d_i(\vartheta)\big) &\le \hat{\xi}_{ij}(\vartheta), \quad \text{for all } j \in M_i \text{ with } \vartheta_{ij} = 0,\end{aligned}$$

$$\begin{aligned} U_i'\big(d_i(\psi)\big) &= \sum_{r \in R_{ij}} C_r'\big(\ell_r(\psi)\big), \quad \text{for all } j \in M_i \text{ with } \psi_{ij} > 0,\\ U_i'\big(d_i(\psi)\big) &\le \sum_{r \in R_{ij}} C_r'\big(\ell_r(\psi)\big), \quad \text{for all } j \in M_i \text{ with } \psi_{ij} = 0.\end{aligned}$$

Proof. The function π_i is differentiable and concave with respect to φ_i. Furthermore, the set of profiles Φ is convex. Since ϑ is a Nash equilibrium, the strategy ϑ_i solves $\max_{\varphi_i \in \Phi_i} \pi_i(\varphi_i;\, \vartheta_{-i})$. Thus, we can invoke the variational inequality as a necessary and sufficient optimality condition giving (1). Note that the derivative of π_i with respect to φ_{ij} is given by $\frac{\partial \pi_i}{\partial \varphi_{ij}}(\varphi_i;\, \varphi_{-i}) = U_i'(d_i(\varphi)) - \hat{\xi}_{ij}(\varphi)$. The conditions (2) and (3) follow directly from the Karush-Kuhn-Tucker conditions for the two problems $\max_{\varphi_i \in \Phi_i} \pi_i(\varphi_i;\, \vartheta_{-i})$ and $\max_{\varphi \in \Phi} \mathcal{U}(\varphi)$, respectively. □

Proof of Lemma 4.1.

Let G be a resource allocation games with n players, cost functions in $\mathcal{C}$, and basic cost sharing methods $\xi_r \in \mathcal{D}_n$ for all $r \in R$. Let Θ_G be the set of Nash equilibria. Then, $\mathcal{U}(\vartheta) \ge 0$ for all $\vartheta \in \Theta_G$.

Proof. Let $\vartheta \in \Theta_G$ be a Nash equilibrium. We deduce the following inequalities.

$$\mathcal{U}(\vartheta) \ge \sum_{i=1}^m \sum_{j=1}^{m_i} \hat{\xi}_{ij}(\vartheta) \cdot \vartheta_{ij} - C(\vartheta) = \sum_{r \in R} \sum_{i=1}^n \frac{\partial \xi_i^r}{\partial \vartheta_i^r}(\vartheta) \cdot \vartheta_i^r - C(\vartheta) \ge \sum_{r \in R} \sum_{i=1}^n \xi_i^r(\vartheta) - C(\vartheta) \ge 0\,.$$

Here, the first inequality follows from (2) in Lemma 3.12. The second equality follows by rearranging terms. The third inequality follows from convexity of $\xi_i^r(\vartheta)$ and condition (3) of basic cost sharing methods implying $0 = \xi_i(0, \vartheta_{-i}^r) \ge \xi_i(\vartheta_i^r, \vartheta_{-i}^r) + \frac{\partial \xi_i^r}{\partial \vartheta_i^r}(\vartheta) \cdot (0 - \vartheta_i^r)$. The last inequality follows from the cost covering condition of ξ. □

Proof of Lemma 4.5.

Consider a game G with basic cost sharing methods and linear utility functions, that is, $U_i(x) = u_i \cdot x$, $u_i \geq 0$, $i \in N$. Let ψ be an optimal profile and ϑ be a Nash equilibrium. Then, ψ and ϑ generate a total surplus of

$$\mathcal{U}(\psi) = \sum_{r \in R} \Big(\ell_r(\psi) \cdot C_r'\big(\ell_r(\psi)\big) - C_r\big(\ell_r(\psi)\big)\Big)$$
$$\mathcal{U}(\vartheta) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \hat{\xi}_{ij}(\vartheta)\,\vartheta_{ij} - C(\vartheta)\,.$$

Proof. Using condition (3) from Lemma 3.12, we get $u_i = \sum_{r \in R_{ij}} C_r'\big(\ell_r(\psi)\big)$ for all $i \in N$, $j \in M_i$, with $\psi_{ij} > 0$. By reordering the summation, we obtain

$$\mathcal{U}(\psi) = \sum_{i=1}^{n} u_i \cdot d_i(\psi) - \sum_{r \in R} C_r\big(\ell_r(\psi)\big) = \sum_{r \in R} \Big(\ell_r(\psi) \cdot C_r'\big(\ell_r(\psi)\big) - C_r\big(\ell_r(\psi)\big)\Big)\,,$$

proving the first claim. Using the optimality condition (2) in Lemma 3.12 we get $u_i = \hat{\xi}_{ij}(\vartheta)$ for all $i \in N$, $j \in M_i$, with $\vartheta_{ij} > 0$, proving the second claim. □

Proof of Proposition 4.7.

For incremental cost sharing, every Nash equilibrium is optimal.

Proof. We use Theorem 4.6 as follows. We define $\lambda = 1$ and show that $\omega_n\big(\mathcal{C}, \mathcal{D}_n, \lambda\big) \leq 0$. To see this, we bound the nominator of (4):

$$\sum_{i=1}^{n} \hat{\xi}_i(x)\,(y_i - x_i) + C\big(\ell(x)\big) - C\big(\ell(y)\big) = C'\big(\ell(x)\big)\,\big(\ell(y) - \ell(x)\big) + C\big(\ell(x)\big) - C\big(\ell(y)\big) \leq 0\,,$$

where the last inequality follows from the convexity of C. □

Proof of Lemma 5.2.

Let ξ be a marginal cost pricing method for n players. Then, $\omega_n\big(C, \xi, \lambda\big) \leq \omega_n^{mcp}(C, \lambda)$, where

$$\omega_n^{mcp}(C, \lambda) := \sup_{\substack{x,y \in \mathbb{R}_+ \\ \mu \in \{0\} \cup [\frac{1}{n}, 1]}} \frac{C'(x)y + C''(x)\mu x y - \lambda\Big(C'(x)x + \big(\mu^2 + \frac{(1-\mu)^2}{n-1}\big)C''(x)x^2\Big) + \lambda C(x) - C(y)}{C'(y) \cdot y - C(y)}.$$

Proof. Using the definition of $\omega_n\big(C, \xi, \lambda\big)$, all we have to show is that

$$\sum_{i=1}^{n} \hat{\xi}_i(x)\,(y_i - \lambda\,x_i) \leq \sup_{\mu \in \{0\} \cup [\frac{1}{n}, 1]} \Big\{C'(\ell(x))\,y + C''(\ell(x))\,\mu\,\ell(x)\,\ell(y)$$
$$- \lambda\,(C'(\ell(x))\,\ell(x) + \Big(\mu^2 + \frac{(1-\mu)^2}{n-1}\Big)\,C''(\ell(x))\,\ell(x)^2)\Big\},$$

where $x, y \in R_+^n$, $\ell(x) = \sum_{i=1}^{n} x_i$, $\ell(y) = \sum_{i=1}^{n} y_i$, and ξ is marginal cost pricing with n players and cost function C. Observe that $\hat{\xi}_i(x) = C''(\ell(x))\,x_i + C'(\ell(x))$. By defining $\mu =$

$\max_{i\in N}\left\{\frac{x_i}{\ell(x)}\right\} \in \left[\frac{1}{n}, n\right]$, if $\ell(x) > 0$ and $\mu = 0$, otherwise, we obtain

$$\sum_{i=1}^{n} \hat{\xi}_i(x)\,(y_i - \lambda\, x_i) = C'(\ell(x))\,\ell(y) - \lambda\, C'(\ell(x))\,\ell(x) + C''(\ell(x)) \sum_{i=1}^{n} \Big(x_i\, y_i - \lambda\, x_i^2\Big)$$
$$\leq C'(\ell(x))\,\ell(y) + C''(\ell(x))\mu\,\ell(x)\,\ell(y) - \lambda\Big(C'(\ell(x))\,\ell(x) + \Big(\mu^2 + \frac{(1-\mu)^2}{n-1}\Big)\, C''(\ell(x))\,\ell(x)^2\Big)\,,$$

where the last inequality follows from $\sum_{i\in N} x_i\, y_i \leq \mu\,\ell(x)\,\ell(y)$ and $\sum_{i\in N} x_i^2 \geq (\mu^2 + \frac{(1-\mu)^2}{n-1})\,\ell(x)^2$. Thus, the claim is proven. $\square$

Proof of Lemma 5.7.

Consider the class $\mathcal{M}_d := \Big\{C(z) = a_d\, z^d,\ a_d \geq 0,\ d \in \{2,3,\dots\}\Big\}$. Then, it holds that

$$\omega_\infty^{mcp}(\mathcal{M}_d;\lambda) \leq \left(\frac{1+\mu(d-1)}{\lambda(1+\mu^2 d)}\right)^{d-1} \left(\frac{d}{d-1} + \mu - 1\right) - \frac{1}{d-1}\,,$$

where $\mu(d) = \frac{1}{\sqrt{d-1}+1}$.

Proof. Using the definition of $\omega_\infty^{mcp}(C;\lambda)$ for $C \in \mathcal{M}_d$ we get

$$\omega_\infty^{mcp}(C;\lambda) = \sup_{\mu\in[0,1]}\ \sup_{\beta\geq 0}\ d\,\beta^{d-1}\,\left(\frac{1}{d-1} + \mu\right) - \lambda\,\beta^d\,\left(1 + \mu^2\, d\right) - \frac{1}{d-1}\,.$$

The unique global maximizer with respect to β is $\beta^* = \frac{1+\mu(d-1)}{\lambda(1+\mu^2 d)}$. Thus, since $C \in \mathcal{M}_d$ was arbitrary, we get

$$\omega_\infty^{mcp}(\mathcal{M}_d;\lambda) \leq \sup_{\mu\in[0,1]} \left(\frac{1+\mu(d-1)}{\lambda(1+\mu^2 d)}\right)^{d-1} \left(\frac{d}{d-1} + \mu - 1\right) - \frac{1}{d-1}\,.$$

The unique maximizer for this supremum is given by $\mu(d) = \frac{1}{\sqrt{d-1}+1}$. $\square$

Proof of Proposition 5.10.

Let $\mathcal{C}^{convD}$ be a class of cost functions with a convex derivative. Consider the set $\mathcal{G}_n(\mathcal{C}^{convD})$ of games with at most $n \in \mathbb{N}^*$ players. Then, $\rho_n(\mathcal{C}^{convD}) \leq \frac{2(n-\sqrt{n})}{\sqrt{n}(n-1)}$ which grows as $O(1/\sqrt{n})$.

Proof. Consider the example in the proof of Theorem 5.8 with polynomial cost functions of degree $d \in \mathbb{N}$. Let $\vartheta^d(n)$ and $\psi^d(n)$be the Nash equilibrium and the optimum profile in the game with n players, respectively. We obtain

$$\rho_n(\mathcal{C}^{convD}) \leq \lim_{d\to\infty} \frac{\mathcal{U}(\vartheta^d(n))}{\mathcal{U}(\psi^d)} = \frac{1 + n\,b^2 - b^2}{1 + n\,b - b}\,,\ \forall b \in [0,1]\,.$$

Since $\frac{1+n\,b^2-b^2}{1+n\,b-b}$ has a global minimum with respect to b with value $\frac{2(n-\sqrt{n})}{\sqrt{n}(n-1)}$, the proposition is proved. $\square$

Proof of Lemma 5.11.

Consider a symmetric game with ξ being marginal cost pricing. Then, it holds that $\omega_n\big(C,\xi,\lambda\big) \leq \omega_n^{mcp,\,sym}(C,\lambda)$, where

$$\omega_n^{mcp,\,sym}(C;\lambda) := \sup_{x,y\in\mathbb{R}_+} \frac{C'(x)\,y + C''(x)\,\frac{x\,y}{n} - \lambda\,\big(C'(x)\,x + C''(x)\,\frac{x^2}{n} - C(x)\big) - C(y)}{C'(y)\cdot y - C(y)}.$$

Proof. The proof is analogous to the proof of Lemma 5.2, except that for symmetric games we use a symmetric optimal profile which implies $\sum_{i\in N} x_i y_i = \frac{\ell(x)\ell(y)}{n}$. Moreover, using $\sum_{i\in N} x_i^2 \geq \frac{\ell(x)^2}{n}$ the claim follows. □

Proof of Proposition 5.12.

Let $\mathcal{C}^{convD}$ be the class of cost functions with a convex derivative. Consider the set $\mathcal{G}_n(\mathcal{C}^{convD})$ of symmetric games with at most $n \in \mathbb{N}^*$ players. Then, $\rho_n(\mathcal{C}^{convD}) \geq \frac{2\,n}{2\,n+1}$.

Proof. The proof proceeds along the lines of the proof of Theorem 5.3, except that $\lambda = \frac{1+2\,n}{2\,n}$ and the values μ is replaced by $\frac{1}{n}$. Then, the only interesting difference occurs for the case $x < y$ in evaluating the following maximum:

$$\max_{\beta\in[0,1)} \left(\frac{\beta}{n} - \frac{\lambda\,\beta^2}{n} - \frac{(1-\beta)^2}{2}\right) \leq \frac{1+2\,n-2\,n\,\lambda}{2\,n\,(2\,\lambda+n)}\,.$$

Thus, since $\lambda = \frac{1+2\,n}{2\,n}$, the claim is proven. □

Proof of Theorem 5.13.

Let $\mathcal{C}_d$ be the class of polynomial cost function with non-negative coefficients and arbitrary degree $d \in \{2,3,\ldots\}$. Consider the set $\mathcal{G}_n(\mathcal{C}_d)$ of symmetric games with n players and cost functions in $\mathcal{C}_d$. Then, $\rho_n(\mathcal{C}_d) = \frac{3}{4}$.

Proof. Let ϑ be a Nash equilibrium profile and ψ the system optimum. Using Remark 5.6, it is sufficient to consider monomial cost functions $C_{d'}(z) = a_{d'}\,z^{d'}$, $a_j \geq 0$, for some $d' \in \{2,3,\ldots\}$. Then, we obtain

$$\omega_n^{mcp,\,sym}(C_{d'};\lambda) \leq \sup_{\beta\geq 0}\, d'\left(\frac{1}{d'-1}+\frac{1}{n}\right)\beta^{d'-1} - \lambda\,\beta^{d'}\left(1+\frac{d'}{n}\right) - \frac{1}{d'-1}\,.$$

The unique maximizer is $\beta^* = \frac{n+d'-1}{\lambda\,(n+d')}$. Thus, we get

$$\omega_n^{mcp,\,sym}(C_{d'};\lambda) \leq \left(\frac{n+d'-1}{\lambda\,(n+d')}\right)^{d'-1}\left(\frac{n+d'-1}{n\,(d'-1)}\right) - \frac{1}{d'-1}\,.$$

We define $\lambda = \lambda(d',n) = \left(\frac{n+d'-1}{n+d'}\right)\left(\frac{n}{n+d'-1}\right)^{-\frac{1}{d'-1}}$ implying $\omega_n^{mcp,\,sym}\big(C_{d'};\lambda(d',n)\big) = 0$. Thus, applying Lemma 5.11 and Theorem 4.6 yields $\mathcal{U}(\psi) \leq \lambda(d',n)\,\mathcal{U}(\vartheta)$ for a Nash equilibrium ϑ and optimal profile ψ. We now observe that $\lambda(d',n)$ is a decreasing function in d' and n. Hence, the worst case occurs for $d' = 2$ and $n = 1$ leading to the desired bound of $3/4$.

To prove that the bound is tight, we consider a resource allocation game with a single resource and cost function $C(z) = \frac{1}{2}z^2$. We consider n players with utility functions $U(\varphi_i) = \varphi_i$. Then, the following conditions hold for a Nash equilibrium ϑ: $1 - \big(\ell(\vartheta) + \vartheta_i\big) = 0 \;\Rightarrow\; \vartheta_i = 1 - \ell(\vartheta)$. Hence, we have: $\ell(\vartheta) = n\,\vartheta_i = n\,\big(1-\ell(\vartheta)\big) \;\;\Rightarrow\;\; \ell(\vartheta) = \frac{n}{n+1}$. The total surplus evaluates to

$\mathcal{U}(\vartheta) = \frac{n}{n+1} - \frac{1}{2}\frac{n^2}{(n+1)^2}$. The optimal profile ψ has value 1 and its total surplus evaluates to $\mathcal{U}(\psi) = \frac{1}{2}$. Evaluating the ratio $\frac{\mathcal{U}(\vartheta)}{\mathcal{U}(\psi)}$ proves the claim. $\square$

Proof of Theorem 6.1.

In order to prove Theorem 6.1, we first present the following two lemmata. The first one establishes closed-form expressions of the total surplus of a Nash equilibrium and an optimal profile.

Lemma 8.1. Consider a game G in which utility functions are linear, that is, $U_i(x) = u_i \cdot x$, $u_i \geq 0$, $i \in N$. Let ψ be an optimal profile and ϑ be a Nash equilibrium. Then, ψ and ϑ generate total surplus of $\mathcal{U}(\psi) = \sum_{r\in R} \big(\ell_r(\psi)\big)^2 \cdot c_r'\big(\ell_r(\psi)\big)$ and $\mathcal{U}(\vartheta) = \sum_{r\in R}\sum_{i=1}^n \big(\vartheta_i^r\big)^2 \cdot c_r'\big(\ell_r(\vartheta)\big)$, where $c_r(\ell(\varphi)) = \frac{C_r(\ell(\varphi))}{\ell(\varphi)}$ is the per-unit cost function.

The proof follows from Lemma 4.5. Next, along the lines of Section 5, we derive an upper bound of the quantity $\omega_n\big(C,\xi,\lambda\big)$, where ξ corresponds to average cost sharing.

Lemma 8.2. For ξ being average cost pricing, it holds $\omega_n\big(C,\xi,\lambda\big) \leq \omega_n^{avg}(C,\lambda)$, where

$$\omega_n^{avg}(C,\lambda) := \sup_{x,y\in\mathbb{R}_+}\ \sup_{\mu\in\{0\}\cup[\frac{1}{n},1]} \frac{c(x)\,y + c'(x)\,x\,y\,\mu - c(y)\,y - \lambda\left(\mu^2 + \frac{(1-\mu)^2}{n-1}\right)c'(x)\,x^2}{c'(y)\,y^2}. \tag{12}$$

Proof. The proof proceeds along the lines of the proof of Lemma 5.2. Using the definition of $\omega_n\big(C,\xi,\lambda\big)$, it remains to show that

$$\begin{aligned}
&\sum_{i=1}^n \hat{\xi}_i(x)\,(y_i - \lambda\,x_i) + \lambda\,\ell(x)\,c(\ell(x)) \\
&\leq \sup_{\mu\in\{0\}\cup[\frac{1}{n},1]} c(\ell(x))\,y + c'(\ell(x))\,\ell(x)\,\ell(y)\,\mu - \lambda\left(\mu^2 + \frac{(1-\mu)^2}{n-1}\right)c'(\ell(x))\,\ell(x)^2\,.
\end{aligned}$$

where $x, y \in R_+^n$, $\ell(x) = \sum_{i=1}^n x_i$, $\ell(y) = \sum_{i=1}^n y_i$, ξ is average cost pricing with n players and cost function C, and $c(\ell(x)) = \frac{C(\ell(x))}{\ell(x)}$. Using the definition of average cost pricing, this reduces to

$$\sum_{i\in N} x_i\,y_i - \lambda \sum_{i\in N} x_i^2 \leq \sup_{\mu\in\{0\}\cup[\frac{1}{n},1]} \ell(x)\,\ell(y)\,\mu - \lambda\left(\mu^2 + \frac{(1-\mu)^2}{n-1}\right)\ell(x)^2\,.$$

Observe that setting $\mu = \max_{i\in N}\left\{\frac{x_i}{\ell(x)}\right\} \in \left[\frac{1}{n}, n\right]$, if $\ell(x) > 0$ and $\mu = 0$, otherwise, we obtain $\sum_{i\in N} x_i\,y_i \leq \mu\,\ell(x)\,\ell(y)$ and $\sum_{i\in N} x_i^2 \geq \left(\mu^2 + \frac{(1-\mu)^2}{n-1}\right)\ell(x)^2$ proving the claim. $\square$

Now we prove the theorem.

Let $\mathcal{C}^{conv}$ be a class of convex cost functions. Consider the set $\mathcal{G}_n(\mathcal{C}^{conv})$ of games with at most $n \in \mathbb{N}$ players. Then, $\rho_n(\mathcal{C}^{conv}) \geq \frac{1}{n}$.

Proof. We first need the following simple observation

$$\omega_n^{avg}(C;\lambda) \leq \sup_{x,y\in\mathbb{R}_+} \frac{c(x)\,y + c'(x)\,x\,y - c(y)\,y - \frac{\lambda}{n}\,c'(x)\,x^2}{c'(y)\,y^2}\,.$$

Then, we define $\lambda = n$ and prove the claim by showing $\omega_n^{avg}(C;\lambda) \leq 0$ for $C \in \mathcal{C}^{conv}$. Let $T_x(y) = c(x)\,x + (c(x)\,x)'\,(y-x)$ be the supporting tangent of $c(y)\,y$ in x. Using $T_x(y) \leq c(y)\,y$ for all $y \geq 0$ (note that $c(x)\,x$ is a convex function), we obtain

$$c(x)\,y + c'(x)\,x\,y - c(y)\,y - c'(x)\,(x)^2 = T_x(y) - c(y)\,y \leq 0.$$

□

Proof of Theorem 6.2.

Let $\mathcal{C}^{convU}$ be the class of cost functions with convex unit costs. Then, $\rho(\mathcal{C}^{convU}) = \frac{4}{n+3}$.

Proof. First, we define $\lambda = \frac{n+3}{4}$ and prove $\rho(\mathcal{C}^{convU}) \geq \frac{4}{n+3}$ by showing $\omega_n^{avg}(C;\lambda) \leq 0$.

We denote the nominator of (12) by $N(x,y,\mu;C,\lambda)$. Using $T_x(y) = c(x)+c'(x)\,(y-x) \leq c(y)$ (using that c is convex), we obtain

$$N(x,y,\mu;C,\lambda) \leq c'(x)\,\left(x\,y\,(1+\mu) - y^2 - \lambda\,\left(\mu^2 + \frac{(1-\mu)^2}{n-1}\right)\,x^2\right)\,.$$

Then, since for $y \in \mathbb{R}_+$ it holds

$$x\,y\,(1+\mu) - y^2 \leq \frac{(1+\mu)^2}{4\,n}\,x^2\,,$$

we obtain

$$N(x,y,\mu;C,\lambda) \leq c'(x)\,x^2\,\left(\frac{(1+\mu)^2}{4\,n} - \lambda\,\left(\mu^2 + \frac{(1-\mu)^2}{n-1}\right)\right)\,.$$

Finally, the inequality

$$\frac{(1+\mu)^2}{4\,n} - \lambda\,\left(\mu^2 + \frac{(1-\mu)^2}{n-1}\right) \leq \frac{\lambda\,(n-4\,\lambda+3)}{1+4\,\lambda\,n-n}$$

for $\lambda = \frac{n+3}{4}$ proves the claim.

To prove $\rho(\mathcal{C}^{convU}) \leq \frac{4}{n+3}$ consider the following example. Assume n users share a single resource with the cost function $C(\ell(\varphi)) = \ell(\phi)$. Further, assume user 1 has the utility function $U_1(\varphi_1) = a\,\varphi_1$, while users $i = 2,\dots,n$ have utility functions $U_i(\varphi_i) = \frac{1-a}{n-1}\,\varphi_i$, with $a = \frac{n+3}{(n+1)^2}$. A system optimum is achieved when all of the resource is allocated to user 1 resulting in a total utility of $\mathcal{U}(\psi) = \frac{a^2}{4}$. A Nash equilibrium is $\vartheta_1 = a - \frac{1}{n+1}$, $\vartheta_i = \frac{1-a}{n-1} - \frac{1}{n+1}$ for $i = 2,\dots,n$. We obtain a total utility of $\mathcal{U}(\vartheta) = a^2 + \frac{(1-a)^2}{n-1} - \frac{n+2}{(n+1)^2}$. Thus, relative efficiency is $\frac{\mathcal{U}(\vartheta)}{\mathcal{U}(\psi)} = \frac{4}{n+3}$. □

Proof of Theorem 6.4.

Let $\mathcal{C}^{convU}$ be the class of cost functions with convex unit costs. Consider the set $\mathcal{G}_n(\mathcal{C}^{convD})$ of symmetric games with at most $n \in \mathbb{N}^*$ players. Then, $\rho(\mathcal{C}^{convD}) = \frac{4\,n}{(n+1)^2}$.

Proof. Using similar arguments as in the proof of Lemma 5.11, we obtain

$$\omega_n(C,\xi,\lambda) \leq \omega_n^{avg,\,sym}(C;\lambda) := \sup_{x,y\in\mathbb{R}_+^2} \frac{c(x)\,y + c'(x)\,\frac{x\,y}{n} - c(y)\,y - \frac{\lambda}{n}\,c'(x)\,x^2}{c'(y)\,y^2}\,.$$

We define $\lambda = \frac{(n+1)^2}{4}$ and prove the theorem by showing $\omega_n^{avg,\,sym}(C;\lambda) \leq 0$. With $T_x(y) = c(x) + c'(x)\,(y-x) \leq c(y)$ (using that c is convex), we obtain

$$c(x)\,y + c'(x)\,\frac{x\,y}{n} - c(y)\,y - \frac{\lambda}{n}\,c'(x)\,(x)^2 \leq c'(x)\,\left(x\,y\,\left(1+\frac{1}{n}\right) - y^2 - \frac{\lambda}{n}\,x^2\right)\,.$$

Then, we use that for $y \in \mathbb{R}_+$ it holds that $x\,y\,(1+\frac{1}{n}) - y^2 \leq \frac{(1+1/n)^2}{4\,n}x^2$. The choice of λ, thus, proves the claim. The upper bound follows by a simple construction and is omitted. □

Paper 4

Optimal Cost Sharing Protocols for Scheduling Games

Philipp von Falkenhausen and Tobias Harks
Optimal Cost Sharing Protocols for Scheduling Games
In Proc. of the 12th ACM Conference on Electronic Commerce (EC), 2011, pp. 285-294
Full version to appear in Mathematics of Operations Research.

Abstract. We consider the problem of designing *cost sharing protocols* to minimize the price of anarchy and stability for a class of scheduling games. Here, we are given a set of players, each associated with a job of certain non-negative weight. Any job fits on any machine, and the cost of a machine is a non-decreasing function of the total load on the machine. We assume that the private cost of a player is determined by a cost sharing protocol. We consider four natural design restrictions for feasible protocols: *stability, budget balance, separability,* and *uniformity.* While budget balance is self-explanatory, the stability requirement asks for the existence of pure-strategy Nash equilibria. Separability requires that the resulting cost shares only depend on the set of players on a machine. Uniformity additionally requires that the cost shares on a machine are instance-independent, that is, they remain the same even if new machines are added to or removed from the instance. We call a cost sharing protocol basic, if it satisfies only stability and budget balance. Separable and uniform cost sharing protocols additionally satisfy separability and uniformity, respectively. For n-player games we show that among all basic and separable cost sharing protocols, there is an optimal protocol with price of anarchy and stability of precisely $\mathcal{H}_n = \sum_{i=1}^{n} 1/i$. For uniform protocols we present a strong lower bound showing that the price of anarchy is unbounded. Moreover, we obtain several results for special cases in which either the cost functions are restricted, or the job sizes are restricted. As a byproduct of our analysis, we obtain a complete characterization of outcomes that can be enforced as a pure-strategy Nash equilibrium by basic and separable cost sharing protocols.

1. Introduction

Congestion games play a fundamental role for many applications, including traffic networks, telecommunication networks and economics. In a congestion game, there is a set of resources and a pure strategy of a player consists of a subset of resources. The cost of a resource depends only on the number of players choosing the resource, and the private cost of a player is the sum of the costs of the chosen resources. Under these assumptions, Rosenthal proved the existence of a pure Nash equilibrium (PNE for short) [20]. An important question in congestion games is the degree of suboptimality caused by selfish resource allocation. Koutsoupias and Papadimitriou [17] introduced a measure to quantify the inefficiency of Nash equilibria which they termed the *price of anarchy*. The price of anarchy is defined as the worst-case ratio of the cost of a Nash equilibrium over the cost of a system optimum. In the past decade, considerable progress has been made in exactly quantifying the price of anarchy for many interesting classes of games. In the context of nonatomic network routing games, the price of anarchy for specific classes of cost functions is well understood, see Roughgarden and Tardos [23], Roughgarden [21] and Correa, Schulz, and Stier-Moses [11]. (For an overview of these results, we refer to the book by Roughgarden [22].) Awerbuch et al. [4], Christodoulou and Koutsoupias [9], Aland et al. [2] and Bhawalkar et al. [5] derived several tight bounds on the price of anarchy for weighted and unweighted congestion games with specific classes of latency functions. Despite these bounds, it is known that the price of anarchy for general latency functions is unbounded even on simple *parallel-arc* networks [5, 23].

Motivated by the fact that pure Nash equilibria may be very inefficient even in parallel-arc networks, we focus in this paper on the design of *cost sharing methods* as a means to leverage the resulting price of anarchy. The concrete scenario that we consider is the problem of scheduling jobs on parallel machines. We are given a set of players, each associated with a job of certain non-negative weight. Any job fits on any machine, and the cost of a machine is a non-decreasing function of the total load on the machine. We assume that the private cost of a player is determined by a cost sharing method. For instance, a simple cost sharing method that has been analyzed in the aforementioned literature is *average cost sharing*, see [5, 11, 23]. In almost all settings in the theory and practice of mechanism or protocol design, a designer may only choose protocols out of a set of *feasible* protocols. Therefore, we have to precisely define the design space of feasible protocols. To this end, we define the following four properties listed below which are defined more formally in Section 2. These properties have been introduced first by Chen et al. [8] in the context of the design of cost sharing protocols for network design games (a formal definition of these games will be given in Section 1.2).

(1) *Stability*: There is at least one pure strategy Nash equilibrium in each scheduling game induced by the cost sharing protocol.
(2) *Budget-balance*: For every outcome of a scheduling game induced by the cost sharing protocol, the cost of each resource is exactly covered by the collected cost shares of the players using the resource.
(3) *Separability*: In each scheduling game induced by the cost sharing protocol, the cost shares of each resource are completely determined by the set of players that use it.
(4) *Uniformity*: Across all scheduling games induced by the cost sharing protocol, the cost shares of a resource (for each potential set of users) depend only on the resource cost, and not on the set of available resources.

A cost sharing method is called *basic* if it satisfies (1)-(2), *separable* if it satisfies (1)-(3), and *uniform* if it satisfies (1)-(4). We briefly discuss the above four properties and refer to [8] for a more detailed treatment. The condition (2) is the least controversial in the context of cost sharing protocols. The stability condition (1) requires the existence of at least one Nash equilibrium in pure strategies. While this requirement restricts the search space for cost sharing protocols, it is certainly the solution concept of choice when mixed or correlated strategies have no meaningful physical interpretation in the game played; see also the discussion in Osborne and Rubinstein [19, § 3.2] about critics of mixed Nash equilibria. While condition (3) seems restrictive, it is crucial for practical applications in which cost sharing methods have only local information about their own resource usage (see for instance the TCP/IP protocol design, where routers drop packets based on some function of the number of packets in the queue, see [24]). Uniformity (4) is the strongest and perhaps the most problematic design restriction. A uniform protocol is not only separable but also *strongly local* in the sense that the cost shares of a resource are independent of the set of resources available to the game designer. This property may be crucial for systems in which the resources can be added or removed over time and a reconfiguration of the system (changing the cost sharing protocol) is too costly.

The goal of this paper is twofold. On the one hand side, we want to systematically analyze the achievable worst-case efficiency of Nash equilibria by basic, separable and uniform cost sharing protocols in the context of scheduling games. Besides this worst-case perspective, we also ask a larger question: Which outcomes can actually be enforced as pure Nash equilibria? More precisely, we call an outcome of a scheduling game *weakly-enforceable* if there is a basic protocol that induces the outcome as a pure Nash equilibrium. We call an outcome of a scheduling game *strongly-enforceable* if there is a basic protocol that induces the outcome as the most expensive pure Nash equilibrium.

1.1. Our Results

We study protocol design problems in the context of scheduling games, where the goal is to minimize the induced price of anarchy and the price of stability. Our results for these problems can be summarized as follows.

Among all basic and separable protocols, we provide an *optimal* protocol minimizing the resulting price of anarchy and price of stability simultaneously. For n-player scheduling games, the optimal value of the price of anarchy and stability is precisely the n-th harmonic number $\mathcal{H}_n = \sum_{i=1}^{n} 1/i$. Moreover, we obtain a complete characterization of weakly-enforceable outcomes. This characterization can be used for designing cost sharing protocols minimizing the price of stability with respect to an arbitrary objective function. We also derive sufficient conditions for an outcome to be strongly-enforceable. Our proof of this result is constructive by providing a cost sharing protocol that strongly enforces an outcome satisfying the sufficient conditions. We then show that this protocol gives rise to an optimal cost sharing protocol minimizing the price of anarchy and stability as mentioned above. For scheduling games with cost functions that have non-decreasing per-unit costs, we derive an optimal cost sharing protocol with price of anarchy equal to 1. We remark that this assumption is quite weak insofar as non-decreasing and convex cost functions satisfy non-decreasing per-unit costs.

We also study the achievable price of anarchy of uniform cost sharing protocols. We show that there is no uniform cost sharing protocol with a bounded price of anarchy. This bound even holds for a family of instances with only 3 players, at most 3 machines and cost functions with non-decreasing costs per unit. Only for instances in which the demands are integer multiples of each other, we present a cost sharing protocol with a bounded price of anarchy of n.

1.2. Related Work

There is a large body of work on scheduling games (or singleton congestion games) with unweighted and weighted players [1, 12, 13, 14, 15, 18]. Most of these papers study the existence and price of anarchy of pure Nash equilibria for the uniform cost sharing protocol in which the private cost of every player is equal to the cost on the resource. These works, however, do not consider the design perspective of cost sharing protocols. Christodoulou et al. [10] and follow-up papers such as [6, 16] study *coordination mechanisms* and their price of anarchy in scheduling games in which n players assign a task to one of m machines. Rather than paying a share of the resulting cost of a machine as in our scenario, the players in these games consider the completion time of their respective job as private cost. This completion time depends on the sequence in which the jobs on a machine are processed which in turn is given by the *coordination mechanism*.

The notion of private cost in these papers establishes an entirely different set of Nash equilibria compared to our work and hence their results concerning the price of anarchy are unrelated to ours.

Our work is motivated by the paper by Chen et al. [8]. In this paper, the authors study the design of cost sharing protocols for network design games, see also Anshelevich et al. [3] and Chen and Roughgarden [7] for earlier work on network design games. In a network design game, each player i has a unit demand that she wishes to send along a path in a (directed or undirected) network connecting her source node s_i to her terminal node t_i. Every edge has a constant non-negative cost and the problem is to design a separable or uniform cost sharing protocol so as to minimize the price of anarchy and stability in this setting. Our approach follows their lead in terms of the feasible protocol space, but we apply cost sharing protocols to the structural different class of scheduling models. On the one hand, such scheduling models are more general in the sense that we allow arbitrary non-decreasing cost functions instead of constant costs on the resources. Moreover, in contrast to [8], we allow players to have different non-negative weights. On the other hand, scheduling models are more restricted in the sense that we consider a relatively simple strategy space for the players, that is, a pure strategy for a player is simply a single resource. Moreover, in contrast to [8], our games are symmetric, that is, every player has access to every resource. These structural differences result in different approaches and also the results of [8] are different to ours. For example, while Chen et al. [8] proved bounds on the price of anarchy for uniform protocols of order $\Theta(\log(n)), \Theta(\text{polylog}(n))$, and n for undirected single-sink instances, undirected multi-commodity instances, and directed single-sink instances, respectively, we show that for scheduling games such results are impossible. The price of anarchy for uniform protocols inducing scheduling games is unbounded. Finally, it is worth noting that, while [8] analyzed separable and uniform protocols, we additionally analyze the larger class of basic protocols.

2. Model and Problem Statement

A *scheduling model* is represented by a tuple (N, M, d, c). Here, $N = \{1, \dots, n\}$ is a nonempty set of players and $M = \{a_1, \dots, a_m\}$ is a nonempty set of machines. Every player is associated with a task of weight d_i and $d = (d_1, \dots, d_n)$ is the combined weight vector. Every machine $a \in M$ has an associated non-negative and non-decreasing cost function $c_a : \mathbb{R}^+ \to \mathbb{R}^+$. We assume $c_a(0) = 0$ for all $a \in M$. The vector of cost functions is denoted by $c = (c_1, \dots, c_m)$. Given a scheduling model (N, M, d, c), we associate a strategic game represented by the tuple (N, X, ξ). Here, it is assumed that every task fits on every machine, thus, the set of pure strategies for player $i \in N$ is $X_i = M$ and the overall strategy space is $X = M^n$. The outcomes $x = (x_1, \dots, x_n) \in M^n$ are vectors of machines where the strategy played by player i is machine x_i. The private cost of player $i \in N$ in such an outcome x is determined by the *cost sharing method* $\xi_i : X \to \mathbb{R}^+$. A *cost sharing protocol* $\Xi : (N, M, d, c) \mapsto \xi$ provides every scheduling model with a vector $\xi = (\xi_i)_{i \in N}$ of such cost sharing methods. For $x \in M^n$, the set of players using some machine $a \in M$ is denoted by $S_a(x) := \{i \in N : x_i = a\}$ and for $a \in M$, the load on machine a is defined as $\ell_a(x) := \sum_{i \in S_a(x)} d_i$. The cost of an outcome is defined as $C(x) := \sum_{a \in M} c_a(\ell_a(x))$. Abusing notation, we will often write $c_a(x)$ instead of $c_a(\ell_a(x))$. We consider cost minimization games, thus, when choosing her strategy, each player strives to minimize her resulting private cost $\xi_i(x)$. We say that the game (N, X, ξ) on a scheduling model (N, M, d, c) is induced by the protocol Ξ.

An important solution concept in non-cooperative game theory for the analysis of strategic games are pure Nash equilibria. Using standard notation in game theory, for an outcome $x \in M^n$ we denote by

$$(a, x_{-i}) := (x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_n) \in M^n$$

the outcome that arises if only player i deviates to strategy a.

Definition 2.1 (Pure Nash Equilibrium)**.** Let (N, X, ξ) be a scheduling game. The outcome x is a pure Nash equilibrium if no player i can strictly reduce her private cost by unilaterally moving to a different machine, that is, for all $i \in N$

$$\xi_i(x) \leq \xi_i(a, x_{-i}) \text{ for all } a \in M. \tag{1}$$

Two well established concepts that quantify the efficiency of Nash equilibria are the *price of anarchy* and the *price of stability*. The price of anarchy measures the largest possible ratio of the cost of a Nash equilibrium and the cost of an optimal outcome. The price of stability measures the smallest ratio of the cost of a Nash equilibrium and the cost of an optimal outcome. For a cost sharing protocol Ξ, we define by $PoA(\Xi)$ and $PoS(\Xi)$ the corresponding worst case price of anarchy and price of stability across games induced by protocol Ξ. The main goal of this paper is to design cost sharing protocols that minimize the price of anarchy and price of stability, respectively. Of course, the attainable objective values crucially depend on the design space that we permit. The following properties have been first proposed by Chen et al. [8] in the context of designing cost sharing methods for network design games.

Definition 2.2 (Properties of cost sharing protocols)**.** A cost sharing protocol Ξ is

(1) *stable* if it induces only games that admit at least one pure Nash equilibrium.

(2) *basic* if it is stable and additionally *budget balanced*, i.e. if it assigns all scheduling models (N, M, d, c) with cost sharing methods $(\xi_i)_{i \in N}$ such that

$$c_a(x) = \sum_{i \in S_a(x)} \xi_i(x) \text{ for all } a \in M, x \in M^n. \tag{2}$$

This property requires $c_a(0) = 0$ for unused machines, which we will assume in the paper.

(3) *separable* if it is basic and if it induces only games (N, M^n, ξ) for which in any two outcomes $x, x' \in M^n$

$$S_a(x) = S_a(x') \Rightarrow \xi_i(x) = \xi_i(x') \; \forall i \in S_a(x), a \in M.$$

(4) *uniform* if it is separable and if it assigns any two models (N, M, d, c), (N, M', d, c') with cost sharing methods $(\xi_i)_{i \in N}$ and $(\xi'_i)_{i \in N}$ such that the following condition holds. For all $a \in M \cap M'$ with $c_a = c'_a$ and all outcomes $x \in M^n$, $x' \in M'^n$

$$S_a(x) = S_a(x') \Rightarrow \xi_i(x) = \xi'_i(x') \text{ for all } i \in S_a(x).$$

Informally, separability means that in an outcome x the value $\xi_i(x)$ depends only on the set $S_{x_i}(x)$ of players sharing machine x_i and disregards all other information contained in x. Still, separable protocols can assign cost sharing methods that are specifically tailored to the given scheduling model, for example based on an optimal outcome. Uniform protocols are not allowed to do this, they even disregard the layout of the model and assign the same cost sharing methods when machines are added to or removed from the model.

We denote by $\mathcal{B}_n, \mathcal{S}_n$ and $\mathcal{U}_n$ the set of basic, separable and uniform cost sharing protocols for scheduling games with n players, respectively. We obtain the following optimization problems that we address in this paper.

$$\min_{\Xi \in \mathcal{B}_n} PoA(\Xi), \; \min_{\Xi \in \mathcal{B}_n} PoS(\Xi), \; \min_{\Xi \in \mathcal{S}_n} PoA(\Xi), \; \min_{\Xi \in \mathcal{S}_n} PoS(\Xi),$$

$$\min_{\Xi \in \mathcal{U}_n} PoA(\Xi), \; \min_{\Xi \in \mathcal{U}_n} PoS(\Xi) \; .$$

3. Basic and Separable Protocols

We start with studying basic and separable cost sharing protocols. While our goal is to find a cost sharing protocol minimizing the induced PoA and PoS, we first study the issue of enforceability of pure Nash equilibria by basic and separable cost sharing protocols. To be more precise, given a scheduling model (N, M, d, c), we first ask which outcomes $x \in M^n$ can be enforced as pure Nash equilibria by some basic or separable cost sharing protocol. We will differentiate between *weakly* enforceable outcomes and *strongly* enforceable outcomes, see the definition below.

Definition 3.1 (Enforceable outcomes)**.** Consider a scheduling model (N, M, d, c) and an outcome $x \in M^n$.

i) x is *weakly-enforceable* if there exists a basic cost sharing protocol Ξ such that x is a Nash equilibrium in the game (N, M^n, ξ) induced by Ξ.
ii) x is *separable weakly-enforceable* if there exists a separable cost sharing protocol Ξ such that x is a Nash equilibrium in the game (N, M^n, ξ) induced by Ξ.
iii) x is *strongly-enforceable* if there exists a separable cost sharing protocol Ξ such that x is the most expensive Nash equilibrium in the game (N, M^n, ξ) induced by Ξ, i.e. $C(x') \leq C(x)$ for all Nash equilibria $x' \in M^n$.

In the following section, we will give an exact characterization of weakly-enforceable and separable weakly-enforceable outcomes. This characterization provides a structural property that can be used to design cost sharing protocols for minimizing the price of stability for *arbitrary* objective functions.

Throughout this section, the players are assumed to be ordered by non-decreasing weights:

$$d_1 \leq d_2 \leq \cdots \leq d_n. \tag{3}$$

3.1. Weakly-Enforceable Outcomes

This section provides an exact characterization of weakly-enforceable outcomes. Our characterization relies on the notion of *decharged outcomes* defined below.

Definition 3.2 (Weakly decharged outcome)**.** Consider a scheduling model (N, M, d, c). A machine $a \in M$ is *weakly decharged* in an outcome $x \in M^n$ if

$$c_a(x) \leq \sum_{i \in S_a(x)} \min_{b \in M} c_b(b, x_{-i}). \tag{4}$$

The outcome x itself is called *weakly decharged* if all machines are weakly decharged.

We further introduce the weak x-enforcing protocol.

Definition 3.3 (Weak x-enforcing protocol)**.** The *weak x-enforcing protocol* takes as input a weakly decharged outcome x. We use x to define for any outcome z and machine a the sets $S_a^0(z) := \{i \in S_a(z) \cap S_a(x)\}$ (*home players on a*) and $S_a^1(z) := \{i \in S_a(z) \backslash S_a(x)\}$ (*foreign players on a*). Then, the weak x-enforcing protocol assigns for all $i \in N, z \in M^n$ the following cost sharing methods

$$\xi_i(z) := \begin{cases} \dfrac{\min\limits_{b \in M} c_b(b, x_{-i})}{\sum\limits_{j \in S_{x_i}(x)} \min\limits_{b \in M} c_b(b, x_{-j})} \cdot c_{x_i}(x), & \text{if } S_{z_i}(z) = S_{z_i}(x) \text{ and } c_{x_i}(x) > 0, \\ c_{z_i}(z), & \text{if } S_{z_i}^1(z) \neq \emptyset \text{ and } i = \min S_{z_i}^1(z), \\ c_{z_i}(z), & \text{if } S_{z_i}^1(z) = \emptyset,\ S_{z_i}(z) \subset S_{z_i}(x) \text{ and } i = \min S_{z_i}(z), \\ \quad 0, & \text{else.} \end{cases}$$

Informally, if $S_a(z) = S_a(x)$, the players on machine a share the cost proportional to their opportunity cost (cost of change) in outcome x. Otherwise, the smallest foreign player (deviating from outcome x) or, if there are none, the smallest home player (not deviating) pays the entire cost of the machine. Observe that in weakly decharged outcomes x we have

$$\sum_{j \in S_a(x)} \min_{b \in M} c_b(b, x_{-j}) > 0 \quad \text{for all } a \in M \text{ with } c_a(x) > 0$$

and thus the protocol is well defined. We are now ready to state our first main result.

Theorem 3.4. For any scheduling model (N, M, d, c) and outcome x, the following statements are equivalent.

(i) the outcome x is weakly decharged,
(ii) the outcome x is weakly-enforceable,
(iii) the outcome x is separable weakly-enforceable.

Observe that (iii) $\Rightarrow$ (ii) holds because by definition separable protocols are a subclass of basic protocols. We prove (i) $\Rightarrow$ (iii) and (ii) $\Rightarrow$ (i) by two lemmas.

Lemma 3.5. For every weakly decharged outcome x, the weak x-enforcing protocol is a separable cost sharing protocol and weakly enforces x.

Proof. Budget balance and separability of the cost sharing methods are clear from the definition of the protocol, thus, we prove only that x is a Nash equilibrium. For all machines $a \in M$ with $c_a(x) > 0$ we are in the first case of the definition of the protocol, thus, we obtain

$$\begin{aligned} \xi_i(x) &= \frac{\min_{b \in M} c_b(b, x_{-i})}{\sum_{j \in S_a(x)} \min_{b \in M} c_b(b, x_{-j})} \cdot c_a(x) \\ &\le \min_{b \in M} c_b(b, x_{-i}) \le \min_{b \in M \setminus \{a\}} \xi_i(b, x_{-i}) \text{ for all } i \in S_a(x), \end{aligned}$$

where the first inequality holds because outcome x is weakly decharged. For all other machines $a \in M$, we have $\xi_i(x) = c_a(x) = 0$ for all $i \in S_a(x)$ and thus x is a pure Nash equilibrium. □

We prove (ii) $\Rightarrow$ (i) from Theorem 3.4 by the following lemma.

Lemma 3.6. Consider the scheduling model (N, M, d, c). Then, any weakly-enforceable outcome x is weakly decharged.

Proof. Say x is a Nash equilibrium under the basic protocol Ξ that assigns cost sharing methods ξ. Then $\xi_i(x) \le \min_{b \in M} \xi_i(b, x_{-i})$ for all $i \in N$ and hence due to budget balance of Ξ,

$$\begin{aligned} c_a(x) &= \sum_{i \in S_a(x)} \xi_i(x) \le \sum_{i \in S_a(x)} \min_{b \in M} \xi_i(b, x_{-i}) \\ &\le \sum_{i \in S_a(x)} \min_{b \in M} c_b(b, x_{-i}) \end{aligned}$$

for all machines $a \in M$. Thus, x is weakly decharged. □

The above characterization has a direct consequence for the design of cost sharing protocols so as to minimize the price of stability with respect to an arbitrary objective function over the strategy space. As formalized below, by Theorem 3.4 this problem reduces to solving a well-structured finite-dimensional optimization problem.

Corollary 3.7. Let (N, M, d, c) be a scheduling model and let $F : M^n :\to \mathbb{R}$ be a social welfare function. Then, $\min_{\xi \in \mathcal{B}_n} PoS(\xi; F)$ and $\min_{\xi \in \mathcal{S}_n} PoS(\xi; F)$ can be reduced to solving the optimization problem

$$\min_{x \in M^n} F(x) \;\; \text{s.t. } c_a(x) \leq \sum_{i \in S_a(x)} \min_{b \in M} c_b(b, x_{-i}) \, \forall \, a \in M.$$

3.2. Strongly-Enforceable Outcomes

In this section, we turn to strongly-enforceable outcomes. We present a slightly extended protocol that we term the the strong x-enforcing protocol. We will show that outcomes that are strongly decharged (a definition will follow shortly) are strongly-enforceable by this protocol.

Definition 3.8 (Strongly Decharged Outcome)**.** Consider a scheduling model (N, M, d, c). A machine $a \in M$ is *strongly decharged* if it is weakly decharged and additionally

$$c_a(x) < \sum_{i \in S_a(x)} \min_{b \in M} c_b(b, x_{-i}), \text{ if } |S_a(x)| > 1 \text{ and } c_a(x) > 0. \tag{5}$$

Machines that are not strongly decharged are called *charged*. The outcome x is called *strongly decharged* if all machines are strongly decharged.

We now introduce the strong x-enforcing protocol.

Definition 3.9 (Strong x-enforcing protocol)**.** The *strong x-enforcing protocol* takes as input a strongly decharged outcome x. As before, we use x to define for any outcome z and machine a the sets $S_a^0(z)$ and $S_a^1(z)$. Additionally, we define the set $S_a^2(z) := \{i \in S_a(z) \backslash S_a(x) : c_{x_i}(x) = 0\}$ that we term *strong foreign players on a*. Then, the strong x-enforcing protocol assigns for all $i \in N, z \in M^n$, the following cost sharing methods:

$$\xi_i(z) := \begin{cases} \dfrac{\min_{b \in M} c_b(b, x_{-i})}{\sum_{j \in S_{x_i}(x)} \min_{b \in M} c_b(b, x_{-j})} \cdot c_{x_i}(x), \text{ if } S_{z_i}(z) = S_{z_i}(x) \text{ and } c_{x_i}(x) > 0, \\ c_{z_i}(z), \text{ if } S_{z_i}^2(z) \neq \emptyset \text{ and } i = \min S_{z_i}^2(z), \\ c_{z_i}(z), \text{ if } S_{z_i}^2(z) = \emptyset, \, S_{z_i}^1(z) \neq \emptyset \text{ and } i = \min S_{z_i}^1(z), \\ c_{z_i}(z), \text{ if } S_{z_i}^1(z) = \emptyset, \, S_{z_i}(z) \subset S_{z_i}(x) \text{ and } i = \min S_{z_i}(z), \\ 0, \qquad \text{else.} \end{cases}$$

The protocol works almost the same as the weak x-enforcing protocol, it only accounts differently for strong foreign players.

Theorem 3.10. Consider a scheduling model (N, M, d, c). If an outcome x is strongly decharged, then the strong x-enforcing protocol is separable and strongly enforces x.

Proof. Separability follows from the definition of the strong x-enforcing protocol. We only show that for any Nash equilibrium $z \neq x$ we have $C(z) \leq C(x)$. To this end, fix such a z and let

$$i := \min \{j \in N : z_j \neq x_j\} \tag{7}$$

be the smallest player who deviates from x. First, note that for all $j > i$,

$$\xi_j(z) \leq \begin{cases} \xi_j(z_i, z_{-j}) = 0 \text{ , if } c_{x_j}(x) > 0 \\ \xi_j(x_j, z_{-j}) = 0 \text{ , if } c_{x_j}(x) = 0, \end{cases} \tag{8}$$

because z is a Nash equilibrium. Hence,

$$c_a(z) = 0 \text{ for all machines } a \neq z_i \text{ with foreign players } S_a^1(z) \neq \emptyset. \tag{9}$$

Also, $c_a(z) \le c_a(x)$ for all machines $a \neq z_i$ that only have home players $S_a^0(z) = S_a(z)$, because for these machines $\ell_a(z) \le \ell_a(x)$. Thus, we already have

$$c_a(z) \le c_a(x) \quad \text{for all machines } a \neq z_i. \tag{10}$$

If there is a strong foreign player on z_i, then even $c_{z_i}(z) = 0$ and we are done. Thus, from now on we assume that there are no strong foreign players on z_i. We can bound $c_{z_i}(z)$ from above using the Nash inequality $c_{z_i}(z) = \xi_i(z) \le \xi_i(x_i, z_{-i})$. The remaining proof focuses on bounding the value $\xi_i(x_i, z_{-i})$ from above.

The value of $\xi_i(x_i, z_{-i})$ assigned by the x-enforcing protocol depends on $S_{x_i}(x_i, z_{-i})$ and $c_{x_i}(x)$, for which there are three possibilities, according to the definition of the strong x-enforcing protocol. These cases are

(1) $S_{x_i}(x_i, z_{-i}) = S_{x_i}(x)$ and $c_{x_i}(x) > 0$, where the protocol returns $\xi_i(x_i, z_{-i}) = \xi_i(x)$.
(2) $S_{x_i}(x_i, z_{-i}) \subset S_{x_i}(x)$ and $i = \min S_{x_i}(x_i, z_{-i})$, where the protocol returns $\xi_i(x_i, z_{-i}) = c_{x_i}(x_i, z_{-i})$.
(3) All cases in which the protocol returns $\xi_i(x_i, z_{-i}) = 0$.

In each case we will find $c_{x_i}(z) + c_{z_i}(z) \le c_{x_i}(x) + c_{z_i}(x)$ and thus with (10) we have $C(z) \le C(x)$, which proves the Theorem. Note that (10) already implies $c_{x_i}(z) \le c_{x_i}(x)$.

We begin with Case 1. The condition $c_{x_i}(x) > 0$ implies that if there is some strong foreign player $j > i$ (with $z_j \neq x_j$ and $c_{x_j}(x) = 0$), then $c_{z_i}(z) = \xi_i(z) \le \xi_i(z_j, z_{-i}) = 0$ and we are done. Thus, we will in the following assume that there are no strong foreign players at all. If $\xi_i(x) = 0$, we obtain $0 = \xi_i(x) = \xi_i(x_i, z_{-i}) \ge \xi_i(z) = c_{z_i}(z)$, because we are in Case 1. Thus, we will also assume

$$\xi_i(x) > 0. \tag{11}$$

We now compare the allocation of load in the outcomes z and x, respectively. First, we consider machines $a \neq z_i$, which host foreign players $j \in S_a(z) \backslash S_a(x)$. For these foreign players we obtain

$$\begin{aligned} \min_{b \in M} c_b(b, x_{-j}) &\ge \min_{b \in M} c_b(b, x_{-i}) && \text{(12a)} \\ &\ge \xi_i(x) && \text{(12b)} \\ &> 0. && \text{(12c)} \end{aligned}$$

Observe that (7) implies $j > i$ and hence (by (3)) $d_j \ge d_i$. As the cost functions are non-decreasing, the first inequality (12a) follows. Inequality (12b) holds since x is decharged. The last inequality (12c) follows from (11). We conclude for machine a

$$\begin{aligned} c_a(a, x_{-j}) &\ge \xi_j(a, x_{-j}) \\ &\ge \xi_j(x) && (13) \\ &= \frac{c_{x_j}(x)}{\sum\limits_{k \in S_{x_j}(x)} \min\limits_{b \in M} c_b(b, x_{-k})} \cdot \min_{b \in M} c_b(b, x_{-j}) && (14) \\ &> 0, && (15) \\ &= c_a(z), && (16) \end{aligned}$$

where (13) holds because x is a Nash equilibrium and (14) stems from the definition of the protocol because there are no strong foreign players and hence $c_{x_j}(x) > 0$. Inequality (15) holds because of (12) and finally (16) holds because of (9).Hence, there must be a non-empty set of players $S_a(x) \backslash S_a(z)$. These players cannot be strong foreign players, thus $c_a(x) > 0$. With $c_a(z) = 0$ and $c_a(x) > 0$ we have $\ell_a(x) > \ell_a(z)$ for all machines $a \neq z_i$ with foreign players. For all machines a without foreign players we know $\ell_a(x) \ge \ell_a(z)$ and for machine x_i even

$\ell_{x_i}(x) = \ell_{x_i}(z) + d_i$ because we are in Case 1. Since the total load is the same in x and z, we have for machine z_i

$$\ell_{z_i}(z_i, x_{-i}) = \ell_{z_i}(x) + d_i \leq \ell_{z_i}(z). \tag{17}$$

Consequently,

$$\begin{aligned} \xi_i(z) = c_{z_i}(z) &\geq c_{z_i}(z_i, x_{-i}) && (18)\\ &\geq \frac{c_{x_i}(x)}{\sum\limits_{j \in S_{x_i}(x)} \min\limits_{b \in M} c_b(b, x_{-j})} \cdot \min_{b \in M} c_b(b, x_{-i}) && (19)\\ &= \xi_i(x) = \xi_i(x_i, z_{-i}), && (20) \end{aligned}$$

where the first inequality (18) holds because of (17) and the second inequality (19) because x is decharged and $c_{z_i}(z_i, x_{-i}) \geq \min_{b \in M} c_b(b, x_{-i})$. Equality (20) holds by the definition of the strong x-enforcing protocol for Case 1 and the last equation holds because we assume Case 1. If $|S_{x_i}(x)| > 1$, then inequality (19) is strict, because x is strongly decharged (i.e., (5) holds) which implies $\xi_i(z) > \xi_i(x_i, z_{-i})$. This contradicts the fact that z is a Nash equilibrium. Thus, $S_{x_i}(x) = \{i\}$ and $c_{z_i}(z) = \xi_i(z) \leq \xi_i(x_i, z_{-i}) = c_{x_i}(x_i, z_{-i}) = c_{x_i}(x)$. Moreover, using $c_{x_i}(z) = 0$, because $\ell_{x_i}(z) = 0$, we obtain $c_{x_i}(z) + c_{z_i}(z) \leq c_{x_i}(x) + c_{z_i}(x)$ as desired.

Case 2 is $S_{x_i}(x_i, z_{-i}) \subset S_{x_i}(x)$ and $i = \min S_{x_i}(x_i, z_{-i})$. Here, we obtain

$$c_{z_i}(z) = \xi_i(z) \leq \xi_i(x_i, z_{-i}) = c_{x_i}(x_i, z_{-i}) \leq c_{x_i}(x),$$

where the first inequality holds because z is a Nash equilibrium. The second inequality holds because Case 2 implies $\ell_{x_i}(x_i, z_{-j}) \leq \ell_{x_i}(x)$. We also get

$$c_{x_i}(z) = \sum_{j \in S_{x_i}(z)} \xi_j(z) \leq \sum_{j \in S_{x_i}(z)} \xi_j(z_i, z_{-j}) = 0.$$

This inequality is a result of (8), because in this case all players $j \in S_{x_i}(z)$ have a higher index $j > i$. Consequently, we have again $c_{x_i}(z) + c_{z_i}(z) \leq c_{x_i}(x) + c_{z_i}(x)$.

Finally, we examine Case 3 where the protocol returns $\xi_i(x_i, z_{-i}) = 0$ and thus for the Nash equilibrium z we have $c_{z_i}(z) = \xi_i(z) \leq \xi_i(x_i, z_{-i}) = 0$. Again, $c_{x_i}(z) + c_{z_i}(z) \leq c_{x_i}(x) + c_{z_i}(x)$. □

Remark 3.11. This upper bound on the price of anarchy does not hold for asymmetric games as inequality (8) does not hold. On the contrary, the price of anarchy of basic protocols is at least n for asymmetric games. Consider the model (N, M, d, c) with n players, $n+1$ machines $M = \{a_1, \ldots, a_{n+1}\}$, unit weights $d_i = 1$, cost functions $c_{a_i}(\ell) \equiv 1$ for $\ell \geq 1$, $i \in N$ and strategy sets $X_i = \{a_i, a_{n+1}\}$. The optimal outcome is $y = (a_{n+1}, \ldots, a_{n+1})$ with $C(y) = 1$ while $x = (a_1, \ldots, a_n)$ with $C(x) = n$ is a Nash equilibrium under every basic protocol.

Remark 3.12. Opposed to the equivalence from Theorem 3.4, a strongly-enforceable outcome is not necessarily strongly decharged. For example consider the model with two players, $d_1 = d_2 = 1$ and two machines with similar cost functions $c_{a_1}(1) = c_{a_2}(1) = 1$ and $c_{a_1}(2) = c_{a_2}(2) = 2$. Here, all outcomes have the same cost and $x = (a_1, a_1)$ is a Nash equilibrium under the x-enforcing protocol, but x is not strongly decharged.

3.3. An Optimal Protocol

Using the insights gained in the previous sections, we show that among all basic and separable protocols, the strong x-enforcing protocol gives rise to an optimal protocol simultaneously minimizing the price of anarchy and stability. Our main result involves the n-th harmonic number $\mathcal{H}_n = \sum_{i=1}^{n} \frac{1}{i}$.

Theorem 3.13.

$$\min_{\Xi \in \mathcal{B}_n} PoA(\Xi) = \min_{\Xi \in \mathcal{B}_n} PoS(\Xi) = \min_{\Xi \in \mathcal{S}_n} PoA(\Xi)$$
$$= \min_{\Xi \in \mathcal{S}_n} PoS(\Xi) = \mathcal{H}_n.$$

We will prove the theorem by two subsequent lemmas. In the first lemma, we prove that $\mathcal{H}_n$ is a lower bound on the price of stability for every basic cost sharing protocol. We then continue by presenting an algorithm that returns for any scheduling model a strongly decharged outcome of cost at most $\mathcal{H}_n$ times the cost of an optimal outcome. Together with the strong x-enforcing protocol we conclude that the price of anarchy of the thus defined protocol is precisely $\mathcal{H}_n$.

Lemma 3.14. The price of stability is at least $\mathcal{H}_n$ for basic cost sharing protocols on scheduling models with n players and non-decreasing cost functions. This lower bounds holds even for models with unit demands.

Proof. Consider the scheduling model (N, M, d, c) with n players that have unit demand $d_i = 1$ for all $i \in N$ and n machines with cost functions as in Table 1.

ℓ	$c_{a_1}(\ell)$	$c_{a_2}(\ell)$	$\dots$	$c_{a_i}(\ell)$	$\dots$	$c_{a_n}(\ell)$
0	0	0	$\dots$	0	$\dots$	0
1	$1+\epsilon$	$\frac{1}{2}$	$\dots$	$\frac{1}{i}$	$\dots$	$\frac{1}{n}$
> 1	$1+\epsilon$	n	$\dots$	n	$\dots$	n

for some small $\epsilon > 0$.

Table 1. Cost functions for machines used in the proof of Lemma 3.14

The only optimal outcome is clearly $y = (a_1, \dots, a_1)$ with $C(y) = 1 + \epsilon$. An outcome z can only be a Nash equilibrium if it is weakly decharged (Lemma 3.6). We show that the cheapest weakly decharged outcomes are those in which each machine is used by exactly one player, which all have the same cost as $x = (a_1, \dots, a_n)$. It is easy to see that outcome x is decharged and with $C(x) = \sum_{i=1}^{n} \frac{1}{i} = \mathcal{H}_n$ this proves the lemma.

If in an outcome z some machine other than a_1 is used by multiple players, then $C(z) \geq n$, thus such outcomes are more expensive than x. If in outcome z multiple players use machine a_1, say k players, then there are at least $k-1$ unused machines and for the cheapest of these, say machine $\hat{a}$, we have

$$c_{\hat{a}}(1) \leq \frac{1}{k}.$$

Thus, z is not weakly decharged as

$$c_{a_1}(z) = 1 + \epsilon > 1 = \sum_{i \in S_{a_1}(z)} \frac{1}{k} \geq \sum_{i \in S_{a_1}(z)} c_{\hat{a}}(\hat{a}, z_{-i}) = \sum_{i \in S_{a_1}(z)} \min_{b \in M} c_b(b, z_{-i}).$$

Altogether, only such outcomes in which all machines are used by exactly one player are cheap weakly-enforceable outcomes. $\square$

While the previous Lemma showed that there sometimes are no weakly-enforceable outcomes cheaper than $\mathcal{H}_n$ times the cost of an optimal outcome, the following lemma shows that we always find strongly decharged outcomes of at most $\mathcal{H}_n$ times the cost of an optimal outcome.

Lemma 3.15. Any scheduling model (N, M, d, c) with an optimal outcome y has a strongly decharged outcome x with $C(x) \leq \mathcal{H}_n \cdot C(y) = \sum_{k=1}^{n} \frac{1}{k} \cdot C(y)$.

Algorithm 1 Find strongly decharged outcome x

1: $k \leftarrow 1$ {stepnumber}
2: $x^1 \leftarrow y$ {starts with optimal outcome y}
3: $t_i \leftarrow 0$ for all $i \in N$ {stores when a player was last moved}
4: **while** there are charged machines **do**
5: $\quad a^k \leftarrow \text{argmax}\,\{c_a(x^k) : a \in M \text{ is charged}\}$ {select the most expensive charged machine}
6: $\quad$ **if** $\min\{c_b(b, x^k_{-i}) : b \in M\} = 0$ for $i = \min S_{a^k}(x^k)$ **then**
7: $\qquad$ {player can move to cost-free machine, case called *Zero-move*}
8: $\qquad i^k \leftarrow \min S_{a^k}(x^k)$ {select smallest player}
9: $\quad$ **else if** $\max\{t_i : i \in S_{a^k}(x^k)\} > 0$ **then**
10: $\qquad$ {some player on a^k was moved before, case called *Shuffle*}
11: $\qquad i^k \leftarrow \text{argmax}\,\{t_i : i \in S_{a^k}(x^k)\}$ {select last moved player}
12: $\quad$ **else**
13: $\qquad$ {no foreign players on a^k, case called *Kick-off*}
14: $\qquad i^k \leftarrow \min S_{a^k}(x^k)$ {select smallest player}
15: $\quad$ **end if**
16: $\quad b^k \leftarrow \text{argmin}\,\{c_b(b, x^k_{-i^k}) : b \in M\}$ {select cheapest machine}
17: $\quad x^{k+1} \leftarrow (b^k, x^k_{-i^k})$ {move player}
18: $\quad t_{i^k} \leftarrow k$ {store stepnumber}
19: $\quad k \leftarrow k + 1$ {iterate}
20: **end while**
21: **return** $x \leftarrow x^k$

Proof. The desired outcome x is found by Algorithm 1. The algorithm takes as input an optimal outcome y. In each cycle k of the algorithm's main loop (lines 4-20), a player i^k on the most expensive charged machine a^k is selected (line 5) and moved to the cheapest available machine b^k (lines 16,17). If possible, the algorithm selects a player who can be moved to a cost-free machine, this is called *Zero-move* (line 6). Otherwise, it selects a player that has been moved before in a last-in/first-out scheme which is maintained through the variables t_i that store the cycle in which each player was last moved. Such moves are called *Shuffles* (line 9). If neither a Zero-move nor a Shuffle is possible, the smallest player on the machine is selected, which is called *Kick-off* (line 12). The algorithm terminates when no charged machines are left.

First, we show that the algorithm terminates. To this end, observe that Shuffles are only performed when Zero-moves are not possible. Hence, if in cycle k a Shuffle is performed, the following inequalities hold.

$$\min_{b \in M} c_b(b, x^k_{-j}) > 0 \quad \text{for all } j \in S_{a^k}(x^k). \tag{21}$$

We now consider two cases. For $|S_{a^k}(x^k)| = 1$, we obtain

$$\begin{aligned} c_{a^k}(x^k) &> \min_{b \in M} c_b(b, x^k_{-i^k}) && (22)\\ &= c_{b^k}(b^k, x^k_{-i^k}) = c_{b^k}(x^{k+1}), && (23) \end{aligned}$$

where (22) follows because a^k is charged in x^k. Equality (23) follows since Algorithm 1 moves i^k to the cheapest available machine.

If $|S_{a^k}(x^k)| > 1$, then we obtain

$$\begin{aligned} c_{a^k}(x^k) &\geq \sum_{j \in S_{a_k}(x^k)} \min_{b \in M} c_b(b, x^k_{-j}) \qquad (24) \\ &> \min_{b \in M} c_b(b, x^k_{-i^k}) \qquad (25) \\ &= c_{b^k}(b^k, x^k_{-i^k}) = c_{b^k}(x^{k+1}), \end{aligned}$$

where (24) is valid because a^k is charged in x^k. The second inequality (25) holds because of (21) and the equalities follow as above. In both cases, a Shuffle moves the player to a strictly cheaper machine. To see that the algorithm terminates, we will now follow some player i over the course of the algorithm. Each Zero-move and each Shuffle take her to a strictly cheaper machine. If the player is moved in cycle k and is next moved by a Shuffle in cycle l, the cost of her machine $x_i^{k+1} = x_i^l$ may increase in the meantime as other players arrive on that machine. The algorithm assures by its last-in/first-out mechanism that these other players have been moved again before the Shuffle in cycle l and consequently the cost has decreased to the original level

$$c_{x_i^{k+1}}(x^{k+1}) \geq c_{x_i^l}(x^l).$$

Since only machines with positive costs can be charged, this implies that after a Zero-move, the player will never again be considered for Shuffles. Hence, a player can be moved by at most one Kick-off, afterwards a sequence of Shuffles and thereafter only Zero-moves. The sequence of Shuffles is finite because each Shuffle takes the player to a strictly cheaper machine. Once the player has has been moved by a Zero-move, further Zero-moves are only possible if in between some other player arrives on the player's machine via a Kick-off or a Shuffle, but again this is only finitely often possible. Altogether, each player can only be moved finitely often and thus the algorithm terminates after a finite number of cycles.

To complete the proof, we show that the final outcome x has cost $C(x) \leq \mathcal{H}_n \cdot C(y)$. The concept of this final part of the proof is that in outcome x the cost of every used machine is determined by the player who has last moved there or, if there are no such players, the home players. For this, some new notation is needed. Let $p_i, i \in N$, correspond to the position (by index) of player i on her optimal machine y_i, i.e., on any machine a we have $p_j = 1$ for player $j = \max S_a(y)$, $p_{j'} = 2$ for $j' = \max(S_a(y) \backslash \{j\})$ and so on. Consequently, when some player i performs her Kick-off in cycle k, there are p_i players sharing her machine $a^k = y_i$ at that moment and she is the smallest of them. We obtain for machine b^k that she is moved to

$$\begin{aligned} c_{b^k}(x^{k+1}) = c_{b^k}(b^k, x^k_{-i}) &= \min_{b \in M} c_b(b, x^k_{-i}) \\ &\leq \frac{1}{p_i} \cdot \sum_{j \in S_{a^k}(x^k)} \min_{b \in M} c_b(b, x^k_{-j}) \qquad (26) \\ &\leq \frac{1}{p_i} \cdot c_{a^k}(x^k) \qquad (27) \\ &\leq \frac{1}{p_i} \cdot c_{y_i}(y), \qquad (28) \end{aligned}$$

where the first inequality (26) is valid because i is the smallest of the p_i players on machine a^k in step k, the second inequality (27) holds because a^k is charged in x^k and the last inequality (28) holds because there are no foreign players on $a^k = y_i$ and hence $\ell_{a^k}(x^k) \leq \ell_{a^k}(y) = \ell_{y_i}(y)$.

Since Shuffles and Zero-moves assign player i to cheaper machines, after her last move in cycle k', she is on machine $b^{k'}$ at cost

$$c_{b^{k'}}(x^{k'}) \leq \frac{1}{p_i} \cdot c_{y_i}(y).$$

Altogether, in the final outcome x, the cost of a machine $a \in M$ to which players have been moved is determined by the last player who was moved there, that is, $i_a := \operatorname{argmax}\{t_i : i \in S_a(x)\}$. We thus obtain $c_a(x) \leq \frac{1}{p_{i_a}} \cdot c_{y_{i_a}}(y)$. For machines $a \in M$ that are used in x but where no player has been moved, the player $i_a := \max S_a(y)$ with $p_{i_a} = 1$ is still on machine a. In this case, the cost is bounded from above by $c_a(x) \leq c_a(y) = \frac{1}{p_{i_a}} \cdot c_{y_{i_a}}(y)$. Unused machines $a \in M$ have cost $c_a(x) = 0$. Altogether, we obtain $c_a(x) = \frac{1}{p_{i_a}} \cdot c_{y_{i_a}}(y)$ for all $a \in M$ with $\ell_a(x) > 0$, and $c_a(x) = 0$ for all $a \in M$ with $\ell_a(x) = 0$. This yields the desired bound for the cost of outcome x, because now every used machine $a \in M$ has a unique player i_a that determines the machine's cost. We obtain

$$\begin{aligned} C(x) = \sum_{a \in M} c_a(x) &\leq \sum_{\substack{a \in M \\ \ell_a(x) > 0}} \frac{1}{p_{i_a}} \cdot c_{y_{i_a}}(y) \leq \sum_{i \in N} \frac{1}{p_i} \cdot c_{y_i}(y) \\ &\leq \sum_{a \in M} \mathcal{H}_{p_{\max}} \cdot c_a(y) = \mathcal{H}_{p_{\max}} \cdot C(y) \leq \mathcal{H}_n \cdot C(y), \end{aligned}$$

where $p_{\max} := \max\{|S_a(y)| : a \in M\}$.

Observe that the bound for the price of anarchy obtained here can be much lower than $\mathcal{H}_n$ for scheduling models that have optimal outcomes, where the players are scattered over the machines and where therefore $p_{\max}$ is smaller than n. □

Remark 3.16. While Lemma 3.15 shows that an optimal outcome can be turned into a strongly decharged outcome of cost at most $\mathcal{H}_n$ times the cost of an optimal outcome, this holds true more generally: Algorithm 1 turns every outcome into a strongly decharged outcome with a cost increase of a factor at most $\mathcal{H}_n$. This may be useful if the computation of an optimal outcome is not possible in polynomial time. Still, Algorithm 1 does not run in polynomial time and this issue deserves further attention.

3.4. Non-Decreasing Cost per Unit

In this section we require that the cost functions are non-negative, non-decreasing and the per-unit costs $\frac{c(x)}{\ell(x)}$ are non-decreasing with respect to the load $\ell(x)$. Such functions are still quite rich as they contain non-negative, non-decreasing and convex functions.

We introduce the *opt-enforcing protocol* for which we prove a price of anarchy of 1. The intuition behind this protocol is similar to the x-enforcing protocols: make all undesired outcomes unstable by charging some player a very high price.

Definition 3.17 (opt-enforcing protocol)**.** For a given scheduling model (N, M, d, c) the *opt-enforcing protocol* takes as input an optimal outcome y. We again denote for any outcome z and machine a the set of foreign players on a by $S_a^1(z) = \{i \in S_a(z) \backslash S_a(y)\}$. Then, the opt-enforcing protocol assigns the cost sharing methods

$$\xi_i(z) := \begin{cases} d_i \cdot \dfrac{c_{z_i}(z)}{\ell_{z_i}(z)}, \text{ if } S^1_{z_i} = \emptyset \\ c_{z_i}(z), \text{ if } S^1_{z_i} \neq \emptyset \text{ and } i = \min S^1_{z_i}(z) \\ 0, \text{ else.} \end{cases}$$

Under the opt-enforcing protocol, the players share the cost proportional to their job weights on all machines without foreign players. On machines with foreign players, the foreign player with the smallest index pays the entire cost of the machine. The opt-enforcing protocol is obviously not uniform, as the private cost function of each player is dependent on an optimal outcome y and thus dependent on the entire set of machines M.

Theorem 3.18. The opt-enforcing protocol is separable and has a price of anarchy of 1 for games with non-decreasing cost per unit.

Theorem 3.18 applies not only to the scheduling games considered so far, but even to the wider class of general congestion games with non-decreasing cost per unit. Therefore, we will first introduce these games and then prove the theorem in this more general context.

In a general congestion game, we again require a scheduling model (N, M, d, c). Additionally, for each player i we are given an individual strategy set $X_i \subseteq 2^M$ that consists of subsets of the available resources M. Hence, a strategy profile $z \in X$ is a vector of sets with $z = (z_1, \dots, z_n)$, $z_i \in X_i$ and $z_i \subseteq M$ for all $i \in N$. A cost sharing protocol assigns cost sharing methods $\xi_{i,a} : X \to \mathbb{R}^+$ for every $i \in N$, $a \in M$ and the players' private cost functions are

$$\bar{\xi}_i(z) := \sum_{a \in z_i} \xi_{i,a}(z), \quad z \in X.$$

As before, an outcome $x \in X$ is a Nash equilibrium if $\bar{\xi}_i(x) \leq \bar{\xi}_i(z_i, x_{-i})$ for all $z_i \in X_i$ and $i \in N$.

Definition 3.19 (generalized opt-enforcing protocol)**.** The generalized opt-enforcing protocol takes as input an optimal outcome y. We again denote for any outcome z and machine a the set of foreign players on a by $S_a^1(z) = \{i \in S_a(z) \backslash S_a(y)\}$. Then, the protocol assigns the cost sharing methods

$$\xi_{i,a}(z) := \begin{cases} d_i \cdot \dfrac{c_a(z)}{\ell_a(z)}, \text{ if } a \in z_i \text{ and } S_a^1 = \emptyset & (29) \\ c_a(z), \text{ if } a \in z_i,\ S_a^1 \neq \emptyset \text{ and } i = \min S_a^1(z) & (30) \\ 0, \text{ else.} & (31) \end{cases}$$

Theorem 3.20. The generalized opt-enforcing protocol is separable and has a price of anarchy of 1 for general congestion games with non-decreasing cost per unit.

Proof. Budget balance and separability are clear from the definition of the cost sharing methods. For stability it can easily be verified that the optimal outcome y is a Nash equilibrium. We only proof the bound on the price of anarchy, showing that all Nash equilibria x are optimal outcomes using the Nash inequalities $\bar{\xi}_i(x) \leq \bar{\xi}_i(y_i, x_{-i})$ for all $i \in N$. Two cases are to be considered for the machines $a \in y_i$: either they host foreign players $S_a^1(x)$ or $S_a^1(x) = \emptyset$. If there are foreign players on a, then one of them will pay for the entire cost and hence (31) gives $\xi_{i,a}(y_i, x_{-i}) = 0$. If there are no foreign players on a, then we have $\ell_a(y_i, x_{-i}) \leq \ell_a(y)$ and thus

$$\frac{c_a(y_i, x_{-i})}{\ell_a(y_i, x_{-i})} \leq \frac{c_a(y)}{\ell_a(y)},$$

because the cost per unit is non-decreasing. Plugging this into (29) we have $\xi_{i,a}(y_i, x_{-i}) \leq \xi_{i,a}(y)$ for all $a \in y_i$ without foreign players. We conclude

$$\bar{\xi}_i(x) \leq \bar{\xi}_i(y_i, x_{-i}) = \sum_{a \in y_i} \xi_{i,a}(y_i, x_{-i}) \leq \sum_{a \in y_i} \xi_{i,a}(y) = \bar{\xi}_i(y)$$

where the first inequality holds because x is a Nash equilibrium. Altogether

$$C(x) = \sum_{i \in N} \bar{\xi}_i(x) \leq \sum_{i \in N} \bar{\xi}_i(y) = C(y)$$

which implies that every Nash equilibrium x is also an optimal outcome. □

4. Uniform Protocols

The separable protocols that we introduced so far were always tailored to some desirable outcome, either an enforceable outcome or even an optimal outcome. Since uniform protocols need to assign cost sharing methods independent of the set M, they cannot be based on specific outcomes. We show in this section that uniformity leads in general to an unbounded price of anarchy. Only for games in which the demands are integer multiples of each other we introduce the semi-ordered protocol that gives a price of anarchy of n. The question of $\min_{\xi \in \mathcal{U}_n} PoS(\xi)$ remains open.

4.1. Lower Bound

Theorem 4.1. There is no uniform protocol for which the price of anarchy has an upper bound. This holds even for models with at most 3 players, 3 machines and non-decreasing costs per unit.

Proof. The essence of uniform protocols is that adding machines to or removing them from the model does not change the cost shares of players using a certain machine, as long as the player set and the weight vector remain the same. This motivates the definition of cost share functions $\hat{\xi}_i$ that return the private cost ξ_i of player i as a function of the machine a that she uses and the set of players $S \subseteq N$ sharing the machine.

$$\hat{\xi}_i(a, S) := \xi_i(x) \; \forall \; a \in M, S \subseteq N, i \in S, x \in M^n : S_a(x) = S. \tag{32}$$

As in Definition 2.1, an outcome x is a Nash equilibrium if none of the players can reduce their private cost by choosing a different machine. This can be expressed via cost share functions as follows. For all $i \in N, a \in M$ it holds that

$$\hat{\xi}_i(x_i, S_{x_i}(x)) \leq \hat{\xi}_i(a, S_a(x) \cup \{i\}). \tag{33}$$

For the proof of the theorem, we propose a number of scheduling models and show that for any uniform cost sharing protocol at least one of these models has a Nash equilibrium of more than q times the cost of an optimal outcome for arbitrary $q \geq 2$. Throughout the entire proof, the player set will always be $N = \{1, 2, 3\}$ with weights $d = (4, 3, 2)$. The machines will be a subset of $M = \{a_1, \dots, a_7\}$ with cost functions as outlined in Table 2.

ℓ	$c_{a_1}(\ell)$	$c_{a_2}(\ell)$	$c_{a_3}(\ell)$	$c_{a_4}(\ell)$	$c_{a_5}(\ell)$	$c_{a_6}(\ell)$	$c_{a_7}(\ell)$
0	0	0	0	0	0	0	0
2	0	0	0	0	0	q^2	0
3	0	0	0	q^3	0	q^4	0
4	1	q	1	$2q^3$	q^3	$2q^4$	q^4
5	2	q^3	q^5		$2q^3$		
6	q^3	q^5					
7	q^4						

Table 2. Cost functions of machines used in the proof of Theorem 4.1

First, consider the model with machines $M_1 = \{a_1, a_2\}$ and their respective cost functions. The optimal outcome $y_1 = (a_2, a_1, a_1)$ has cost $C(y_1) = q+2$, while the outcome $x_1 = (a_1, a_2, a_2)$ has cost $C(x_1) = q^3 + 1$. Either x_1 is a Nash equilibrium and hence the protocol has a price of anarchy greater than q or or one of the three players can reduce her private cost by choosing a different machine, which results by (33) in the following three cases.

a) $\hat{\xi}_1(a_2, \{1,2,3\}) < \hat{\xi}_1(a_1, \{1\}) = 1$. In this case, consider $M_2 = \{a_2, a_3\}$. The optimal outcome $y_2 = (a_3, a_2, a_2)$ with $C(y_2) = q^3 + 1$ is not a Nash equilibrium and due to stability some other outcome x with cost $C(x) \geq q^5$ has to be a Nash equilibrium.

b) $\hat{\xi}_2(a_1, \{1,2\}) < \hat{\xi}_2(a_2, \{2,3\}) \leq q^3$. In this case, consider $M_3 = \{a_1, a_2, a_4\}$. The optimal outcome $y_3 = (a_1, a_2, a_4)$ has cost $C(y_3) = 1$, while the outcome $x_3 = (a_2, a_1, a_4)$ has cost $C(x_3) = q$. Either x_3 is a Nash equilibrium or, again, one of the players can reduce her private cost by choosing a different machine, which leads to the following cases.

 b.1) $\hat{\xi}_1(a_1, \{1,2\}) < \hat{\xi}_1(a_2, \{1\}) = q$. This contradicts b), that is $\hat{\xi}_1(a_1, \{1,2\}) = c_1(d_1 + d_2) - \hat{\xi}_2(a_1, \{1,2\}) > q^4 - q^3 > q$.

 b.2) $\hat{\xi}_1(a_4, \{1,3\}) < \hat{\xi}_1(a_1, \{1\}) = q$. In this case, consider $M_4 = \{a_2, a_4, a_5\}$. The optimal outcome $y_4 = (a_2, a_5, a_4)$ with $C(y_4) = q$ is not a Nash equilibrium and due to stability some other outcome x with cost $C(x) \geq q^3$ has to be a Nash equilibrium.

 b.3) Players 2 and 3 cannot reduce their private cost as $\xi_2(x_3) = c_{a_1}(x_3) = 0$ and $\xi_3(x_3) = c_{a_4}(x_3) = 0$.

c) $\hat{\xi}_3(a_1, \{1,3\}) < \hat{\xi}_3(a_2, \{2,3\}) \leq q^3$. In this case, consider $M_5 = \{a_1, a_2, a_6\}$. The optimal outcome $y_5 = (a_2, a_1, a_1)$ has cost $C(y_5) = q + 1$, while the outcome $x_5 = (a_1, a_2, a_6)$ has cost $C(x_5) = q^2 + 1$. Either x_5 is a Nash equilibrium or, again, one of the players can reduce her private cost by choosing a different machine.

 c.1) $\hat{\xi}_1(a_2, \{1,2\}) < \hat{\xi}_1(a_1, \{1\}) = 1$. In this case, consider again $M_3 = \{a_1, a_2, a_4\}$. The optimal outcome $y_3 = (a_1, a_2, a_4)$ with $C(y_3) = 1$ is not a Nash equilibrium and due to stability some other outcome x with cost $C(x) \geq q$ has to be a Nash equilibrium.

 c.2) $\hat{\xi}_1(a_6, \{1,3\}) < \hat{\xi}_1(a_1, \{1\}) = 1$. In this case, consider $M_6 = \{a_3, a_5, a_6\}$. The optimal outcome $y_6 = (a_3, a_5, a_6)$ with $C(y_6) = q^2 + 1$ is not a Nash equilibrium and due to stability some other outcome x with cost $C(x) \geq q^3$ has to be a Nash equilibrium.

 c.3) Player 2 cannot reduce her private cost because $\xi_2(x_5) = c_{a_2}(x_5) = 0$.

 c.4) $\hat{\xi}_3(a_1, \{1,3\}) < \hat{\xi}_3(a_6, \{3\}) = q^2$. In this case, consider $M_7 = \{a_1, a_6, a_7\}$. The optimal outcome $y_7 = (a_1, a_7, a_6)$ with $C(y_7) = q^2 + 1$ is not a Nash equilibrium and due to stability some other outcome x with cost $C(x) \geq q^3$ has to be a Nash equilibrium.

 c.5) $\hat{\xi}_3(a_2, \{2,3\}) < \hat{\xi}_3(a_6, \{3\}) = q^2$. This extends the original assumption from c) which is $\hat{\xi}_3(a_1, \{1,3\}) < \hat{\xi}_3(a_2, \{2,3\}) < q^2$. Therefore this case implies c.4).

Altogether, every uniform cost sharing protocol allows in at least one of the analyzed cases a Nash equilibrium of at least q times the cost of an optimal outcome for an arbitrary $q \geq 2$. Consequently, the price of anarchy is not bounded. □

4.2. Models with Restricted Weights

Although uniform protocols in general do not allow a bound on the price of anarchy, the following class of games permits uniform cost sharing protocols with a bounded price of anarchy. We assume that the player's weights are either uniform, i.e. $d_1 = \ldots = d_n$, or they are multiples of each other. In the following, we propose a semi-ordered protocol that a has a price of anarchy of at most n for such games. In this section, we assume that the players are indexed with their weights in non-increasing order: $d_1 \geq d_2 \geq \ldots \geq d_n$. The semi-ordered protocol lets the players one after the other choose a machine and lets them pay only for the additional cost they cause on that machine, thus making the choice of a player i independent of the choices of all players $j > i$.

Definition 4.2 (Semi-ordered Protocol)**.** The semi-ordered protocol assigns for all $i \in N$

$$\xi_i(x) := c_{x_i}\Big(\sum_{j\in S_{x_i}(x):j\le i} d_j\Big) - c_{x_i}\Big(\sum_{j\in S_{x_i}(x):j<i} d_j\Big). \tag{34}$$

Theorem 4.3. The semi-ordered protocol is uniform and its price of anarchy is at most n for instances, where the players' weights are multiples of each other, that is

$$d_i = q_i \cdot d_{i+i} \quad \text{for some} \quad q_i \in \mathbb{N} \quad \text{and all} \quad i = 1,\dots,n-1.$$

Proof. Budget balance, separability and uniformity of the cost sharing methods are clear. A Nash equilibrium can be found by asking the players in the order of their index to choose a machine that minimizes their private cost considering the choice of all previous players. For proving the bound on the price of anarchy, consider a model (N, M, d, c) which fulfills the restriction on the players' weights. Suppose y is an optimal outcome and x a Nash equilibrium. First, we show that $\xi_i(x) \le \max_{j\le i} \xi_j(y)$ holds for all $i \in N$, which is motivated by the idea that a player can always choose a machine that she or one of the larger players had chosen in the optimal outcome. To this end, fix player $i \in N$. On some machine $a \in \{y_1,\dots,y_{i-1}\}$ there is in outcome x less load from the first $i-1$ players than in the optimal outcome y from the first i players. Due to the restriction on the players' weights this difference in load on machine a has to be at least d_i yielding

$$\sum_{j\in S_a(x):j<i} d_j + d_i \le \sum_{j\in S_a(y):j\le i} d_j.$$

Also, there is a player $k \le i$ (hence $d_k \ge d_i$), $k \in S_a(y)$ who in outcome y uses machine a and for whom due to the weight restrictions

$$\sum_{j\in S_a(y):j<k} d_j \le \sum_{j\in S_a(x):j<i} d_j < \sum_{j\in S_a(x):j<i} d_j + d_i \le \sum_{j\in S_a(y):j\le k} d_j.$$

Combining the above inequalities with (34) yields

$$\begin{aligned}
&\xi_i(a, x_{-i}) \\
&\qquad = c_a\big(\sum_{j\in S_a(x):j<i} d_j + d_i\big) - c_a\big(\sum_{j\in S_a(x):j<i} d_j\Big) \\
&\qquad \le c_a\big(\sum_{j\in S_a(y):j\le k} d_j\big) - c_a\big(\sum_{j\in S_a(y):j<k} d_j\big) \\
&\qquad = \xi_k(y) \le \max_{j\le i} \xi_j(y)
\end{aligned}$$

and because x is a Nash Equilibrium we have

$$\xi_i(x) \le \xi_i(a, x_{-i}) \le \xi_k(y) \le \max_{j\le i} \xi_j(y).$$

This implies

$$C(x) = \sum_{i\in N} \xi_i(x) \le \sum_{i\in N} \max_{j\le i} \xi_j(y) \le \sum_{i\in N} \max_{j\in N} \xi_j(y) = n \cdot \max_{j\in N} \xi_j(y) \le n \cdot C(y),$$

proving the claim. ∎

References

[1] H. Ackermann, H. Röglin, and B. Vöcking, *Pure Nash equilibria in player-specific and weighted congestion games*, Theor. Comput. Sci. **410**, no. 17 (2009), pp. 1552–1563.

[2] S. Aland, D. Dumrauf, M. Gairing, B. Monien, and F. Schoppmann, *Exact price of anarchy for polynomial congestion games.*, in Proc. of STACS, 2006, pp. 218–229.

[3] E. Anshelevich, A. Dasgupta, J. Kleinberg, É. Tardos, T. Wexler, and T. Roughgarden, *The price of stability for network design with fair cost allocation*, SIAM J. Comput. **38**, no. 4 (2008), pp. 1602–1623.

[4] B. Awerbuch, Y. Azar, and A. Epstein, *The price of routing unsplittable flow*, in Proc. of STOC, ACM, 2005, pp. 57–66.

[5] K. Bhawalkar, M. Gairing, and T. Roughgarden, *Weighted congestion games: Price of anarchy, universal worst-case examples, and tightness*, in Proc. of ESA, Springer, 2010, pp. 17–28.

[6] I. Caragiannis, *Efficient coordination mechanisms for unrelated machine scheduling*, in Proc. of SODA, 2009, pp. 815–824.

[7] H.-L. Chen and T. Roughgarden, *Network design with weighted players*, Theor. Comput. Syst. **45**, no. 2 (2009), pp. 302–324.

[8] H.-L. Chen, T. Roughgarden, and G. Valiant, *Designing network protocols for good equilibria*, SIAM J. Comput. **39**, no. 5 (2010), pp. 1799–1832.

[9] G. Christodoulou and E. Koutsoupias, *The price of anarchy of finite congestion games*, in Proc. of STOC, 2005, pp. 67–73.

[10] G. Christodoulou, E. Koutsoupias, and A. Nanavati, *Coordination mechanisms*, in Proc. of ICALP, Springer, 2004, pp. 45–56.

[11] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *Selfish routing in capacitated networks*, Math. of Oper. Res. **29** (2004), pp. 961–976.

[12] E. Even-Dar, A. Kesselman, and Y. Mansour, *Convergence time to Nash equilibrium in load balancing*, ACM Trans. Algorithms **3**, no. 3 (2007), p. 32.

[13] D. Fotakis, S. Kontogiannis, E. Koutsoupias, M. Mavronicolas, and P. Spirakis, *The structure and complexity of Nash equilibria for a selfish routing game*, in Proc. of ICALP, 2002, pp. 123 – 134.

[14] M. Gairing, B. Monien, and K. Tiemann, *Routing (un-) splittable flow in games with player-specific linear latency functions*, in Proc of ICALP, 2006, pp. 501–512.

[15] S. Ieong, R. McGrew, E. Nudelman, Y. Shoham, and Q. Sun, *Fast and compact: A simple class of congestion games*, in Proc. AAAI, 2005, pp. 489–494.

[16] N. Immorlica, L. Li, V. Mirrokni, and A. Schulz, *Coordination mechanisms for selfish scheduling*, Theor. Comput. Sci. **410**, no. 17 (2009), pp. 1589–1598.

[17] E. Koutsoupias and C. H. Papadimitriou, *Worst-case equilibria*, in STACS, 1999, pp. 404–413.

[18] I. Milchtaich, *Congestion games with player-specific payoff functions*, Games Econom. Behav. **13**, no. 1 (1996), pp. 111 – 124.

[19] M. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, 1994.

[20] R. Rosenthal, *A class of games possessing pure-strategy Nash equilibria*, Inter. J. Game Theory **2**, no. 1 (1973), pp. 65–67.

[21] T. Roughgarden, *The price of anarchy is independent of the network topology*, Journal of Computer and System Sciences **67** (2002), pp. 341–364.

[22] T. Roughgarden, *Selfish Routing and the Price of Anarchy*, The MIT Press, 2005.

[23] T. Roughgarden and É. Tardos, *How bad is selfish routing?*, J. ACM **49**, no. 2 (2002), pp. 236–259.

[24] R. Srikant, *The Mathematics of Internet Congestion Control*, Birkhäuser Boston, 2003.

Paper 5

Characterizing the Existence of Potential Functions in Weighted Congestion Games

Tobias Harks, Max Klimm and Rolf H. Möhring
Characterizing the Existence of Potential Functions in Weighted Congestion Games
Theory of Computing Systems **49** (2011), no. 1, pp. 46–70

Abstract. Since the pioneering paper of Rosenthal a lot of work has been done in order to determine classes of games that admit a potential. First, we study the existence of potential functions for weighted congestion games. Let $\mathcal{C}$ be an arbitrary set of locally bounded functions and let $\mathcal{G}(\mathcal{C})$ be the set of weighted congestion games with cost functions in $\mathcal{C}$. We show that every weighted congestion game $G \in \mathcal{G}(\mathcal{C})$ admits an exact potential if and only if C contains only affine functions. We also give a similar characterization for w-potentials with the difference that here $\mathcal{C}$ consists either of affine functions or of certain exponential functions. We finally extend our characterizations to weighted congestion games with facility-dependent demands and elastic demands, respectively.

1. Introduction

In many situations, the state of a system is determined by a large number of independent agents, each pursuing selfish goals optimizing an individual objective function. A natural framework for analyzing such decentralized systems are noncooperative games. It is well known that an equilibrium point in pure strategies (if it exists) need not optimize the social welfare as individual incentives are not always compatible with social objectives. Fundamental goals in algorithmic game theory are to decide whether a Nash equilibrium in pure strategies (PNE for short) exists, how efficient it is in the worst case, and how fast an algorithm (or protocol) converges to an equilibrium.

One of the most successful approaches in accomplishing these goals is the potential function approach initiated by Rosenthal [28] and generalized by Monderer and Shapley in [25]: one defines a function P on the set of possible strategies of the game and shows that every strictly improving move by one defecting player strictly reduces (increases) the value of P. If the set of outcomes of such a game is finite, every sequence of improving moves converges to a PNE. In particular, the global minimum (maximum) of P is a PNE.

A function P with the property above is called a *potential function* of the game. If one can associate a weight w_i to each player such that w_iP decreases about the same value as the private cost of the defecting player i, then P is called a *w-potential*. If, in addition, $w_i = 1$ for each player, then P is called an *exact potential*.

1.1. Framework

An important class of games studied in the game theory, operations research, computer science and economics literature is the class of congestion games. This class of games has several concrete applications such as scheduling games, routing games, facility location games, and network design games, see [1, 3, 7, 16, 19, 24]. Congestion games, as introduced by Rosenthal [28], model the interaction of a finite set of strategic agents that compete over a finite set of facilities. A pure strategy of each player is a set of facilities. We consider cost minimization games. Here, the cost of facility f is given by a real-valued cost function c_f that depends on the number of players using f and the private cost of every player equals the sum of the costs of the facilities in the strategy that she chooses.[1] Rosenthal [28] proved in a seminal paper that such congestion games always admit a PNE by showing these games posses an exact potential function.

In a *weighted congestion game*, every player has a demand $d_i \in \mathbb{R}_{>0}$ that she places on the chosen facilities. The cost of a facility is a function of the total demand of the facility. In contrast to unweighted congestion games, weighted congestion games, even with two players, do not always admit a PNE, see the examples given by Fotakis et al. [14], Goemans et al. [17], and Libman and Orda [21]. It is worth noting that the instance in [14] only relies on cost functions that are either linear or maxima of two linear functions. The instance in [17] only uses polynomial cost functions with nonnegative coefficients and degree of at most two.

On the positive side, Fotakis et al. [14, 15] proved that every weighted congestion game with affine cost functions possesses an exact potential function and thus, a PNE. Panagopoulou and Spirakis [27] proved existence of a weighted potential function for the case that all costs are determined by the exponential function. The results of [14, 15] and [27] are particularly appealing as they establish existence of a potential function *independent* of the underlying game structure, that is, *independent* of the underlying strategy set, demand vector, and number of players, respectively. To further stress this independence property, we rephrase the result of Fotakis et al. as follows: Let $\mathcal{C}$ be a set of affine cost functions and let $\mathcal{G}(\mathcal{C})$ be the set of *all* weighted congestion games with cost functions in $\mathcal{C}$. Then, *every* game in $\mathcal{G}(\mathcal{C})$ possesses an exact potential.

A natural open question is to decide whether there are further functions guaranteeing the existence of an exact or weighted potential. We thus investigate the following question: How large is the class $\mathcal{C}$ of (continuous) cost functions such that every game in the set of weighted congestion games $\mathcal{G}(\mathcal{C})$ with cost functions in $\mathcal{C}$ does admit a potential function and hence a PNE?

Before we outline our results we present related work and explain, why it is important to characterize weighted congestion games admitting a potential function.

1.2. Related Work

Fundamental issues in algorithmic game theory are the computability of Nash equilibria and the design of distributed dynamics (for instance best-response) that provably converge in reasonable time to a Nash equilibrium (in pure or mixed strategies).

[1] Since we allow the cost of a facility to be positive or negative, we also cover the maximization games.

Monderer and Shapley [25] formalized Rosenthal's approach of using potential functions to determine the existence of PNE. Furthermore, they show that one-side better response dynamics, i.e., sequences of unilateral deviations strictly reducing the deviating player's private costs, always converge to a PNE provided the game is finite and admits a potential. In addition, they proved that w-potential games have other desirable properties, e.g., the Fictitious Play Process introduced by Brown [6] converges to a PNE [26]. For recent progress on convergence towards approximate Nash equilibria using potential functions, see Awerbuch et al. [4] and Fotakis et al. [12].

Fabrikant et al. [11] proved that one can efficiently compute a PNE for symmetric network congestion games with nondecreasing cost functions. Their proof uses a potential function argument, similar to Rosenthal [28]. Fotakis et al. [14] proved that one can compute a PNE for weighted network games with affine cost (with nonnegative coefficients) in pseudo-polynomial time (again using a potential function).

Milchtaich [23] introduced weighted congestion games with player-specific cost functions. He presented, among other results, a game on 3 parallel links with 3 players, which does not possess a PNE. On the other hand, he proved that such games with 2 players do possess a PNE. Ackermann et al. [1] characterized conditions on the strategy space in weighted congestion games that guarantee the existence of PNE. They also considered the case of player-specific cost functions.

Gairing et al. [16] derive a potential function for the case of unweighted congestion games with player-specific linear latency functions (without a constant term). Mavronicolas et al. [22] prove that every unweighted congestion game with player-specific (additive or multiplicative) constants on parallel links has an ordinal potential. Even-Dar et al. [10] consider a variety of load balancing games with makespan objectives and prove among other results that games on unrelated machines possess a generalized ordinal potential function. For related results, see the survey by Vöcking [29] and references therein.

Potential functions also play a central role in Shapley cost sharing games with weighted players, which are special cases of weighted congestion games, see Anshelevich et al. [3] and Albers et al. [2]. In the variant with weighted players, each player i has a demand d_i that she wishes to place on each facility of an allowable subset of facilities (e.g., a path in a network connecting her source node s_i to her terminal node t_i). When facility $f \in F$ is stressed with a load of $\ell_f(x)$ in strategy profile x, there exists a cost of $k_f(\ell_f(x))$. Under Shapley cost sharing, this cost is shared fairly with respect to the demands among the users. Thus the cost of player i for using facility f is defined as $c_{i,f}(x) = k_f(\ell_f(x))\, d_i/\ell_f(x)$ and clearly, the private cost of player i in strategy profile x is given as $\pi_i(x) = \sum_{f \in x_i} c_{i,f}(x)$. For the unweighted case ($d_i = 1, i \in N$), Anshelevich et al. [3] proved existence of PNE and derived bounds on the worst-case efficiency of Nash equilibria using a potential function argument. This argument fails in general for games with weighted players, see the counterexamples given by Chen and Roughgarden [7]. Determining subclasses of Shapley cost sharing games with weighted players that admit a potential, however, is an open problem that we address in this paper.

1.3. Our Results for Weighted Congestion Games

Our first two results provide a characterization of the existence of exact and w-potential functions for the set of weighted congestion games with locally bounded and continuous cost functions, respectively. Let $\mathcal{C}$ be an arbitrary set of locally bounded functions and let $\mathcal{G}(\mathcal{C})$ be the set of weighted congestion games with cost functions in $\mathcal{C}$. We show that every weighted congestion game $G \in \mathcal{G}(\mathcal{C})$ admits an exact potential if and only if $\mathcal{C}$ contains only affine functions. Our proof relies on a seminal result of Monderer and Shapley [25] stating that a finite strategic game is an exact potential game if and only if the discrete integral over the player's utility functions along every 4-cycle is zero. We apply this 4-cycle condition for a generic weighted congestion game and obtain a functional equation in terms of the used cost functions and the used demands, respectively. By varying the demands we obtain a necessary and sufficient condition on the used cost functions expressed by a differential equation. Finally, we show that only affine functions fulfill this differential equation. We note that while the if-part of our characterization follows already from the potential function given by Fotakis et al. [14], our complete characterization also delivers an alternative non-constructive proof for the if-part.

For an arbitrary set $\mathcal{C}$ of continuous functions, we show that every weighted congestion game $G \in \mathcal{G}(\mathcal{C})$ possesses a weighted potential if and only if exactly one of the following cases hold: (i) $\mathcal{C}$ contains only affine functions; (ii) $\mathcal{C}$ contains only exponential functions such that $c(\ell) = a_c\, e^{\phi\, \ell} + b_c$ for some $a_c, b_c, \phi \in \mathbb{R}$, where a_c and b_c may depend on c, while ϕ must be equal for every $c \in \mathcal{C}$. To derive this result we use a scaling technique that transforms a w-potential game into an exact potential game. This allows us to apply the 4-cycle criterion of Monderer and Shapley on the transformed game which again gives rise to a functional equation. However, due to the degree of freedom of our scaling technique it is not possible to derive a differential equation. By discretizing the demands we can express the necessary and sufficient conditions on the cost functions as a recurrence relation for which we show that it is only satisfied by either affine or exponential only.

We additionally show that the above characterizations for exact and w-potentials are valid even if we restrict the set $\mathcal{G}(\mathcal{C})$ to two-player games (four-player games for w-potentials), three-facility games (four-facility games for w-potentials), games with symmetric strategies, games with singleton strategies, games with integral demands. Moreover, we derive a result for two-player weighted congestion games, showing that every such game with cost functions in $\mathcal{C}$ admits a weighted potential if $\mathcal{C} = \{(c : \mathbb{R}_{>0} \to \mathbb{R}) : c(x) = a\, m(x) + b,\ \ a, b \in \mathbb{R}\}$, where $m : \mathbb{R}_{>0} \to \mathbb{R}$ is a strictly monotonic function.

Our results have a series of consequences. First, using a result of Monderer and Shapley [25, Lemma 2.10], our characterization of w-potentials in weighted congestion games carries over to the mixed extension of weighted congestion games.

Second, we obtain the following characterizations for Shapley cost sharing games. Let $\mathcal{K}$ be a set of continuous functions. Then, the set $\mathcal{S}(\mathcal{K})$ of Shapley cost sharing games with weighted players and construction cost functions in $\mathcal{K}$ are w-potential games if and only if $\mathcal{K}$ contains either quadratic construction cost functions $k(\ell) = a_k\, \ell^2 + b_k\, \ell$ or functions of type $k(\ell) = a_k\, e^{\phi\, \ell}\, \ell + b_k\, \ell$ for some $a_k, b_k, \phi \in \mathbb{R}$, where a_k and b_k may depend on k, while ϕ must be equal for every $k \in \mathcal{K}$. Notice that these results hold for arbitrary coefficients $a_k, b_k, \phi \in \mathbb{R}$. Thus, we obtain the existence of PNE for a family of games with nondecreasing and strictly concave construction costs modeling the effect of economies of scale.

After the initial publication of this paper, Harks and Klimm [18] explored the existence of PNE in weighted congestion games. For a class $\mathcal{C}$ of twice continuously differentiable cost functions, they showed that the conditions given in Theorem 3.9 are in fact necessary for the existence of PNE in all weighted congestion games contained in $\mathcal{G}(\mathcal{C})$. Their characterization, however, requires new techniques based on the analysis of generic improvement cycles, see [18] for details.

1.4. Our Results for Extended Models

In the second part of this paper, we introduce two non-trivial extensions of weighted congestion games.

First, we study weighted congestion games with *facility-dependent* demands, that is, the demand $d_{i,f}$ of player i depends on the facility f. These games contain, among others, scheduling games on identical, restricted, related and unrelated machines. In contrast to classical load balancing games, we do not consider makespan objectives. In our model, the private cost of a player is a function of the machine load multiplied with the demand of the player.

We show the following: Let $\mathcal{C}$ be a set of continuous functions and let $\mathcal{G}^{fd}(\mathcal{C})$ denote the set of weighted congestion games with facility-dependent demands and cost functions in $\mathcal{C}$. Every $G \in \mathcal{G}^{fd}(\mathcal{C})$ has a w-potential if and only if $\mathcal{C}$ contains only affine functions. In this case the w-potential is an exact potential. To the best of our knowledge, our characterization establishes for the first time the existence of an exact potential function (and hence the existence of a PNE) for affine cost functions and *arbitrary* strategy sets and demands, respectively.

Second, we study weighted congestion games with *elastic* demands. Here, each player i is allowed to choose both a subset of the set of facilities and her demand d_i out of a compact set $D_i \subset \mathbb{R}_{>0}$ of demands that are allowable for her. This congestion model can be interpreted as a generalization of Cournot games [9], where multiple producers strategically determine quantities they will produce. The cost of a producer is given by her offered quantity multiplied with the market price, which is usually a decreasing function of the total quantity offered by all producers. Weighted congestion games with elastic demands generalize Cournot games in the sense that there are multiple markets (facilities) and each player may offer her quantity on allowable subsets of these markets.

Weighted congestion games with elastic demands have several additional applications: they model, e.g., routing problems in the Internet, where each user wants to route data along a path in the network and adjusts the injected data rate according to the level of congestion in the network. Most mathematical models for routing and congestion control rely on fractional routing, see Kelly [20] and Cole et al. [8]. In practice, however, routing protocols use single path routing, see, e.g., the current TCP/IP protocol. Weighted congestion games with elastic demands model both congestion control and unsplittable routing. Yet another application is that of Shapley cost sharing games with players that may vary their requested demand.

Let $\mathcal{G}^{e}(\mathcal{C})$ be the set of weighted congestion games with elastic demands where each player may chose her demand out of a compact space and where the cost of each facility is determined by a function in $\mathcal{C}$. Here, our main contribution is to show that all games $G \in \mathcal{G}^{e}(\mathcal{C})$ are w-potential games if and only if $\mathcal{C}$ contains only affine functions. For this important class of games, this result also establishes for the first time the existence of PNE.

2. Preliminaries

A *finite strategic game* is a tuple $G = (N, X, \pi)$ where $N = \{1, \dots, n\}$ is the non-empty finite set of players, $X = \bigtimes_{i \in N} X_i$ where X_i is the finite and non-empty set of strategies of player i, and $\pi : X \to \mathbb{R}^n$ is the combined private cost function.

We will call an element $x \in X$ strategy profile. For $S \subset N$, $-S$ denotes the complementary set of S, and we define for convenience of notation $X_S = \bigtimes_{j \in S} X_j$. Instead of $X_{-\{i\}}$ we will write X_{-i}, and with a slight abuse of notation we will write sometimes a strategy profile as $x = (x_i, x_{-i})$ meaning that $x_i \in X_i$ and $x_{-i} \in X_{-i}$. A strategy profile x is a pure Nash equilibrium if for all $i \in N$ the condition $\pi_i(x) \leq \pi_i(y_i, x_{-i})$ holds for all $y_i \in X_i$. A sufficient condition for the existence of a pure Nash equilibrium is the existence of a potential function, see Monderer and Shapley [25].

Definition 2.1 (Weighted and exact potential games)**.** A strategic game $G = (N, X, \pi)$ is called *weighted potential game* if there is a vector $w = (w_i)_{i \in N} \in \mathbb{R}^n_{>0}$ and a function $P : X \to \mathbb{R}$ such that $\pi_i(x_i, x_{-i}) - \pi_i(y_i, x_{-i}) = w_i \left(P(x_i, x_{-i}) - P(y_i, x_{-i})\right)$ for all $i \in N$, $x_{-i} \in X_{-i}$, and all $x_i, y_i \in X_i$. The function P together with the vector w is then called a w-potential of the game G. The function P is called an *exact* potential if $w_i = 1$ for all $i \in N$.

Monderer and Shapley [25, Theorem 2.8] characterized exact potentials by the use of certain cycles defined below. For this, let a finite strategic game $G = (N, X, \pi)$ be given. A *path* in X is a sequence $\gamma = (x^0, x^1, \dots x^m)$ with $x^k \in X$, $k = 0, \dots, m$, such that for all $k \in \{1, \dots, m\}$ there exists a unique player $i_k \in N$ such that $x^k = (x^k_{i_k}, x^{k-1}_{-i_k})$ for some $x^k_{i_k} \neq x^{k-1}_{i_k}$, $x^k_{i_k} \in X_i$. A path is called closed if $x^0 = x^m$ and is called simple closed if in addition $x^k \neq x^l$ for $0 \leq k \neq l \leq m-1$. The length of a closed path is defined as the number of its distinct elements. For a set of strategy profiles X let $\Gamma(X)$ denote the set of all simple closed paths in X that have length 4. For a finite path $\gamma = (x^0, x^1, \dots, x^m)$ let the discrete path integral of π along γ be defined as $I(\gamma, \pi) = \sum_{k=1}^m \big(\pi_{i_k}(x^k) - \pi_{i_k}(x^{k-1})\big)$ where i_k is the deviator at step k in γ, that is $x^k_{i_k} \neq x^{k-1}_{i_k}$.

Theorem 2.2 (Monderer and Shapley)**.** Let $G = (N, X, \pi)$ be a finite strategic game. Then, G is an exact potential game if and only if $I(\gamma, \pi) = 0$ for all $\gamma \in \Gamma(X)$.

In the following, we will use this characterization in order to study the existence of potentials in weighted congestion games.

3. Weighted Congestion Games

Definition 3.1 (Congestion model)**.** A tuple $\mathcal{M} = (N, F, X = \times_{i\in N} X_i, (c_f)_{f\in F})$ is called a *congestion model*, where $N = \{1, \dots, n\}$ is a non-empty, finite set of players, F is a non-empty, finite set of facilities, for each player $i \in N$, her collection of pure strategies X_i is a non-empty, finite set of subsets of F and $(c_f)_{f\in F}$ is a set of cost functions.

In the following, we will define weighted congestion games similar to Goemans et al. [17].

Definition 3.2 (Weighted congestion game)**.** Let $\mathcal{M} = (N, F, X, (c_f)_{f\in F})$ be a congestion model and $(d_i)_{i\in N} \in \mathbb{R}^n_{>0}$ be a vector of demands. The corresponding *weighted congestion game* is the strategic game $G(\mathcal{M}) = (N, X, \pi)$, where π is defined as $\pi = \times_{i\in N} \pi_i$, $\pi_i(x) = \sum_{f\in x_i} d_i\, c_f\big(\ell_f(x)\big)$ and $\ell_f(x) = \sum_{j\in N: f\in x_j} d_j$.

We call $\ell_f(x)$ the *load* on facility f in strategy x. In case there is no confusion on the underlying congestion model, we will write G instead of $G(\mathcal{M})$.

A slightly different class of games has been considered by (among others) Fotakis et al. [14, 15], Gairing et al. [16] and Mavronicolas et al. [22]. They considered games that almost coincide with Definition 2.1 except that the private cost of every player is not scaled by her demands. We call such games *normalized* if they comply with Definition 2.1 except that the private costs are defined as $\bar{\pi}_i(x) = \sum_{f\in x_i} c_f\big(\ell_f(x)\big)$ for all $i \in N$.

Fotakis et al. [14] show that there are normalized weighted congestion games with $c_f(\ell) = \ell$ for all $f \in F$ that are not exact potential games. They also show that any normalized weighted congestion game with linear costs on the facilities admits a w-potential.

We state the following trivial relations between weighted congestion games and normalized weighted congestion games: Let $G = (N, X, \pi)$ and $\bar{G} = (N, X, \bar{\pi})$ be a weighted congestion game and a normalized weighted congestion game with demands $(d_i)_{i\in N}$, respectively. Moreover, let them share the same congestion model and the same demands. Then G and $\bar{G}$ coincide in the following sense: (i) A strategy profile $x \in X$ is a PNE in G if and only if x is a PNE in $\bar{G}$; (ii) A real-valued function $P : X \to \mathbb{R}$ is a $(w_i/d_i)_{i\in N}$-potential for G if and only if P is a $(w_i)_{i\in N}$-potential for $\bar{G}$; (iii) A real-valued function $P : X \to \mathbb{R}$ is an ordinal potential for G (see [25] for a definition) if and only if P is an ordinal potential for $\bar{G}$; (iv) The real-valued function $P : X \to \mathbb{R}$ is an exact potential for G if and only if P is a $(d_i)_{i\in N}$-potential for $\bar{G}$; (v) The real-valued function $P : X \to \mathbb{R}$ is an exact potential for $\bar{G}$ if and only if P is a $(1/d_i)_{i\in N}$-potential for G. All proofs rely on the simple observation that $\pi_i(x) = d_i\, \bar{\pi}_i(x)$ for all $i \in N, x \in X$.

3.1. Characterizing the Existence of an Exact Potential

In the following, we will examine necessary and sufficient conditions for a weighted congestion game G to be a potential game. The criterion in Theorem 2.2 states that the existence of an exact potential for $G = (N, X, \pi)$ is equivalent to the fact that $I(\gamma, \pi) = 0$ for all $\gamma \in \Gamma(X)$. In such paths, either one or two players deviate. It is easy to verify that $I(\gamma, \pi) = 0$ for all paths γ with only one deviating player. Considering a path γ with two deviating players, say i and j, each of them uses two different strategies, say $x_i, y_i \in X_i$ and $x_j, y_j \in X_j$. We denote by $z_{-\{i,j\}} \in X_{-\{i,j\}}$ the strategy profile of all players except i and j that remains constant in γ. Then, a generic path $\gamma \in \Gamma(X)$ can be written as $\gamma = \big((x_i, x_j, z_{-\{i,j\}}), (y_i, x_j, z_{-\{i,j\}}), (y_i, y_j, z_{-\{i,j\}}), (x_i, y_j, z_{-\{i,j\}}), (x_i, x_j, z_{-\{i,j\}})\big)$. For a facility $f \in F$, we define $r_f = \sum_{k \in N \setminus \{i,j\} : f \in (z_{-\{i,j\}})_m} d_m$ as the sum of the demands on f in the partial strategy profile $z_{-\{i,j\}}$. The following lemma provides an explicit formula for the calculation of $I(\gamma, \pi)$ for such a path.

Lemma 3.3. Let $\mathcal{M} = (N, F, X, (c_f)_{f \in F})$ be a congestion model and $G(\mathcal{M})$ a corresponding weighted congestion game with demands $(d_i)_{i \in N}$. Moreover, let

$$\gamma = \big((x_i, x_j, z_{-\{i,j\}}), (y_i, x_j, z_{-\{i,j\}}), (y_i, y_j, z_{-\{i,j\}}), (x_i, y_j, z_{-\{i,j\}}), (x_i, x_j, z_{-\{i,j\}})\big)$$

be an arbitrary path in $\Gamma(X)$ with two deviating players. Then,

$$\begin{aligned} I(\gamma, \pi) = &\sum_{f \in F_1 \cup F_{11}} (d_j - d_i) c_f(d_i + d_j + r_f) - d_j c_f(d_j + r_f) + d_i c_f(d_i + r_f) \\ &+ \sum_{f \in F_3 \cup F_9} (d_i - d_j) c_f(d_i + d_j + r_f) - d_i c_f(d_i + r_f) + d_j c_f(d_j + r_f), \end{aligned} \tag{1}$$

where $F_1 = (x_i \setminus y_i) \cap (x_j \setminus y_j)$, $F_3 = (x_i \setminus y_i) \cap (y_j \setminus x_j)$, $F_9 = (y_i \setminus x_i) \cap (x_j \setminus y_j)$, and $F_{11} = (y_i \setminus x_i) \cap (y_j \setminus x_j)$.

Proof. We fix $i, j \in N, x_i, y_i \in X_i, x_j, y_j \in X_j$, and $z_{-\{i,j\}} \in X_{-\{i,j\}}$ arbitrarily and consider the path $\gamma = \big((x_i, x_j, z_{-\{i,j\}}), (y_i, x_j, z_{-\{i,j\}}), (y_i, y_j, z_{-\{i,j\}})(x_i, y_j, z_{-\{i,j\}}), (x_i, x_j, z_{-\{i,j\}})\big)$. We compute straightforwardly that

$$\begin{aligned} I(\gamma, \pi) = &\pi_i(y_i, x_j, z_{-\{i,j\}}) - \pi_i(x_i, x_j, z_{-\{i,j\}}) + \pi_j(y_i, y_j, z_{-\{i,j\}}) - \pi_j(y_i, x_j, z_{-\{i,j\}}) \\ &+ \pi_i(x_i, y_j, z_{-\{i,j\}}) - \pi_i(y_i, y_j, z_{-\{i,j\}}) + \pi_j(x_i, x_j, z_{-\{i,j\}}) - \pi_j(x_i, y_j, z_{-\{i,j\}}). \end{aligned} \tag{2}$$

For fixed x_i, y_i, x_j and y_j, every facility $f \in F$ can be chosen by player i in both strategy x_i and strategy y_i, in one of these strategies or not at all. The same holds for player j and strategies x_j and y_j. We can thus decompose F into 16 disjoint sets $F_1, \dots, F_{16}$. The first set, F_1, comprises all facilities that are in $(x_i \setminus y_i) \cap (x_j \setminus y_j)$. F_2 contains all facilities that are in $(x_i \setminus y_i) \cap (x_j \cap y_j)$, and so on. The comprehensive description of all 16 cases is given in Table 1.

	$x_j \setminus y_j$	$x_j \cap y_j$	$y_j \setminus x_j$	$F \setminus (x_j \cup y_j)$
$x_i \setminus y_i$	F_1	F_2	F_3	F_4
$x_i \cap y_i$	F_5	F_6	F_7	F_8
$y_i \setminus x_i$	F_9	F_{10}	F_{11}	F_{12}
$F \setminus (x_i \cup y_i)$	F_{13}	F_{14}	F_{15}	F_{16}

Table 1. Decomposition of F into 16 disjoint subsets $F_k, k = 1, \dots, 16$.

In order to compute for instance the first term of equation (2), we notice that in strategy profile $x = (y_i, x_j, z_{-\{i,j\}})$ the load on each facility $f \in F_5 \cup F_6 \cup F_9 \cup F_{10}$ equals $\ell_f(x) = d_i + d_j + r_f$, while the load on each facility $g \in F_7 \cup F_8 \cup F_{11} \cup F_{12}$ equals $\ell_g(x) = d_i + r_g$. These considerations lead to the following equation. We will use the notation $\sum_{F,G}$ for $\sum_{f \in F \cup G}$.

$$\begin{aligned} I(\gamma,\pi) = d_i & \left(\sum_{F_9,F_{10}} c_f(d_i+d_j+r_f) + \sum_{F_{11},F_{12}} c_f(d_i+r_f) - \sum_{F_1,F_2} c_f(d_i+d_j+r_f) - \sum_{F_3,F_4} c_f(d_i+r_f) \right) \\ & + d_j \left(\sum_{F_7,F_{11}} c_f(d_i+d_j+r_f) + \sum_{F_3,F_{15}} c_f(d_j+r_f) - \sum_{F_5,F_9} c_f(d_i+d_j+r_f) - \sum_{F_1,F_{13}} c_f(d_j+r_f) \right) \\ & + d_i \left(\sum_{F_2,F_3} c_f(d_i+d_j+r_f) + \sum_{F_1,F_4} c_f(d_i+r_f) - \sum_{F_{10},F_{11}} c_f(d_i+d_j+r_f) - \sum_{F_9,F_{12}} c_f(d_i+r_f) \right) \\ & + d_j \left(\sum_{F_1,F_5} c_f(d_i+d_j+r_f) + \sum_{F_9,F_{13}} c_f(d_j+r_f) - \sum_{F_3,F_7} c_f(d_i+d_j+r_f) - \sum_{F_{11},F_{15}} c_f(d_j+r_f) \right). \end{aligned}$$

By reordering the summation many terms cancel out and we obtain

$$\begin{aligned} I(\gamma,\pi) = & \sum_{f \in F_1 \cup F_{11}} (d_j - d_i) c_f(d_i + d_j + r_f) - d_j c_f(d_j + r_f) + d_i c_f(d_i + r_f) \\ & + \sum_{f \in F_3 \cup F_9} (d_i - d_j) c_f(d_i + d_j + r_f) - d_i c_f(d_i + r_f) + d_j c_f(d_j + r_f), \end{aligned}$$

establishing the result. □

Using Lemma 3.3, we can derive a sufficient condition on the existence of an exact potential in a weighted congestion game.

Proposition 3.4. Let $\mathcal{M} = (N, F, X, (c_f)_{f \in F})$ be a congestion model and $G(\mathcal{M})$ a corresponding weighted congestion game with demands $(d_i)_{i \in N}$. For each facility $f \in F$, we denote by $N^f = \{i \in N : (\exists x_i \in X_i : f \in x_i))\}$ the set of players potentially using f, and by $\mathcal{R}^f_{-\{i,j\}} = \left\{\sum_{k \in P} d_k : P \subseteq N^f \setminus \{i,j\}\right\}$ the set of possible residual demands by all players except i and j. If for all $f \in F$ and all $i, j \in N^f$

$$(d_j - d_i) c_f(d_i + d_j + r_f) - d_j c_f(d_j + r_f) + d_i c_f(d_i + r_f) = 0 \quad \forall r_f \in \mathcal{R}^f_{-\{i,j\}}, \tag{3}$$

then G admits an exact potential.

Proof. Using the criterion of Monderer and Shapley, it is enough to prove that $I(\gamma, \pi) = 0$ for all $\gamma \in \Gamma(X)$. By Lemma 3.3, $I(\gamma, \pi)$ evaluates to

$$\begin{aligned} I(\gamma,\pi) = & \sum_{f \in F_1 \cup F_{11}} (d_j - d_i) c_f(d_i + d_j + r_f) - d_j c_f(d_j + r_f) + d_i c_f(d_i + r_f) \\ & + \sum_{f \in F_3 \cup F_9} (d_i - d_j) c_f(d_i + d_j + r_f) - d_i c_f(d_i + r_f) + d_j c_f(d_j + r_f), \end{aligned} \tag{4}$$

for some $i, j \in N^f$ and $r_f \in \mathcal{R}^f_{-\{i,j\}}$. Using (3) each summand of (4) equals 0, establishing the result. □

It follows easily that the above condition is satisfied if all demands are equal (this corresponds to unweighted congestion games, see Rosenthal's potential [28]).

$$c(\ell) = \begin{cases} 0 & \text{if } \ell \in (0,1] \\ 5\ell - 5 & \text{if } \ell \in (1,3] \\ -\ell + 13 & \text{if } \ell \in (3,5] \\ 2\ell - 2 & \text{if } \ell \in (5,6] \\ 10 & \text{if } \ell \in (6,\infty) \end{cases}$$

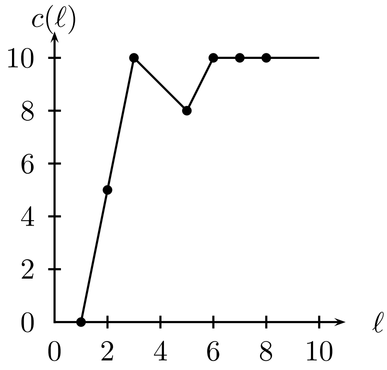


Figure 1. A non-linear cost function c_f that gives rise to an exact potential in all weighted congestion games with demands $d_1 = 1, d_2 = 3$ and $d_3 = 5$.

For different demands $d_i \neq d_j$ it is a useful observation that we can write the condition of Proposition 3.4 as

$$\frac{c_f(d_i + d_j + r_f) - c_f(d_j + r_f)}{d_i} = \frac{c_f(d_j + r_f) - c_f(d_i + r_f)}{d_j - d_i} \tag{5}$$

for all $i, j \in N^f$ and $r_f \in \mathcal{R}^f_{-\{i,j\}}$. Thus, the difference quotients of c_f between the points $d_i + r_f$ and $d_j + r_f$ as well as $d_j + r_f$ and $d_i + d_j + r_f$ must have the same value. For *arbitrary* demands (weighted congestion games) and *affine* cost functions, one can check that the above condition is also satisfied, see the positive result of Fotakis et al. [14].

For a single weighted congestion game, the linearity condition on cost functions, however, is only sufficient but not necessary. In Example 3.5, we show that it is possible to construct a non-affine cost function that satisfies the condition of Proposition 3.4 for all 3 player games with demand vector $(1, 2, 5)$.

Example 3.5. Let $\mathcal{M} = (N = \{1, 2, 3\}, X, F, (c_f)_{f \in F})$ be an arbitrary congestion model with three players and let $G(\mathcal{M})$ be a corresponding weighted congestion game with demands $d_1 = 1, d_2 = 2, d_3 = 5$.

We want to construct a non-linear cost function that gives rise to an exact potential in G. To this end, we consider an arbitrary 4-cycle γ. We apply Lemma 3.3 and obtain that $I(\gamma, \pi)$ evaluates to

$$\begin{aligned} I(\gamma, \pi) = &\sum_{f \in F_1 \cup F_{11}} (d_j - d_i) c_f(d_i + d_j + r_f) - d_j c_f(d_j + r_f) + d_i c_f(d_i + r_f) \\ &+ \sum_{f \in F_3 \cup F_9} (d_i - d_j) c_f(d_i + d_j + r_f) - d_i c_f(d_i + r_f) + d_j c_f(d_j + r_f), \end{aligned} \tag{6}$$

Regarding (6), only the following realizations of (d_i, d_j, r_f) are possible:

$$\begin{matrix} (1,2,0), & (1,5,0), & (2,5,0), \\ (1,2,5), & (1,5,2), & (2,5,1). \end{matrix} \tag{7}$$

Note that only realizations with $d_i < d_j$ are considered, the others are symmetric and, thus, omitted. Proposition 3.4 establishes that it is sufficient for the existence of an exact potential that in each cost function c_f, the values to the arguments shown in (7) lie on a straight line. It is easy to construct a non-linear cost function $c : \mathbb{R}_{>0} \to \mathbb{R}$ satisfying this property. An example of such a function is given in Fig. 1.

We derive that $I(\gamma, \pi) = 0$ for any 4-cycle γ in any such game regardless of the structure of the set of strategies.

There is, however, an important question left: Are there non-affine cost functions that give rise to an exact potential in *all* weighted congestion games, i.e., in weighted congestion games with arbitrary strategy spaces and number of players, respectively? Under mild assumptions on feasible cost functions, we will give in Theorem 3.7 a negative answer to this question. First, we need the following lemma.

Lemma 3.6. Let $\mathcal{C}$ be a set of functions and let $\mathcal{G}(\mathcal{C})$ be the set of all weighted congestion games with cost functions in $\mathcal{C}$. Every $G \in \mathcal{G}(\mathcal{C})$ has an exact potential if and only if for all $c \in \mathcal{C}$

$$(x - y)\, c(x + y + z) - x\, c(x + z) + y\, c(y + z) = 0 \tag{8}$$

for all $x, y \in \mathbb{R}_{>0}$ and $z \in \mathbb{R}_{\geq 0}$.

Proof. Suppose $G = (N, X, \pi) \in \mathcal{G}(\mathcal{C})$ is a weighted congestion game with cost functions in $\mathcal{C}$ and every $c \in \mathcal{C}$ satisfies (8). First, we will show that G has an exact potential. To this end, let $\gamma \in \Gamma(X)$ be an arbitrary simple closed path in X of length 4. $I(\gamma, \pi)$ evaluates to (1), which is zero using (8).

For the opposite direction suppose that there is a $\tilde{c} \in \mathcal{C}$ that does not satisfy equation (8). This implies that there are $x, y \in \mathbb{R}_{>0}$ and $z \in \mathbb{R}_{\geq 0}$ such that

$$(x - y)\tilde{c}(x + y + z) - x\tilde{c}(x + z) + y\tilde{c}(y + z) \neq 0.$$

Now consider the congestion model $\mathcal{M} = (N, F, X, (c_f)_{f\in F})$ where $N = \{1, 2, 3\}$, $F = \{f, g, h\}$, $X_1 = \big\{\{f\}, \{g\}\big\}$, $X_2 = \big\{\{f\}, \{h\}\big\}$, $X_3 = \big\{f\big\}$, and $c_f = c_g = c_h = \tilde{c}$. Let $G = (N, X, \pi)$ be a corresponding weighted congestion game with demands $d_1 = y$, $d_2 = x$ and $d_3 = z$. We will investigate the value of $I(\gamma, \pi)$ for

$$\gamma = \big((\{g\}, \{h\}, \{f\}), (\{f\}, \{h\}, \{f\}), (\{f\}, \{f\}, \{f\}), (\{g\}, \{f\}, \{f\}), (\{g\}, \{h\}, \{f\})\big).$$

This value equals $(x - y)\tilde{c}(x + y + z) - x\tilde{c}(x + z) + y\tilde{c}(y + z) \neq 0$ implying that this game does not possess an exact potential function. $\square$

We will now solve the functional equation (8) in order to characterize all cost functions that guarantee an exact potential in all weighted congestion games. We require the following property: A function $c : \mathbb{R}_{>0} \to \mathbb{R}$ is *locally bounded*, if for every compact set $K \subset \mathbb{R}_{>0}$, $|c(x)| < M_K$ for all $x \in K$ and a constant $M_K \in \mathbb{R}_{>0}$ potentially depending on K.

Theorem 3.7. Let $\mathcal{C}$ be a set of locally bounded functions and let $\mathcal{G}(\mathcal{C})$ be the set of weighted congestion games with cost functions in $\mathcal{C}$. Then, every $G \in \mathcal{G}(\mathcal{C})$ admits an exact potential function if and only if $\mathcal{C}$ contains affine functions only, that is, every $c \in \mathcal{C}$ can be written as $c(\ell) = a_c\, \ell + b_c$ for some $a_c, b_c \in \mathbb{R}$.

Proof. Fotakis et al. [14] derived an exact potential function for weighted congestion games with affine cost functions. We can provide an alternative non-constructive proof by checking that affine functions fulfill functional equation (8) and, thus, we may conclude that they give rise to an exact potential. We will prove the reverse direction in two steps.

In Step 1, we prove the following: Let c fulfill (8). Then, c is differentiable and $c'(x + z) = \big(c(x + z) - c(z)\big)/x$ holds for all $x \in \mathbb{R}_{>0}$ and $z \in \mathbb{R}_{\geq 0}$.

First, we will show continuity of c on $\mathbb{R}_{>0}$. Let $x \in \mathbb{R}_{>0}$ and $z \in \mathbb{R}_{\geq 0}$ be arbitrary and let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{>0}$ such that $y_n \overset{n\to\infty}{\longrightarrow} 0$ and both $y_n + z > 0$ and $y_n + x > 0$ for all $n \in \mathbb{N}$. Then, using (8) we get $x\,(c(x + z + y_n) - c(x + z)) = y_n\,(c(x + z + y_n) - c(z + y_n))$. As c is bounded on any compact set, the right hand side of the previous equation goes to 0 as n goes to infinity and hence $x \lim_{n\to\infty}(c(x + z + y_n) - c(x + z)) = 0$. This shows continuity in $x + z$.

Moreover, (8) implies that $x\,\big(c(x+z+y_n)-c(x+z)\big)/y_n = c(x+z+y_n)-c(z+y_n)$. As c is continuous we know that the limits on the right hand side of the previous equation exist and, thus, $c'(x+z)=\big(c(x+z)-c(z)\big)/x$ holds for all $x\in\mathbb{R}_{>0}$.

So c satisfies the differential equation $c'(x+z)=\big(c(x+z)-c(z)\big)/x$. We will show in Step 2 that only affine functions solve this differential equation. To see this, we set $t=x+z$, which leads to the differential equation $c'(t)=\big(c(t)-c_0\big)\,/\big(t-t_0\big), t\in\mathbb{R}_{>0}$, where $c_0=c(z)$ and $t_0=z$ are constants. Standard calculus shows that for every initial value c_1 for the initial time $t_1>t_0$, this ordinary linear differential equation admits a unique solution $c(t)=(t-t_0)\,C+c_0$, where $C=(c_1-c_0)/(t_1-t_0)$. □

3.2. Characterizing the Existence of a w-Potential

Our next goal is to determine whether weaker notions of potential functions will enrich the class of cost functions giving rise to a potential game. The idea of a w-potential allows a player specific scaling of the private cost π_i by a strictly positive w_i. It is a useful observation that the existence of a w-potential function is equivalent to the existence of a strictly positive-valued vector $w=(w_i)_{i\in N}$ such that the game G^w with private costs $\bar{\pi}=\bigtimes_{i\in N}\pi_i/w_i$ has an exact potential.

Using this equivalent formulation and Theorem 2.2 it follows that the existence of an exact potential function for the game $G^w=(N,X,\bar{\pi})$ is equivalent to $I(\gamma,\bar{\pi})=0$ for all $\gamma\in\Gamma(X)$ suggesting that G^w has an exact potential if and only if there are $w_i,w_j\in\mathbb{R}_{>0}$ such that

$$\left(\frac{d_i}{w_i}-\frac{d_j}{w_j}\right)c_f(d_i+d_j+r_f)=\frac{d_i}{w_i}c_f(d_i+r_f)-\frac{d_j}{w_j}c_f(d_j+r_f)$$

for all $i,j\in N$ and all $r_f\in\mathcal{R}^f_{-\{i,j\}}$. In particular it is necessary that either $c_f(d_i+d_j+r_f)=c_f(d_j+r_f)=c_f(d_i+r_f)$ or the value $\alpha(d_i,d_j)$ defined as

$$\alpha(d_i,d_j)=\frac{w_i}{w_j}=\frac{d_i}{d_j}\cdot\frac{c_f(d_i+d_j+r_f)-c_f(d_i+r_f)}{c_f(d_i+d_j+r_f)-c_f(d_j+r_f)} \tag{9}$$

is strictly positive and independent of both f and r_f. This observation leads us to the following lemma.

Lemma 3.8. Let $\mathcal{C}$ be a set of functions. Let $\mathcal{G}(\mathcal{C})$ be the set of weighted congestion games with cost functions in $\mathcal{C}$. Every $G\in\mathcal{G}(\mathcal{C})$ has a w-potential if and only if for all $x,y\in\mathbb{R}_{>0}$, there exists an $\alpha(x,y)\in\mathbb{R}_{>0}$ such that

$$\alpha(x,y)\cdot\big(c(x+y+z)-c(y+z)\big)=\frac{x}{y}\cdot\big(c(x+y+z)-c(x+z)\big) \tag{10}$$

for all $z\in\mathbb{R}_{\geq 0}$ and $c\in\mathcal{C}$.

Proof. Let $\mathcal{M}=(N,F,X,(c_f)_{f\in F})$ be a congestion model where for all $f\in F$ the cost function c_f satisfies equation (10). Let $(d_i)_{i\in N}, d_i\in\mathbb{R}_{>0}$, be an arbitrary vector of demands and $G(\mathcal{M})$ the corresponding weighted congestion game. We will show that this game possesses a w-potential. Lemma 3.8 implies that there for any two distinct players $i,j\in N$ there is $\alpha(d_i,d_j)\in\mathbb{R}_{>0}$ such that

$$\alpha(d_i,d_j)\cdot\big(c_f(d_i+d_j+z)-c_f(d_j+z)\big)=\frac{d_i}{d_j}\cdot\big(c_f(d_i+d_j+z)-c_f(d_i+z)\big) \tag{11}$$

for all $z\in\mathbb{R}_{\geq 0}$ and $f\in F$. If $c_f(d_i+d_j+z)-c_f(d_j+z)=0$ for all $f\in F$ then $\alpha(d_i,d_j)$ can be chosen arbitrarily. If, in contrast, there is $f'\in F$ such that $c_{f'}(d_i+d_j+z)-c_{f'}(d_j+z)\neq 0$ then $\alpha(d_i,d_j)=d_i/d_j\cdot\big(c_{f'}(d_i+d_j)-c_{f'}(d_i)\big)/\big(c_{f'}(d_i+d_j)-c_{f'}(d_j)\big)$. In both cases, we can chose the values of $\alpha(d_i,d_j)$ such that $\alpha(d_i,d_k)=\alpha(d_i,d_j)\cdot\alpha(d_j,d_k)$ for all $d_i,d_j,d_k\in\mathbb{R}_{>0}$. In particular, we can find a vector of weights $(w_i)_{i \in N} \in \mathbb{R}_{>0}$ such that $\alpha(d_i, d_j) = w_i/w_j$ for all $i, j \in N$ with $i \neq j$.

Using Monderer and Shapley's criterion we will show that the corresponding game $G^w = (N, X, \bar{\pi})$ has an exact potential. For this, we consider an arbitrary path $\gamma \in \Gamma(X)$. Without loss of generality only two players, say i and j, change their strategies in γ while the sum of the demands of all other players is equal to a facility-specific value r_f. Analogously to the proof of Lemma 3.6, we get

$$\begin{aligned} I(\gamma, \bar{\pi}) = &\sum_{f \in F_1, F_{11}} \left(\frac{d_j}{w_j} - \frac{d_i}{w_i}\right) c_f(d_i + d_j + r_f) - \frac{d_j}{w_j} c_f(d_j + r_f) + \frac{d_i}{w_i} c_f(d_i + r_f) \\ &+ \sum_{f \in F_3, F_9} \left(\frac{d_i}{w_i} - \frac{d_j}{w_j}\right) c_f(d_i + d_j + r_f) - \frac{d_i}{w_i} c_f(d_i + r_f) + \frac{d_j}{w_j} c_f(d_j + r_f). \end{aligned}$$

We multiply with w_i, use $\alpha(d_i, d_j) = w_i/w_j$ and obtain

$$\begin{aligned} w_i\, I(\gamma, \bar{\pi}) = &\sum_{f \in F_1, F_{11}} \alpha(d_i, d_j) d_j \left(c_f(d_i + d_j + r_f) - c_f(d_j + r_f)\right) - d_i \left(c_f(d_i + d_j + r_f) + c_f(d_i + r_f)\right) \\ + &\sum_{f \in F_3, F_9} -\alpha(d_i, d_j) d_j \left(c_f(d_i + d_j + r_f) - c_f(d_j + r_f)\right) + d_i \left(c_f(d_i + d_j + r_f) - c_f(d_i + r_f)\right). \end{aligned}$$

Using equation (11) shows that $I(\gamma, \bar{\pi}) = 0$ proving the first result.

To show the other direction, assume that the condition on $\mathcal{C}$ does not hold, that is, there are $x_0, y_0 \in \mathbb{R}_{>0}$ such that for every $\alpha > 0$ there is a cost function $c_\alpha \in \mathcal{C}$ and a value $z_\alpha \in \mathbb{R}_{\geq 0}$ with

$$\alpha \cdot \big(c_\alpha(x_0 + y_0 + z_\alpha) - c_\alpha(y_0 + z_\alpha)\big) \neq \frac{x_0}{y_0} \cdot \big(c_\alpha(x_0 + y_0 + z) - c_\alpha(x_0 + z_\alpha)\big). \tag{12}$$

First assume that $c_\alpha(x_0 + y_0 + z_\alpha) - c_\alpha(y_0 + z_\alpha) = 0$. Note that (12) implies that $c_\alpha(x_0 + y_0 + z) - c_\alpha(x_0 + z_\alpha) \neq 0$. Let us consider the congestion model $\mathcal{M} = (N, F, X, (c_f)_{f \in F})$ where $N = \{1, 2, 3\}$, $F = \{f, g, h\}$, $X_1 = \big\{\{f\}, \{g\}\big\}$, $X_2 = \big\{\{f, h\}\big\}$, $X_3 = \big\{\{f\}\big\}$, and $c_f = c_g = c_h = c_\alpha$. Furthermore, let $d_1 = x_0$, $d_2 = y_0$ and $d_3 = z_\alpha$ and consider the corresponding weighted congestion game $G = (N, X, \pi)$. For the path

$$\gamma = \big((\{g\}, \{h\}, \{f\}), (\{f\}, \{h\}, \{f\}), (\{f\}, \{f\}, \{f\}), (\{g\}, \{f\}, \{f\}), (\{g\}, \{h\}, \{f\})\big)$$

in X we get for any strictly positive vector w

$$\begin{aligned} I(\gamma, \pi/w) &= \left(\frac{y_0}{w_2} - \frac{x_0}{w_1}\right) c_\alpha(x_0 + y_0 + z_\alpha) - \frac{y_0}{w_2} c_\alpha(y_0 + z_\alpha) + \frac{x_0}{w_1} c_\alpha(x_0 + z_\alpha) \\ &= \frac{x_0}{w_1} \big(c_\alpha(x_0 + z_\alpha) - c_\alpha(x_0 + y_0 + z_\alpha)\big) \neq 0. \end{aligned}$$

The second equality follows from the assumption $c_\alpha(x_0 + y_0 + z_\alpha) = c_\alpha(y_0 + z_\alpha)$ and the contradiction follows from $c_\alpha(x_0 + y_0 + z) - c_\alpha(x_0 + z_\alpha) \neq 0$. Using the criterion of Monderer and Shapley the game does not admit a weighted potential and we thus may assume in the following that $c_\alpha(x_0 + y_0 + z_\alpha) \neq c_\alpha(y_0 + z_\alpha)$ for all $c_\alpha \in \mathcal{C}$ and $z_\alpha \in \mathbb{R}_{\geq 0}$.

Let α be arbitrary and consider $c_\alpha \in \mathcal{C}$ and $z_\alpha \in \mathbb{R}_{\geq 0}$ such that (12) does not hold. As we may assume that $c_\alpha(x_0 + y_0 + z_\alpha) \neq c_\alpha(y_0 + z_\alpha)$, the value

$$\beta = \frac{x_0}{y_0} \cdot \frac{c_\alpha(x_0 + y_0 + z_\alpha) - c_\alpha(x_0 + z_\alpha)}{c_\alpha(x_0 + y_0 + z_\alpha) - c_\alpha(y_0 + z_\alpha)}$$

is well defined. If $\beta < 0$, it follows that $I(\gamma, \pi \,/\, w) \neq 0$ for any strictly positive vector w. If $\beta > 0$, we find $c_\beta \in \mathcal{C}$ and z_β such that

$$\beta \cdot \big(c_\beta(x_0 + y_0 + z_\beta) - c_\beta(y_0 + z_\beta)\big) \neq \frac{x_0}{y_0} \cdot \big(c_\beta(x_0 + y_0 + z) - c_\beta(x_0 + z_\beta)\big).$$

Finally, let us consider the congestion model $\mathcal{M} = (N, F, X, (c_f)_{f \in F})$, where $N = \{1, 2, 3, 4\}$, $F = \{f, g, h, \iota\}$, $X_1 = \big\{\{f\}, \{h\}, \{\iota\}\big\}$, $X_2 = \big\{\{g\}, \{h\}, \{\iota\}\big\}$, $X_3 = \big\{\{h\}\big\}$, $X_4 = \big\{\{\iota\}\big\}$ and $c_f = c_g = c_h = c_\alpha, c_\iota = c_\beta$. Now we regard the game $G(\mathcal{M}) = (N, X, \pi)$ with demands $d_1 = x_0$, $d_2 = y_0$, $d_3 = z_\alpha$, $d_4 = z_\beta$. Assuming that G admits a weighted potential we can find a strictly positive vector w such that G^w admits an exact potential. To this end, we will apply the criterion of Monderer and Shapley to the paths

$$\begin{aligned}\gamma_1 = \big(&(\{f\}, \{g\}, \{h\}, \{\iota\}), (\{h\}, \{g\}, \{h\}, \{\iota\}), (\{h\}, \{h\}, \{h\}, \{\iota\}),\\ &(\{f\}, \{h\}, \{h\}, \{\iota\}), (\{f\}, \{g\}, \{h\}, \{\iota\})\big),\\ \gamma_2 = \big(&(\{f\}, \{g\}, \{h\}, \{\iota\}), (\{\iota\}, \{g\}, \{h\}, \{\iota\}), (\{\iota\}, \{\iota\}, \{h\}, \{\iota\}),\\ &(\{f\}, \{\iota\}, \{h\}, \{\iota\}), (\{f\}, \{g\}, \{h\}, \{\iota\})\big),\end{aligned}$$

and compute that

$$I(\gamma_1, \pi/w) = \left(\frac{y_0}{w_2} - \frac{x_0}{w_1}\right) c_\alpha(x_0 + y_0 + z_\alpha) - \frac{y_0}{w_2} c_\alpha(y_0 + z_\alpha) + \frac{x_0}{w_1} c_\alpha(x_0 + z_\alpha) = 0, \tag{13}$$

$$I(\gamma_2, \pi/w) = \left(\frac{y_0}{w_2} - \frac{x_0}{w_1}\right) c_\beta(x_0 + y_0 + z_\beta) - \frac{y_0}{w_2} c_\beta(y_0 + z_\beta) + \frac{x_0}{w_1} c_\beta(x_0 + z_\beta) = 0. \tag{14}$$

We derive from equations (13) and (14) that

$$\beta = \frac{x_0}{y_0} \cdot \frac{c_\alpha(x_0 + y_0 + z_\alpha) - c_\alpha(x_0 + z_\alpha)}{c_\alpha(x_0 + y_0 + z_\alpha) - \tilde{c}_1(y_0 + z_\alpha)} = \frac{w_1}{w_2} = \frac{x_0}{y_0} \cdot \frac{c_\beta(x_0 + y_0 + z_\beta) - c_\beta(x_0 + z_\beta)}{c_\beta(x_0 + y_0 + z_\beta) - c_\beta(y_0 + z_\beta)} \neq \beta,$$

which is a contradiction. $\square$

Although condition (10) seems to be similar to the functional equation (8) characterizing the existence of an exact potential, it is not possible to proceed using differential equations. As $\alpha(x, y)$ need not be bounded it is not possible to prove continuity and differentiability of c. Instead, we will use the discrete counterpart of differential equations, that is, difference equations.

Theorem 3.9. Let $\mathcal{C}$ be a set of continuous functions. Let $\mathcal{G}(\mathcal{C})$ be the set of weighted congestion games with cost functions in $\mathcal{C}$. Then every $G \in \mathcal{G}(\mathcal{C})$ admits a w-potential if and only if exactly one of the following cases holds:

(1) $\mathcal{C}$ contains only affine functions,
(2) $\mathcal{C}$ contains only exponential functions $c(\ell) = a_c\, e^{\phi \ell} + b_c$ for some $a_c, b_c, \phi \in \mathbb{R}$, where a_c and b_c may depend on c, while ϕ must be equal for every $c \in \mathcal{C}$.

Proof. First, we will prove that these functions guarantee the existence of a w-potential in all such games. We have shown in Section 3.1 that affine cost functions c_f give rise to an exact potential. As every exact potential function is also a w-potential for $w = (1, \ldots, 1)$, we may conclude that affine cost functions give rise to a weighted potential in weighted congestion games.

So let us check the case $c(\ell) = a_c\, e^{\phi\, \ell} + b_c$ for $\phi \neq 0$. It is easy to verify that

$$\alpha(x, y) = \frac{x}{y} \cdot \frac{a_c e^{\phi(x+y+z)} + b_c - a_c e^{\phi(x+z)} - b_c}{a_c e^{\phi(x+y+z)} + b_c - a_c e^{\phi(y+z)} - b_c} = \frac{x}{y} \cdot \frac{e^{\phi\,(x+y)} - e^{\phi\,(x)}}{e^{\phi\,(x+y)} - e^{\phi\,(y)}} > 0.$$

Note in particular that $\alpha(x, y)$ does neither depend on a_c, b_c, nor z. Thus, it is unambiguously defined and strictly positive. Theorem 3.8 then yields the result.

To show the opposite direction, we assume that the conditions on $\mathcal{C}$ do not hold but that every $G \in \mathcal{G}(\mathcal{C})$ admits a w-potential.

First, suppose that there is a function $\tilde{c} \in \mathcal{C}$ that is neither affine nor exponential. This implies that there are four points $p_1 < p_2 < p_3 < p_4$ following neither an exponential nor a affine law, that is, there are neither a, b and $\phi \in \mathbb{R}$ such that

$$\tilde{c}(p_1) = ae^{\phi p_1} + b, \qquad \dots, \qquad \tilde{c}(p_4) = ae^{\phi p_4} + b,$$

nor are there s and $t \in \mathbb{R}$ such that

$$\tilde{c}(p_1) = sp_1 + t, \qquad \dots, \qquad \tilde{c}(p_4) = sp_4 + t.$$

As $\tilde{c}$ is continuous, we may assume without loss of generality that the above conditions hold for rational $p_1, \dots, p_4$ and we can write them as $p_1 = 2m_1/(2\,k), \dots, p_4 = 2m_4/(2\,k)$ for some $m_1, m_2, m_3, m_4, k \in \mathbb{N}$.

We regard a congestion model $\mathcal{M} = (N = \{1,2,3\}, F, X, c)$ and a series of games $G_m(\mathcal{M}) = (N, X, \pi), 0 \leq m \leq 2\,m_4$. We set the demands of the players as $d_1 = 1/(2\,k)$, $d_2 = 2/(2\,k)$ and $d_3 = m/(2\,k)$. By assumption each game G_m admits a w-potential. By Lemma 3.8 this implies that for each game

$$\alpha(d_1, d_2) = \frac{d_1}{d_2} \cdot \frac{\tilde{c}(d_1 + d_2 + d_3) - \tilde{c}(d_1 + d_3)}{\tilde{c}(d_1 + d_2 + d_3) - \tilde{c}(d_2 + d_3)} = \frac{d_1}{d_2} \cdot \frac{\tilde{c}(d_1 + d_2) - \tilde{c}(d_1)}{\tilde{c}(d_1 + d_2) - \tilde{c}(d_2)}.$$

In particular $\alpha(d_1, d_2)$ is the same for each game G_m. Now, we introduce $f_n = \tilde{c}(n/(2\,k))$ and consider the sequence $(f_n)_{n \in N}$. Thus, we can write

$$\alpha(d_1, d_2) = \frac{1}{2} \cdot \frac{f_{m+3} - f_{m+1}}{f_{m+3} - f_{m+2}}.$$

If $\alpha(d_1, d_2) = 1/2$, we conclude that $\tilde{c}$ is constant, which contradicts our assumption. So we may assume that $\alpha(d_1, d_2) \neq 1/2$ and we obtain

$$f_{m+3} - \frac{2\alpha(d_1, d_2)}{2\alpha(d_1, d_2) - 1} f_{m+2} + \frac{1}{2\alpha(d_1, d_2) - 1} f_{m+1} = 0. \tag{15}$$

Equation (15) defines a recursively defined sequence on $\{1, \dots, 2m_4\}$.

The main result in [[5], Chapter 4] gives sufficient conditions on the uniqueness of the general solution of such sequences. First, we define the characteristic equation of a general second-order recurrence relation $a_{m+2} + \beta_2 a_{m+1} + \beta_1 a_m = 0$ as $x^2 + \beta_2 x + \beta_1 = 0$.

Now let x_1 and x_2 be the distinct and real roots of the characteristic equation. Then every general solution a_m of the recurrence relation is a linear combination with coefficients independent of m of powers $(x_i)^m$ of the solutions $x_i, i = 1, 2$. In addition, if x is the double root of the characteristic equation, every general solution a_m of the recurrence relation is a linear combination of x^m and mx^m. In both cases, if two consecutive initial values a_k and a_{k+1} of the recurrence relation are known, a solution can be obtained by evaluating the two constants of the linear combination using the two initial values and the fact that this solution is unique.

The characteristic equation of the recurrence relation (15) equals

$$x^2 - \frac{2\alpha(d_1, d_2)}{2\alpha(d_1, d_2) - 1} x + \frac{1}{2\alpha(d_1, d_2) - 1} = (x - 1) \left(x - \frac{1}{2\alpha(d_1, d_2) - 1} \right).$$

So if $\alpha(d_1, d_2) \neq 1$, two different roots occur and f_m can be computed explicitely and uniquely for even m as

$$f_m = b \cdot 1^m + a \cdot \left(\frac{1}{2\alpha(d_1, d_2) - 1} \right)^m = b + a \cdot exp \left(m \ln \left(\frac{1}{2\alpha(d_1, d_2) - 1} \right) \right)$$

for some constants a and $b \in \mathbb{R}$. If $\alpha(d_1, d_2) = 1$, we can evaluate f_m as

$$f_m = t \cdot 1^m + m\, s \cdot 1^m = t + s\, m$$

for some constants $s, t \in \mathbb{R}$ showing that $\tilde{c}$ follows either an exponential or affine law on $p_1, \ldots, p_4$. So it remains to show that neither affine and exponential functions nor exponential function with different exponents can occur simultaneously.

Let us first assume on the contrary that $\mathcal{C}$ contains both affine and exponential functions, that is, there are $\tilde{c}_1, \tilde{c}_2 \in \mathcal{C}$ such that $\tilde{c}_1(\ell) = s\,\ell + t$ for some constants $s, t \in \mathbb{R}$ and $\tilde{c}_2 = a\,e^{\phi \ell} + b$ for some constants $a, b \in \mathbb{R}$ and $\phi \neq 0$. Let us fix $x = 1$ and $y = 2$. We calculate that

$$\alpha_1(1,2) = \frac{1}{2} \cdot \frac{\tilde{c}_1(3+z) - \tilde{c}_1(1+z)}{\tilde{c}_1(3+z) - \tilde{c}_1(2+z)} = 1, \tag{16}$$

while

$$\alpha_2(1,2) = \frac{1}{2} \cdot \frac{\tilde{c}_2(3+z) - \tilde{c}_2(1+z)}{\tilde{c}_2(3+z) - \tilde{c}_2(2+z)} = \frac{1}{2} \cdot \frac{e^{3\phi} - e^{\phi}}{e^{3\phi} - e^{2\phi}} = \frac{1}{2}\left(1 + e^{-\phi}\right) \neq 1. \tag{17}$$

Thus, $\alpha(1,2)$ is not independent of $\tilde{c}_1$ and $\tilde{c}_2$, respectively, which contradicts Theorem 3.8.

To finish the proof, let us finally assume that $\mathcal{C}$ contains two exponential functions with different exponents $\tilde{c}_2, \tilde{c}_3 \in \mathcal{C}$ where $\tilde{c}_2 = a\,e^{\phi \ell} + b$ $\tilde{c}_3(\ell) = s\,e^{\psi \ell} + t$ for some constants $a, b, s, t \in R$ and $0 \neq \psi \neq \phi \neq 0$. As in (17), we obtain, $\alpha_2(1,2) = (1 + e^{-\phi})\,/\,2$ and $\alpha_3(1,2) = (1 + e^{-\psi})\,/\,2$. Using that the exponential function is bijective, we derive $\alpha_2(1,2) \neq \alpha_3(1,2)$ for $\phi \neq \psi$, which is a contradiction to the conditions of Lemma 3.8. $\square$

Panagopoulou and Spirakis [27] showed that the function $P : X \to \mathbb{R}$ defined as $P(x) = \sum_{f \in F} c_f(x)$ is a w-potential if the cost functions on all facilities are equal the exponential function, that is, $c_f(x) = e^x$ for all $f \in F$. It directly follows that $P(x)$ is also a w-potential if the cost functions are of type $c_f(x) = a_f e^{\phi x}$ for all $f \in F$, where ϕ is a constant that does not depend on the facility. The function $P(x)$ is not a $w-$potential (not even a generalized ordinal potential) if the costs are of type $c_f(x) = a_f e^{\phi x} + b_f$ for all $f \in F$. For this more general case, the function $\tilde{P} : X \to \mathbb{R}$ defined as $\tilde{P}(x) = \sum_{f \in F} c_f(x) + \sum_{i \in N} \sum_{f \in x_i} \frac{e^{\phi d_i} - 1}{e^{\phi}} b_f$ is a w-potential. The proof uses standard arguments and is omitted.

3.3. Implications of Our Characterizations

It is natural to ask whether these results remain valid if additional restrictions on the set $\mathcal{G}(\mathcal{C})$ are made. A natural restriction is to assume that all players have an integral demand. As we used infinitesimally small demands in the proof of Lemma 3.6, our results for exact potentials do not apply directly to integer demands. With a slight variation of the proof of Theorem 3.9, where only the case $\alpha(\cdot,\cdot) = 1$ is considered, however, we still obtain the same result provided $\mathcal{C}$ contains only continuous functions.

Another natural restriction on $\mathcal{G}(\mathcal{C})$ are games with symmetric sets of strategies or games with a bounded number of players or facilities. Since the proofs of Lemma 3.6 and 3.8 and Theorems 3.7 and 3.9 rely on mild assumptions, we can strengthen our characterizations as follows.

Corollary 3.10. Let $\mathcal{C}$ be a set of continuous functions. Let $\mathcal{G}(\mathcal{C})$ be the set of weighted congestion games with cost functions in $\mathcal{C}$ satisfying one or more of the following properties

(1) Each game $G = (N, X, \pi) \in \mathcal{G}(\mathcal{C})$ has two (four) players.
(2) Each game $G = (N, X, \pi) \in \mathcal{G}(\mathcal{C})$ has three (four) facilities.
(3) For each game $G = (N, X, \pi) \in \mathcal{G}(\mathcal{C})$ and each player $i \in N$ the set of her strategies X_i contains a single facility only.

(4) Each game $G = (N, X, \pi) \in \mathcal{G}(\mathcal{C})$ has symmetric strategies, that is $X_i = X_j$ for all $i, j \in N$.
(5) In each game $G = (N, X, \pi) \in \mathcal{G}(\mathcal{C})$, the demands of all players are integral.

Then, every $G = (N, X, \pi) \in \mathcal{G}(\mathcal{C})$ has an exact potential (a w-potential) if and only if $\mathcal{C}$ contains only affine functions (only affine functions or only exponential functions as in Theorem 3.8). Note that for the case of a w-potential, the conditions in parentheses must hold.

We remark that the conditions given in Corollary 3.10 are not necessary for the existence of a generalized ordinal potential function, see Monderer and Shapley [25] for a definition. In fact, for singleton congestion games with non-decreasing cost functions, there exists a generalized ordinal potential function, see Fotakis et al. [13].

Yet, we are able to deduce an interesting result concerning the existence of w-potentials in weighted congestion games, where each facility can be chosen by at most two players. By adapting the proof of Lemma 3.8 for two-player games, the following lemma follows.

Lemma 3.11. Let $\mathcal{C}$ be a set of functions and let $\mathcal{G}^2(\mathcal{C})$ be the set of weighted congestion games where each facility lies in the strategy sets of at most two players and cost functions are in $\mathcal{C}$. Every $G \in \mathcal{G}^2(\mathcal{C})$ has a w-potential if and only if for all $x, y \in \mathbb{R}_{>0}$ there exists an $\alpha(x, y) \in \mathbb{R}_{>0}$ such that

$$\alpha(x, y) \cdot \big(c(x + y) - c(y)\big) = \frac{x}{y} \cdot \big(c(x + y) - c(x)\big) \tag{18}$$

for all $c \in \mathcal{C}$.

Using this lemma, we can prove that in games where each facility lies in the strategy sets of at most two players also non-affine and non-exponential functions give rise to a weighted potential.

Theorem 3.12. Let $m : \mathbb{R}_{>0} \to \mathbb{R}$ be a strictly monotonic function and let $\mathcal{C}_m = \{a\, m(x) + b : a, b \in \mathbb{R}\}$. Let $\mathcal{G}^2(\mathcal{C}_m)$ be the set of weighted congestion games where each facility lies in the strategy sets of at most two players and cost functions are in $\mathcal{C}_m$. Then every such game $G \in \mathcal{G}^2(\mathcal{C}_m)$ admits a w-potential.

Proof. Let $c \in \mathcal{C}_m$ be arbitrary. By definition of $\mathcal{C}_m$, we can write $c(x) = a_c\, m(x) + b_c$ for some $a_c, b_c \in \mathbb{R}$. If $a_c = 0$, the function c is constant and thus fulfills the requirements of Lemma 3.11. If $a_c \neq 0$, it is easy to check that

$$\begin{aligned} \alpha &= \frac{x}{y} \cdot \frac{c(x + y) - c(x)}{c(x + y) - c(y)} = \frac{x}{y} \cdot \frac{a_c\, m(x + y) + b_c - (a_c\, m(x) + b_c)}{a_c\, m(x + y) + b_c - (a_c\, m(y) + b_c)} \\ &= \frac{x}{y} \cdot \frac{m(x + y) - m(x)}{m(x + y) - m(y)} > 0 \end{aligned}$$

for all $c \in \mathcal{C}_m$ and hence the conditions of Lemma 3.11 are fulfilled implying the existence of a w-potential. □

This result generalizes a result of Anshelevich et al. in [3], who showed that a weighted congestion game with two players and $c_f(\ell) = b_f/\ell$ for a constant $b_f \in \mathbb{R}_{>0}$ has a potential. Moreover, this result shows that the characterization of Corollary 3.10 is tight in the sense that weighted congestion games with two players admit a w-potential even if cost functions are neither affine nor exponential.

4. Extensions of the Model

In the last section, we developed a new technique to characterize the set of functions that give rise to a potential in weighted congestion games. In this section, we will introduce two generalizations of weighted congestion games and investigate the set of cost functions that assure the existence of potential functions.

Definition 4.1 (Facility-dependent demands)**.** Let $\mathcal{M} = (N, F, X, (c_f)_{f\in F})$ be a congestion model and let $(d_{i,f})_{i\in N, f\in F}$ be a matrix of facility-dependent demands. The corresponding *weighted congestion game with facility-dependent demands* is the strategic game denoted by $G(\mathcal{M}) = (N, X, \pi)$, where π is defined as $\pi = \bigtimes_{i\in N} \pi_i$, $\pi_i(x) = \sum_{f\in x_i} d_{i,f}\, c_f\big(\ell_f(x)\big)$ and $\ell_f(x) = \sum_{j\in N: f\in x_j} d_{j,f}$.

Restricting the strategy sets to singletons, we obtain scheduling games. In a scheduling game, players are jobs that have machine-dependent demands and can be scheduled on a set of admissible machines (restricted scheduling on unrelated machines). In contrast to the classical approach, where each job strives to minimize its makespan, we consider a different private cost function: Machines charge a price per unit given by a load-dependent cost function c_f and each job minimizes its cost defined as the price of the chosen machine multiplied with its machine-dependent demand.

Theorem 4.2. Let $\mathcal{C}$ be a set of continuous functions and let $\mathcal{G}^{fd}(\mathcal{C})$ be the set of weighted congestion games with facility-dependent demands and cost functions in $\mathcal{C}$. Then, every $G \in \mathcal{G}^{fd}(\mathcal{C})$ admits a w-potential if and only if $\mathcal{C}$ contains only affine functions, that is, every $c \in \mathcal{C}$ can be written as $c(\ell) = a_c\,\ell + b_c$ for some $a_c, b_c \in \mathbb{R}$. For a game G with affine cost functions, the potential function is given by $P(x) = \sum_{i\in N}\sum_{f\in x_i} c_f\left(\sum_{j\in\{1,\dots,i\}: f\in x_j} d_{j,f}\right) d_{i,f}$.

Proof. For any set of functions $\mathcal{C}$, the set $\mathcal{G}(\mathcal{C})$ of weighted congestion games with cost functions in $\mathcal{C}$ is contained in the set of weighted congestion games with facility-dependent demands. Thus, we can restrict $\mathcal{C}$ to the set of affine functions or exponential functions as in Theorem 3.9.

We first show that if $\mathcal{C}$ contains an exponential function, then there is a weighted congestion game with facility-dependent demands that does not admit a weighted potential. To this end, suppose that there is a cost function $\tilde{c} \in \mathcal{C}$ that can be written as $\tilde{c}(\ell) = a_c e^{\phi\ell} + b_c$ for some $a_c, b_c, \phi \in \mathbb{R}$ with $a_c \neq 0$ and $\phi \neq 0$. We consider the congestion model $\mathcal{M} = (N, F, X, (c_f)_{f\in F})$, where $N = \{1, 2\}$, $F = \{f, g, h, \iota\}$, $X_1 = \big\{\{f\}, \{h\}, \{\iota\}\big\}$, $X_2 = \big\{\{g\}, \{h\}, \{\iota\}\big\}$, and $c_f = c_g = c_h = c_\iota = \tilde{c}$. In addition, we specify $d_{1,\iota} = 2\,/\,\phi$ and $d_{i,f} = 1\,/\,\phi$ for all $(i, f) \in N \times F \setminus \{(1, \iota)\}$. Let $G = (N, X, \pi)$ denote the corresponding weighted congestion game with facility-dependent demands. Regarding the 4-cycle

$$\begin{aligned}\gamma_1 = \big(&(\{f\}, \{g\}, \{h\}, \{\iota\}), (\{h\}, \{g\}, \{h\}, \{\iota\}), (\{h\}, \{h\}, \{h\}, \{\iota\}),\\ &(\{f\}, \{h\}, \{h\}, \{\iota\}), (\{f\}, \{g\}, \{h\}, \{\iota\})\big)\end{aligned}$$

we obtain $I(\gamma_1, \pi\,/\,w) = 0$ if and only if $w_1 = w_2$. In contrast, for the 4-cycle

$$\begin{aligned}\gamma_2 = \big(&(\{f\}, \{g\}, \{h\}, \{\iota\}), (\{\iota\}, \{g\}, \{h\}, \{\iota\}), (\{\iota\}, \{\iota\}, \{h\}, \{\iota\}),\\ &(\{f\}, \{\iota\}, \{h\}, \{\iota\}), (\{f\}, \{g\}, \{h\}, \{\iota\})\big)\end{aligned}$$

we derive $I(\gamma_2, \pi\,/\,w) = 0$ if and only if the equation $w_1\,/\,w_2 = (2e^3 - 2e^2)\,/\,(e^3 - e^1)$ is fulfilled. Thus, $I(\gamma_1, \pi\,/\,w) = I(\gamma_2, \pi\,/w\,) = 0$ implies $0 = e^2 - 2e + 1$, a contradiction. We conclude that G does not admit a weighted potential. We proceed by showing that $P(x)$ is an exact potential for affine costs.

Assume $c_f(\ell) = a_f\ell + b_f$, with $a_f, b_f \in \mathbb{R}$ for all $f \in F$. We define the function $c_f^{\leq i}(x) = c_f(\sum_{j\in\{1,\dots,i\}: f\in x_j} d_{j,f})$ and rewrite $P(x)$ as $P(x) = \sum_{i\in N} P_i(x)$, where $P_i(x) = \sum_{f\in x_i} c_f^{\leq i}(x) d_{i,f}$. Let $G = (N, X, \pi)$ be an arbitrary weighted congestion game with facility-dependent demands and let $x, y \in X$ be two strategy profiles such that $x = (x_k, x_{-k})$ and $y = (y_k, y_{-k})$ with $x_{-k} = y_{-k}$ for some $x_k, y_k \in X_k$ and $x_{-k} \in X_{-k}$. We notice that $P_i(x) = P_i(y)$ for all

$i < k$. Now consider a player $i > k$. When computing $P_i(x) - P_i(y)$, we observe that all costs corresponding to facilities not contained in $x_k \cup y_k$ cancel out. For each facility $f \in (x_i \cap x_k) \setminus y_k$, we get $c_f^{\leq i}(x) - c_f^{\leq i}(y) = a_f\, d_{k,f}$. Analogously, for each facility $f \in (x_i \cap y_k) \setminus x_k$, it holds that $c_f^{\leq i}(x) - c_f^{\leq i}(y) = -a_f d_{k,f}$. For each facility $f \in x_i \cap x_k \cap y_k$, we have $c_f^{\leq i}(x) = c_f^{\leq i}(y)$. Hence,

$$P_i(x) - P_i(y) = \sum_{f \in x_i \cap x_k} a_f\, d_{k,f}\, d_{i,f} - \sum_{f \in x_i \cap y_k} a_f\, d_{k,f}\, d_{i,f}.$$

Moreover, we can calculate straightforwardly that

$$P_k(x) - P_k(y) = \sum_{f \in x_k} c_f \Big(\sum_{j \in \{1,\dots,k\}: f \in x_j} d_{j,f} \Big) d_{k,f} - \sum_{f \in y_k} c_f \Big(\sum_{j \in \{1,\dots,k\}: f \in y_j} d_{j,f} \Big) d_{k,f}.$$

We thus obtain

$$\begin{aligned}
&P(x) - P(y) \\
&\quad = \sum_{i \in N} P_i(x) - \sum_{i \in N} P_i(y) \\
&\quad = \sum_{i>k}^{n} \Big(\sum_{f \in x_i \cap x_k} a_f d_{k,f} d_{i,f} - \sum_{f \in y_i \cap y_k} a_f d_{k,f} d_{i,f} \Big) + \sum_{f \in x_k} a_f \Big(\sum_{j \in \{1,\dots,k\}: f \in x_j} d_{j,f} \Big) d_{k,f} \\
&\qquad - \sum_{f \in y_k} a_f \Big(\sum_{j \in \{1,\dots,k\}: f \in y_j} d_{j,f} \Big) d_{k,f} + d_{k,f} \sum_{f \in x_k} b_f - d_{k,f} \sum_{f \in y_k} b_f \\
&\quad = \sum_{f \in x_k} a_f \Big(\sum_{j \in \{1,\dots,n\}: f \in x_j} d_{j,f} \Big) d_{k,f} - \sum_{f \in y_k} a_f \Big(\sum_{j \in \{1,\dots,n\}: f \in y_j} d_{j,f} \Big) d_{k,f} \\
&\qquad + d_{k,f} \sum_{f \in x_k} b_f - d_{k,f} \sum_{f \in y_k} b_f \\
&\quad = \pi_k(x) - \pi_k(y).
\end{aligned}$$

Hence, P is an exact potential function. $\square$

Note that the potential function used is a natural generalization of Rosenthal's potential function [28]. We will now introduce an extension to weighted congestion games allowing players to also choose their demand.

Definition 4.3 (Elastic demands)**.** Let $\mathcal{M} = (N, F, X, (c_f)_{f \in F})$ be a congestion model. Together with $D = \bigtimes_{i \in N} D_i$, where $D_i \subset \mathbb{R}_{>0}$ are compact for all $i \in N$, we define the *weighted congestion game with elastic demands* as the strategic game $G(\mathcal{M}) = (N, \bar{X}, \pi)$ with $\bar{X} = (X, D)$, $\pi = \bigtimes_{i \in N} \pi_i$, and $\pi_i(\bar{x}) = \sum_{f \in x_i} d_i c_f\big(\ell_f(\bar{x})\big)$ and $\ell_f(\bar{x}) = \sum_{j \in N: f \in x_j} d_j$. A strategy of player i is a tuple $\bar{x}_i = (x_i, d_i)$ where $x_i \in X_i$ and $d_i \in D_i$.

In our definition of weighted congestion games with elastic demands, we explicitly allow for positive and negative, and for increasing and decreasing cost functions. Thus, an increase in the demand may increase or decrease the player's private cost. Note that in weighted congestion games with elastic demands, the strategy sets are topological spaces and are in general infinite. By restricting the sets D_i to singletons $D_i = \{d_i\}, i \in N$, we obtain weighted congestion games as a special case of weighted congestion games with elastic demands. The proof of the following result is omitted as it is similar to the case of facility-dependent demands.

Theorem 4.4. Let $\mathcal{C}$ be a set of continuous functions and let $\mathcal{G}^e(\mathcal{C})$ be the set of weighted congestion games with elastic demands and cost functions in $\mathcal{C}$. Then, every $G \in \mathcal{G}^e(\mathcal{C})$ admits a w-potential function if and only if $\mathcal{C}$ contains only affine functions. For a game G with affine cost functions, the potential function is given by the function

$$P(\bar{x}) = \sum_{i \in N} \sum_{f \in x_i} c_f \left(\sum_{j \in \{1,\dots,i\}: f \in x_j} d_j \right) d_i.$$

As an immediate consequence, we obtain the existence of a PNE if cost functions are affine. Note that the existence of a potential is not sufficient for proving existence of a PNE as we are considering infinite games. However, as $\bar{X}$ is compact and P is continuous, P has a minimum $\bar{x}^* \in \bar{X}$ and $\bar{x}^*$ is a PNE.

Corollary 4.5. Let $\mathcal{C}$ be a set of affine functions and let $\mathcal{G}^e(\mathcal{C})$ be the set of weighted congestion games with elastic demands and cost functions in $\mathcal{C}$. Then, every $G \in \mathcal{G}^e(\mathcal{C})$ admits a PNE.

References

[1] H. Ackermann, H. Röglin, and B. Vöcking, *Pure Nash equilibria in player-specific and weighted congestion games*, Theor. Comput. Sci. **410**, no. 17 (2009), pp. 1552–1563.

[2] S. Albers, S. Eilts, E. Even-Dar, Y. Mansour, and L. Roditty, *On nash equilibria for a network creation game*, in Proc. of the 17th Symposium on Discrete Algorithms (SODA), 2006, pp. 89–98.

[3] E. Anshelevich, A. Dasgupta, J. Kleinberg, É. Tardos, T. Wexler, and T. Roughgarden, *The price of stability for network design with fair cost allocation*, SIAM J. Comput. **38**, no. 4 (2008), pp. 1602–1623.

[4] B. Awerbuch, Y. Azar, A. Epstein, V. S. Mirrokni, and A. Skopalik, *Fast convergence to nearly optimal solutions in potential games*, in EC, 2008, pp. 264–273.

[5] V. K. Balakrishnan, *Introductory Discrete Mathematics*, Courier Dover Publications, 1996.

[6] G. Brown, *Iterative Solutions of Games by Fictitious Play*, Wiley, 1951, ch. 24.

[7] H. Chen and T. Roughgarden, *Network design with weighted players*, Theor. Comput. Syst. **45**, no. 2 (2009), pp. 302–324.

[8] R. Cole, Y. Dodis, and T. Roughgarden, *Bottleneck links, variable demand, and the tragedy of the commons*, in Proc. 17th Annual ACM-SIAM Sympos. on Discrete Algorithms, 2006, pp. 668–677.

[9] A. Cournot, *Recherches Sur Les Principes Mathematiques De La Theorie De La Richesse*, Hachette, Paris, 1838.

[10] E. Even-Dar, A. Kesselman, and Y. Mansour, *Convergence time to Nash equilibrium in load balancing*, ACM Trans. Algorithms **3**, no. 3 (2007), p. 32.

[11] A. Fabrikant, C. Papadimitriou, and K. Talwar, *The complexity of pure Nash equilibria*, in Proc. 36th Annual ACM Sympos. Theory Comput., 2004, pp. 604–612.

[12] D. Fotakis, A. Kaporis, and P. Spirakis, *Atomic congestion games: Fast, myopic and concurrent*, in Proc. 1st Internat. Sympos. Algorithmic Game Theory, 2008, pp. 121–132.

[13] D. Fotakis, S. Kontogiannis, E. Koutsoupias, M. Mavronicolas, and P. Spirakis, *The structure and complexity of Nash equilibria for a selfish routing game*, in Proc. 29th Internat. Colloquium on Automata, Languages and Programming, P. Widmayer, F. Triguero, R. Morales, M. Hennessy, S. Eidenbenz, and R. Conejo, eds., LNCS 2380, 2002, pp. 123 – 134.

[14] D. Fotakis, S. Kontogiannis, and P. Spirakis, *Selfish unsplittable flows*, Theor. Comput. Sci. **348**, no. 2-3 (2005), pp. 226–239.

[15] D. Fotakis, S. Kontogiannis, and P. Spirakis, *Atomic congestion games among coalitions*, in Proc. 33rd Internat. Colloquium on Automata, Languages and Programming, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., LNCS 4052, 2006, pp. 572–583.

[16] M. Gairing, B. Monien, and K. Tiemann, *Routing (un-) splittable flow in games with player-specific linear latency functions*, in Proc. 33rd Internat. Colloquium on Automata, Languages and Programming, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., LNCS 4051, 2006, pp. 501–512.

[17] M. Goemans, V. Mirrokni, and A. Vetta, *Sink equilibria and convergence*, in Proc. 46th Annual IEEE Sympos. Foundations Comput. Sci., 2005, pp. 142–154.

[18] T. Harks and M. Klimm, *On the existence of pure Nash equilibria in weighted congestion games*, in Proc. 37rd Internat. Colloquium on Automata, Languages and Programming, S. Abramsky, C. Gavoille, C. Kirchner, F. M. auf der Heide, and P. Spirakis, eds., LNCS 6198, 2010, pp. 79 – 89.

[19] S. Ieong, R. McGrew, E. Nudelman, Y. Shoham, and Q. Sun, *Fast and compact: A simple class of congestion games*, in Proc. 20th National Conf. Artificial Intelligence and the 17th Innovative Appl. Artificial Intelligence Conf., 2005, pp. 489–494.

[20] F. Kelly, A. Maulloo, and D. Tan, *Rate Control in Communication Networks: Shadow Prices, Proportional Fairness, and Stability*, J. Oper. Res. Soc. **49** (1998), pp. 237–52.

[21] L. Libman and A. Orda, *Atomic resource sharing in noncooperative networks*, Telecommunication Systems **17**, no. 4 (2001), pp. 385–409.

[22] M. Mavronicolas, I. Milchtaich, B. Monien, and K. Tiemann, *Congestion games with player-specific constants*, in Proc. 32nd Internat. Sympos. Math. Foundations Comput. Sci., L. Kucera and A. Kucera, eds., LNCS 4708, 2007.

[23] I. Milchtaich, *Congestion games with player-specific payoff functions*, Games Econom. Behav. **13**, no. 1 (1996), pp. 111 – 124.

[24] I. Milchtaich, *The equilibrium existence problem in finite network congestion games*, in Proc. 2nd Internat. Workshop on Internet and Network Econom., M. Mavronicolas and S. Kontogiannis, eds., LNCS 4286, 2006, pp. 87 – 98.

[25] D. Monderer and L. Shapley, *Potential games*, Games Econom. Behav. **14**, no. 1 (1996), pp. 124 – 143.

[26] D. Monderer and L. S. Shapley, *Fictitious play property for games with identical interests*, Journal of Economic Theory **68**, no. 1 (1996), pp. 258 – 265.

[27] P. Panagopoulou and P. Spirakis, *Algorithms for pure Nash equilibria in weighted congestion games*, ACM J. Exp. Algorithmics **11**, no. 2.7 (2006), pp. 1–19.

[28] R. Rosenthal, *A class of games possessing pure-strategy Nash equilibria*, Internat. J. Game Theory **2**, no. 1 (1973), pp. 65–67.

[29] B. Vöcking, *Selfish load balancing*, in Algorithmic Game Theory, N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, eds., Cambridge University Press, 2007, ch. 20, pp. 517–542.

Paper 6

On the Existence of Pure Nash Equilibria in Weighted Congestion Games

Tobias Harks and Max Klimm
On the Existence of Pure Nash Equilibria in Weighted Congestion Games
Math. Oper. Res. **37** (2012), no. 3, pp. 419-436

Abstract. We study the existence of pure Nash equilibria in weighted congestion games. Let $\mathcal{C}$ denote a set of cost functions. We say that $\mathcal{C}$ is *consistent* if every weighted congestion game with cost functions in $\mathcal{C}$ possesses a pure Nash equilibrium. Our main contribution is a complete characterization of consistency of continuous cost functions. We prove that a set $\mathcal{C}$ of continuous functions is consistent for two-player games if and only if $\mathcal{C}$ contains only monotonic functions and for all non-constant functions $c_1, c_2 \in \mathcal{C}$, there are constants $a, b \in \mathbb{R}$ such that $c_1(x) = a\, c_2(x) + b$ for all $x \in \mathbb{R}_{\geq 0}$. For games with at least three players, we prove that $\mathcal{C}$ is consistent if and only if exactly one of the following cases hold: (a) $\mathcal{C}$ contains only affine functions; (b) $\mathcal{C}$ contains only exponential functions such that $c(x) = a_c\, e^{\phi\, x} + b_c$ for some $a_c, b_c, \phi \in \mathbb{R}$, where a_c and b_c may depend on c, while ϕ must be equal for every $c \in \mathcal{C}$. The latter characterization is even valid for three-player games. Finally, we derive several characterizations of consistency of cost functions for games with restricted strategy spaces, such as weighted network congestion games or weighted congestion games with singleton strategies.

1. Introduction

In many situations, the state of a system is determined by a finite number of independent players, each optimizing an individual objective function. A natural framework for analyzing such decentralized systems are noncooperative games. While it is well known that for finite noncooperative games a Nash equilibrium in mixed strategies always exists, this need not be true for Nash equilibria in pure strategies (PNE for short). One of the fundamental goals in game theory is to characterize conditions under which a Nash equilibrium in pure strategies exists. In this paper, we study this question for weighted congestion games.

Congestion games, as introduced by Rosenthal [32], model the interaction of a finite set of players that compete over a finite set of facilities. A pure strategy of each player is a set of facilities. The cost of facility f is given by a real-valued cost function c_f that depends on the number of players using f and the private cost of every player equals the sum of the costs of the facilities in the strategy that she chooses. Rosenthal [32] proved in a seminal paper that such congestion games always admit a PNE by showing that these games posses an exact potential function. In a *weighted* congestion game, every player has a demand $d_i \in \mathbb{R}_{>0}$ that she places on the chosen facilities. The cost of a facility is then a function of the total load on the facility. An important subclass of weighted congestion games are weighted *network* congestion games. Here, every player is associated with a positive demand that she wants to route from her origin to her destination on a path of minimum cost. In contrast to unweighted congestion games, weighted congestion games do not always admit a PNE. Fotakis et al. [16] and Libman and Orda [24] each constructed a single-commodity network instance with two players having demands one and two, respectively, and showed that these games do not have a PNE. Their instances use different non-decreasing cost values per edge that are defined at the three possible loads $1, 2, 3$. Goemans et al. [19] constructed a two-player single-commodity instance without a PNE that uses different polynomial cost functions with nonnegative coefficients and degree of at most two. Interestingly, Anshelevich et al. [5] showed that for cost functions of the form $c_f(x) = \bar{c}_f/x$, $\bar{c}_f \in \mathbb{R}_{\geq 0}$, every two-player game possesses a PNE. For games with affine cost functions, Fotakis et al. [16, 17] proved that every weighted congestion game possesses a PNE. Later, Panagopoulou and Spirakis [30] proved that PNE always exist for instances with uniform exponential cost functions ($c_f(x) = e^x$). Harks et al. [22] generalized this existence result to non-uniform exponential cost functions of the form $c_f(x) = a_f\, e^{\phi\, x} + b_f$ for some $a_f, b_f, \phi \in \mathbb{R}$, where a_f and b_f may depend on the facility f, while ϕ must be equal for all facilities. It is worth noting that the positive results of [16, 17, 22, 30] are particularly important as they establish existence of PNE for the respective sets of cost functions *independent* of the underlying game structure, that is, *independent* of the underlying strategy set, demand vector, and number of players, respectively.

In this paper, we further explore the equilibrium existence problem in weighted congestion games. Our goal is to precisely characterize which types of cost functions actually guarantee the existence of PNE. To formally capture this issue, we introduce the notion of *PNE-consistency* or simply *consistency* of a set of cost functions. Let $\mathcal{C}$ be a set of cost functions and let $\mathcal{G}(\mathcal{C})$ be the set of *all* weighted congestion games with cost functions in $\mathcal{C}$. We say that $\mathcal{C}$ is *consistent* if every game in $\mathcal{G}(\mathcal{C})$ possesses a PNE. Using this terminology, the results of [16, 17, 22, 30] yield that $\mathcal{C}$ is consistent if $\mathcal{C}$ contains either affine functions or certain exponential functions. A natural open question is to decide whether there are further consistent functions, that is, functions guaranteeing the existence of a PNE. We thus investigate the following question: How large is the set $\mathcal{C}$ of consistent cost functions? We also introduce a stricter notion of consistency which we term *FIP-consistency*. Formally, we say that a set $\mathcal{C}$ of cost functions is FIP-consistent, if every game in $\mathcal{G}(\mathcal{C})$ possesses the Finite Improvement Property, that is, every sequence of unilateral improvements is finite, see Monderer and Shapley [28].

1.1. Our results

In order to obtain a complete characterization of the equilibrium existence problem in weighted congestion games, we first derive necessary conditions. Let $\mathcal{C}$ be a set of continuous functions. We show that if $\mathcal{C}$ is consistent, then $\mathcal{C}$ may only contain monotonic functions. We here use "monotonic" in the literal sense, i.e., every function $c \in \mathcal{C}$ is either non-decreasing or non-increasing. We further show that monotonicity of cost functions is necessary for consistency even in singleton games with only two players, two facilities, identical cost functions and symmetric strategies. As our first main result we show that a set of continuous cost functions $\mathcal{C}$ is consistent for two-player games if and only if $\mathcal{C}$ contains only monotonic functions and for all non-constant $c_1, c_2 \in \mathcal{C}$, there are constants $a, b \in \mathbb{R}$ such that $c_1(x) = a\, c_2(x) + b$ for all $x \in \mathbb{R}_{\geq 0}$. This characterization precisely explains the seeming dichotomy between the positive result of Anshelevich et al. [5] for two-player games and the two-player instances without PNE given by [16, 19, 24]. Our second main result establishes a characterization for the general case. We prove that a set $\mathcal{C}$ of continuous functions is consistent for games with at least three players if and only if exactly one of the following cases hold: (a) $\mathcal{C}$ contains only affine functions; (b) $\mathcal{C}$ contains only exponential functions such that $c(x) = a_c\, e^{\phi\, x} + b_c$ for some $a_c, b_c, \phi \in \mathbb{R}$, where a_c and b_c may depend on c, while ϕ must be equal for every $c \in \mathcal{C}$. This characterization is even valid for three-player games. We further show that for both two player games and games with at least three players consistency of $\mathcal{C}$ is equivalent to FIP-consistency.

While the above characterizations hold for *arbitrary* strategy spaces, we also study consistency of cost functions for *restricted* strategy spaces. First, we consider weighted network congestion games. Assuming strictly positive costs, we show that essentially all results translate to directed network congestion games. For games on undirected networks, we give respective characterizations for games with two players and at least four players leaving a gap for three-player games. For singleton weighted congestion games with two players we show that $\mathcal{C}$ is consistent if and only if $\mathcal{C}$ contains only monotonic functions. This characterization does not extend to games with three players. We give an instance with three players and monotonic cost functions without a PNE. For symmetric singleton weighted congestion games, however, we prove that $\mathcal{C}$ is consistent if and only if $\mathcal{C}$ contains only monotonic functions. In contrast to the characterizations for arbitrary strategy spaces, both characterizations do not carry over to FIP-consistency. We provide corresponding instances with improvement cycles.

1.2. Techniques and Outline of the Paper

The proofs of our main results essentially rely on two ingredients. First, we derive in Section 3 for continuous and consistent cost functions two necessary conditions (Monotonicity Lemma and Extended Monotonicity Lemma). The Monotonicity Lemma states that any continuous and consistent cost function must be monotonic. The lemma is proved by constructing a generic two-player weighted congestion game in which we identify a unique 4-cycle of deviations of two players. Then, we show that for any non-monotonic cost function, there is a weighted congestion game with a unique improvement cycle. By adding additional players and carefully choosing the players' weights and strategy spaces, we then derive the Extended Monotonicity Lemma, which ensures that the set of cost functions contained in a certain *finite integer linear hull* of the considered cost functions must be monotonic. By analyzing functions contained in the finite integer linear hull corresponding to two-player games, we prove in Section 4 that a set of continuous cost functions is consistent for two-player games if and only if all cost functions are monotone and every two non-constant cost functions are affine transformations of each other. In Section 5, we consider games with at least three players. We show that the Extended Monotonicity Lemma for games with at least three players implies that consistent and continuous cost functions must be either affine or exponential. In Section 6 and Section 7, we derive characterizations of consistency and FIP-consistency of cost functions for games with restricted strategy spaces, such as weighted network congestion games and weighted singleton congestion games, respectively.

1.3. Significance

Weighted congestion games are among the core topics in the game theory, operations research, computer science and economics literature. This class of games has several applications such as scheduling games, routing games, facility location games, network design games, etc. see [1, 5, 11, 18, 23, 27]. In all of the above applications there are two fundamental goals from a system design perspective: (i) the system must be *stabilizable*, that is, there must exist a stable point (PNE) from which no player wants to unilaterally deviate; (ii) myopic play of the players should guide the system to a *stable* state. Because the number of players and their types (expressed by the demands and the strategy spaces) are only known to the players and not available to the system designer, it is very natural to study the above two issues with respect to the used *cost functions*. In fact, in most of the above mentioned applications, the cost functions are under control of the system designer since they represent the technology associated with the resources, e.g., queuing discipline at routers, latency function in transportation networks, etc. Therefore, our results may help to predict and explain unstable traffic distributions in telecommunication networks and road networks. For instance in telecommunication networks, relevant cost functions are the so-called $M/M/1$-delay functions (see also [35]). These functions are of the form $c_a(x) = 1/(u_a - x)$, where u_a represents the capacity of arc a. In road networks, for instance, the most frequently used functions are monomials of degree 4 put forward by the U.S. Bureau of Public Roads [10]. Our results imply, that for these special types of cost functions, there is *always* a multi-commodity instance (with three players and identical cost functions) that is unstable in the sense that a PNE does not exist. On the other hand, our characterizations can be used to design a stable system: for instance, uniform $M/M/1$-delay functions are consistent for two-player games.

Our results are also relevant for the large body of work quantifying the worst-case efficiency loss of PNE for different sets of cost functions, see Awerbuch et al. [6], Christodoulou and Koutsoupias [12], and Aland et al. [3]. While mixed Nash equilibria are guaranteed to exist, their use is often unrealistic in practice. On the other hand, our work reveals that for most classes of cost functions pure Nash equilibria as the stronger solution concept may fail to exist in weighted congestion games. Thus, our work provides additional justification to study the worst-case efficiency loss for different solution concepts, such as sink equilibria [19], correlated and coarse correlated equilibria [9],[34].

1.4. Related Work

In contrast to ordinary congestion games as introduced by Rosenthal [32], games with weighted players and/or player-specific cost functions need not possess a PNE. As for weighted players, even two-players games may fail to admit a PNE, see the examples given by Fotakis et al. [16], Goemans et al. [19] and Libman and Orda [24]. Also related is the early work of Rosenthal [33] who showed that in weighted congestion games where players are allowed to split their demand integrally, a PNE need not exist. On the positive side, Fotakis et al. [16] and Panagopoulou and Spirakis [30] proved the existence of a PNE in games with affine and exponential costs, respectively. Dunkel and Schulz [13] showed that it is strongly NP-hard to decide whether or not a weighted congestion game with nonlinear cost functions possesses a PNE. If the strategy of every player contains a single facility only (singleton games), Fotakis et al. [15] showed the existence of PNE for linear cost functions (without a constant). Even-Dar et al. [14] derived the existence of PNE for load balancing games on parallel unrelated machines. Andelman et al. [4] proved even the existence of a strong Nash equilibrium - a strengthening of the pure Nash equilibrium to resilience against coalitional deviations - in scheduling games on unrelated machines. In fact, strong Nash equilibria exist in all singleton weighted congestion games with non-decreasing costs, see Harks et al. [21]. This also holds for non-increasing cost functions as proven by Rozenfeld and Tennenholtz [36]. Allowing for player-specific cost functions, Milchtaich [25] showed that unweighted singleton congestion games with player-specific cost functions possess at least one PNE. He also presented an instance with weighted players and player-specific cost functions without a PNE. Gairing et al. [18] showed that best response dynamics do not cycle if the player-specific cost functions are linear without a constant term. Milchtaich [27] further showed that general network games with player-specific costs need not admit a PNE in general. In fact, the corresponding decision problem turns out to be NP-complete, as shown by Ackermann and Skopalik [2]. Ieong et al. [23] proved that in congestion games with singleton strategies and non-decreasing cost functions, best response dynamics converge in polynomial time to a PNE. Ackermann et al. [1] extended this result to weighted congestion games with a so called *matroid property*, that is, the strategy of every player forms a basis of a matroid. In the same paper, they showed that the matroid property is the maximal property that gives rise to a PNE for all non-decreasing cost functions, that is, for any strategy space not satisfying the matroid property, there is an instance of a weighted congestion game not having a PNE. The consistency approach that we pursue in this paper is orthogonal to that of Ackermann et al. [1]. While they characterize the structure of the strategy space guaranteeing the existence of a PNE assuming arbitrary positive and non-decreasing costs, we characterize the structure of cost functions guaranteeing the existence of a PNE assuming arbitrary strategy spaces. Orda et al. [29] study the issue of uniqueness of PNE in weighted network congestion games with splittable demands (see also Fleischer et al. [8], Milchtaich [26], Richman and Shimkin [31] and Yang and Zhang [37]). They give sufficient conditions for uniqueness of PNE for several classes of cost functions. Interestingly, in the final section of their paper, the authors raise the issue about the existence of pure Nash equilibria in such games (depending on the cost functions) under the assumption that the flow is unsplittable. The results in this paper give a complete answer to their question.

An extended abstract of parts of this paper appeared in the Proceedings of the 37th International Colloquium on Automata, Languages and Programming, 2010, see [20].

2. Preliminaries

We consider finite strategic games $G = (N, S, \pi)$, where $N = \{1, \dots, n\}$ is the non-empty and finite set of players, $S = \bigtimes_{i \in N} S_i$ is the non-empty strategy space, and $\pi : S \to \mathbb{R}^n$ is the combined *private cost* function that assigns a private cost vector $\pi(s)$ to each strategy profile $s \in S$. We consider cost minimization games and (unless specified otherwise) we allow private cost functions to be negative or positive. We call an element $s \in S$ strategy profile. For $i \in N$, we write $S_{-i} = \bigtimes_{j \neq i} S_j$ and $s = (s_i, s_{-i})$ meaning that $s_i \in S_i$ and $s_{-i} \in S_{-i}$. A strategy profile s is a *pure Nash equilibrium (PNE)* if $\pi_i(s) \leq \pi_i(t_i, s_{-i})$ for all $i \in N$ and $t_i \in S_i$. A pair $\big(s, (t_i, s_{-i})\big) \in S \times S$ is called an *improving move* (or *profitable deviation*) of player i if $\pi_i(s_i, s_{-i}) > \pi_i(t_i, s_{-i})$. We call a sequence of strategy profiles $\gamma = (s^1, s^2, \dots)$ an *improvement path* if for every k the tuple (s^k, s^{k+1}) is an improving move for some player i. A closed path $(s^1, \dots, s^l, s^1)$ is referred to as an *l-improvement cycle*. A game has the *Finite Improvement Property (FIP)* if no such cycle exists. A function $P : S \to \mathbb{R}$ with $P(s) > P(t)$ for all improving moves (s, t) is called *potential function*. As noticed by Monderer and Shapley [28], every game that admits a potential function has the FIP and every finite game with the FIP possesses a PNE.

A tuple $\mathcal{M} = (N, F, S = \bigtimes_{i \in N} S_i, (c_f)_{f \in F})$ is called a *congestion model*, where N is the set of players, F is a non-empty, finite set of facilities and for each player $i \in N$, her collection of pure strategies S_i is a non-empty, finite set of subsets of F. A cost function $c_f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is associated with every facility $f \in F$. In contrast to most previous works, we do neither assume monotonicity nor positivity of costs. In the following, we define weighted congestion games similar to Goemans et al. [19].

Definition 2.1 (Weighted congestion game)**.** Let $\mathcal{M} = (N, F, S, (c_f)_{f \in F})$ be a congestion model and $(d_i)_{i \in N}$ be a vector of demands with $d_i \in \mathbb{R}_{>0}$. The corresponding *weighted congestion game* is the strategic game $G(\mathcal{M}) = (N, S, \pi)$, where π is defined as $\pi = \bigtimes_{i \in N} \pi_i$, $\pi_i(s) = \sum_{f \in s_i} d_i \, c_f\big(\ell_f(s)\big)$ and $\ell_f(s) = \sum_{j \in N : f \in s_j} d_j$.

We sometimes write G instead of $G(\mathcal{M})$. Let $\mathcal{C}$ be a set of cost functions and let $\mathcal{G}(\mathcal{C})$ be the set of all weighted congestion games with cost functions in $\mathcal{C}$. Then, we say that $\mathcal{C}$ is *consistent* if every $G \in \mathcal{G}(\mathcal{C})$ admits a PNE; we call $\mathcal{C}$ *FIP-consistent* if every $G \in \mathcal{G}(\mathcal{C})$ has the FIP. If the set $\mathcal{G}(\mathcal{C})$ is restricted, for instance to two player games etc., we say that $\mathcal{C}$ is *consistent for* $\mathcal{G}(\mathcal{C})$ if every $G \in \mathcal{G}(\mathcal{C})$ possesses a PNE.

3. Necessary Conditions on the Existence of a PNE

As a first result, we prove that if $\mathcal{C}$ is consistent, then every function $c \in \mathcal{C}$ is monotonic. We first need a technical lemma.

Lemma 3.1. Let $c : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a continuous function. Then, the following two statements are equivalent:

(1) c is monotonic on $\mathbb{R}_{\geq 0}$.
(2) The following two conditions hold:
   (a) For all $x > 0$ with $c(x) > c(0)$ there is $\epsilon > 0$ such that $c(y) \geq c(x)$ for all $y \in (x, x+\epsilon)$.
   (b) For all $x > 0$ with $c(x) < c(0)$ there is $\epsilon > 0$ such that $c(y) \leq c(x)$ for all $y \in (x, x+\epsilon)$.

Proof. *1* $\Rightarrow$ *2*: Trivial.

For proving *2* $\Rightarrow$ *1*, we first derive a useful property of functions satisfying *2*. Let $c : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a continuous function satisfying *2*. Moreover, assume that there is an open interval (α, ω) with $c(x) \neq c(0)$ for all $x \in (\alpha, \omega)$. We claim that c is non-decreasing on (α, ω) if $c(x) > c(0)$ for all $x \in (\alpha, \omega)$ and that c is non-increasing on (α, ω) if $c(x) < c(0)$ for all $x \in (\alpha, \omega)$. We prove only the first case because the second follows by the same arguments. Let $c(x) > c(0)$ for all $x \in (\alpha, \omega)$. For a contradiction, assume that there are $p_1, p_2 \in (\alpha, \omega)$ with $p_1 < p_2$ and $c(p_1) > c(p_2)$. We define $p_1' = \max\{x \in [p_1, p_2] : c(x) \geq c(p_1)\}$. Note that the set $\{x \in [p_1, p_2] : c(x) \geq c(p_1)\}$ is non-empty because it contains p_1 and closed because c is continuous. Using *2*, there is $\epsilon = \epsilon(p_1') > 0$ such that $c(y) \geq c(p_1') \geq c(p_1)$ for all $y \in (p_1', p_1' + \epsilon)$, contradicting the maximality of p_1'.

Now we prove *2* $\Rightarrow$ *1*. Let $\alpha = \inf\{x > 0 : c(x) \neq c(0)\}$. If $\alpha = \infty$, we are done as c is constant. Otherwise, we claim that $c(x) \neq c(0)$ for all $x > \alpha$. For a contradiction, let $\omega = \min\{x > \alpha : c(x) = c(0)\}$ and $\delta = c(\frac{\omega+\alpha}{2})$ (the minimum is attained because c is continuous). By construction, $c(x) \neq c(0)$ for all $x \in (\alpha, \omega)$. If $c(x) > c(0)$ for all $x \in (\alpha, \omega)$, we have $c(x) \geq \delta > c(0)$ for all $x \in (\frac{\omega+\alpha}{2}, \omega)$ and thus $c(0) = c(\omega) = \lim_{x \nearrow \omega} c(x) \geq \delta > c(0)$, a contradiction. If on the other hand $c(x) < c(0)$ for all $x \in (\alpha, \omega)$ we get $c(0) = c(\omega) = \lim_{x \nearrow \omega} c(x) \leq \delta < c(0)$ again a contradiction. We conclude that $c(x) \neq c(0)$ for all $x > \alpha$. Thus, for every $\omega > \alpha$, the function c is monotonic on the open interval (α, ω) and thus, c is monotonic on $\mathbb{R}_{\geq 0}$. $\square$

The following existence result for continuous, non-monotonic functions can be derived directly from Lemma 3.1 and will be very useful in the remainder of this paper.

Lemma 3.2. Let $c : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a continuous, non-monotonic function. Then, there are $x, y \in \mathbb{R}_{>0}$ with $y > x$ such that either $c(y-x) < c(y) < c(x)$ or $c(y-x) > c(y) > c(x)$.

Proof. Using the characterization of monotonic functions of Lemma 3.1, for every continuous non-monotonic function c, there is $x > 0$ such that one of the following holds: $c(x) > c(0)$ and for every $\epsilon > 0$ there is $y = y(\epsilon) \in (x, x+\epsilon)$ such that $c(y) < c(x)$; or $c(x) < c(0)$ and for every $\epsilon > 0$ there is $y = y(\epsilon) \in (x, x+\epsilon)$ such that $c(y) > c(x)$. Fix such x. Because of the continuity of c, we have $c(y-x) \to c(0)$ and $c(y) \to c(x)$ for $\epsilon \to 0$. For sufficiently small ϵ, x and $y(\epsilon)$ have the desired property. $\square$

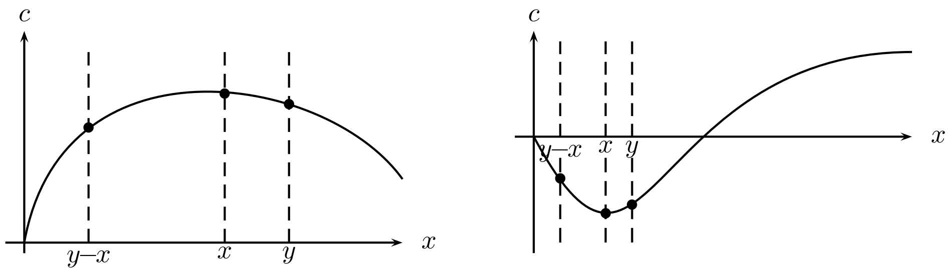


Figure 1. As shown in Lemma 3.2, for every continuous non-monotonic function there are $x, y \in \mathbb{R}_{>0}$ with $y > x$ such that one of the following cases holds: (a) $c(y-x) < c(y) < c(x)$; (b) $c(y-x) > c(y) > c(x)$.

The two cases occurring in Lemma 3.2 are depicted in Figure 1. Now consider a facility f with a non-monotonic cost function and two players with demands $d_1 = y - x$ and $d_2 = x$, where x and y are as in Lemma 3.2. Clearly, in case $c(y-x) < c(y) < c(x)$ player 1 prefers to be alone on f while player 2 would like to share the facility with player 1. If $c(y-x) > c(y) > c(x)$, the argumentation works the other way round. This observation is the key to construct a two-player weighted congestion game with singleton strategies that does not admit a PNE.

Lemma 3.3 (Monotonicity Lemma)**.** Let $\mathcal{C}$ be a set of continuous functions. If $\mathcal{C}$ is consistent, then every $c \in C$ is monotonic.

Proof. For a contradiction, suppose that $c \in \mathcal{C}$ is a non-monotonic function and consider the congestion model $\mathcal{M} = (N, F, S, (c_f)_{f\in F})$ with $N = \{1, 2\}$, $F = \{f, g\}$, $S_1 = S_2 = \big\{\{f\}, \{g\}\big\}$, $c_f = c_g = c$. Since c is non-monotonic, by Lemma 3.2 we can find $x, y \in \mathbb{R}_{>0}$ with $y > x$ such that either $c(y-x) < c(y) < c(x)$ or $c(y-x) > c(y) > c(x)$. Regard the game $G(\mathcal{M})$ with $d_1 = y - x$ and $d_2 = x$. Calculating the differences of the deviating players' private costs along the 4-cycle $\gamma = \big(\big(\{f\},\{f\}\big), \big(\{g\},\{f\}\big), \big(\{g\},\{g\}\big), \big(\{f\},\{g\}\big), \big(\{f\},\{f\}\big)\big)$, we obtain

$$\pi_1(\{g\},\{f\}) - \pi_1(\{f\},\{f\}) = (y-x)\big(c(y-x) - c(y)\big), \tag{1}$$
$$\pi_1(\{f\},\{g\}) - \pi_1(\{g\},\{g\}) = (y-x)\big(c(y-x) - c(x)\big), \tag{2}$$
$$\pi_2(\{g\},\{g\}) - \pi_2(\{g\},\{f\}) = x\big(c(y) - c(x)\big), \tag{3}$$
$$\pi_2(\{f\},\{f\}) - \pi_2(\{f\},\{g\}) = x\big(c(y) - c(x)\big). \tag{4}$$

If $c(y-x) < c(y) < c(x)$, the differences (1)-(4) are negative and γ is an improvement cycle. If $c(y-x) > c(y) > c(x)$, we can reverse the direction of γ and still get an improvement cycle. Using that every strategy combination is contained in γ, the claimed result follows. $\square$

Besides the continuity of the functions in $\mathcal{C}$, the proof of Lemma 3.3 relies on rather mild assumptions and, thus, this result can be strengthened in the following way.

Corollary 3.4. Let $\mathcal{C}$ be a set of continuous functions. Let $\mathcal{G}(\mathcal{C})$ be the set of weighted congestion games with cost functions in $\mathcal{C}$ satisfying one or more of the following properties: (i) Each game $G \in \mathcal{G}(\mathcal{C})$ has two players; (ii) Each game $G \in \mathcal{G}(\mathcal{C})$ has two facilities; (iii) For each game $G \in \mathcal{G}(\mathcal{C})$ and each player $i \in N$, the set of her strategies S_i contains a single facility only; (iv) Each game $G \in \mathcal{G}(\mathcal{C})$ has symmetric strategies; (v) Cost functions are identical, that is, $c_f = c_g$ for all $f, g \in F$. If $\mathcal{C}$ is consistent for $\mathcal{G}(\mathcal{C})$, then, each $c \in \mathcal{C}$ must be monotonic.

We now extend the Monotonicity Lemma to obtain an even stronger result by regarding more players and more complex strategies. To this end, for $K \in \mathbb{N}$ we consider those functions that can be written as the integral linear combination of K functions in $\mathcal{C}$, possibly with an offset.

Formally, we define the *finite integer linear hull* of $\mathcal{C}$ as

$$\mathcal{L}_{\mathbb{Z}}(\mathcal{C}) = \Big\{c : \mathbb{R}_{\geq 0} \to \mathbb{R} : c(x) = \sum_{k=1}^{K} a_k\, c_k(x + b_k) : K \in \mathbb{N}, a_k \in \mathbb{Z}, b_k \in \mathbb{R}_{\geq 0}, c_k \in \mathcal{C}\Big\}, \qquad (5)$$

and show that consistency of $\mathcal{C}$ implies that $\mathcal{L}_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

Lemma 3.5 (Extended Monotonicity Lemma)**.** Let $\mathcal{C}$ be a set of continuous functions. If $\mathcal{C}$ is consistent, then $\mathcal{L}_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

Proof. Assume by contradiction that $c \in \mathcal{L}_{\mathbb{Z}}(\mathcal{C})$ is not monotonic. By allowing $c_k = c_l$ for $k \neq l$, we can omit the integer coefficients a_k and rewrite c as $c(x) = \sum_{k=1}^{m_+} c_k(x+b_k) - \sum_{k=1}^{m_-} \bar{c}_k(x+\bar{b}_k)$ for some $c_k, \bar{c}_k \in \mathcal{C}, m_+, m_- \in \mathbb{N}$ and $b_k, \bar{b}_k \in \mathbb{R}_{\geq 0}$.

We define the congestion model $\mathcal{M} = (N, F, S, (c_f)_{f\in F})$, where $N = N_p \cup N^+ \cup N^-$ and $F = F^1 \cup F^2 \cup F^3 \cup F^4$. The set of players N^+ contains for each c_k, $1 \leq k \leq m_+$, a player with demand b_k and the set of players N^- contains for each $\bar{c}_k$, $1 \leq k \leq m_-$, a player with demand $\bar{b}_k$. We call the players in $N^- \cup N^+$ *offset* players. The set $N_p = \{1, 2\}$ contains two additional (non-trivial) players. Offset players with demand equal to 0 can be removed from the game. For ease of exposition, we assume that all offsets b_k, $k = 1 \ldots, m_+$ and $\bar{b}_k$, $k = 1, \ldots, m_-$ are strictly positive.

We now explain the strategy spaces and the sets F^1, F^2, F^3, F^4. For each function c_k, $1 \leq k \leq m_+$, we introduce two facilities f_k^2, f_k^3 with cost function c_k. For each function $\bar{c}_k$, $1 \leq k \leq m_-$, we introduce two facilities f_k^1, f_k^4 with cost function $\bar{c}_k$. To model the offsets b_k in (5), for each offset player $k \in N^+$, we define $S_k = \{f_k^2, f_k^3\}$. Similarly, for each offset player $k \in N^-$, we set $S_k = \{f_k^1, f_k^4\}$. The non-trivial players in N_p have strategies $S_1 = \{F^1 \cup F^2, F^3 \cup F^4\}$ and $S_2 = \{F^1 \cup F^3, F^2 \cup F^4\}$, where $F^1 = \{f_1^1, \ldots, f_{m_-}^1\}$, $F^2 = \{f_1^2, \ldots, f_{m_+}^2\}$, $F_3 = \{f_1^3, \ldots, f_{m_+}^3\}$, and $F^4 = \{f_1^4, \ldots, f_{m_-}^4\}$.

As c is assumed to be non-monotonic, by Lemma 3.2, there are $x, y \in \mathbb{R}_{>0}$ with $y > x$ such that either $c(y-x) < c(y) < c(x)$ or $c(y-x) > c(y) > c(x)$. We consider the weighted congestion game $G(\mathcal{M})$ with $d_1 = y - x$ and $d_2 = x$ for $1, 2 \in N_p$. For the 4-cycle

$$\gamma = \Big(\big(F^1 \cup F^2, F^1 \cup F^3, \ldots\big), \big(F^3 \cup F^4, F^1 \cup F^3, \ldots\big), \big(F^3 \cup F^4, F^2 \cup F^4, \ldots\big), \\ \big(F^1 \cup F^2, F^2 \cup F^4, \ldots\big), \big(F^1 \cup F^2, F^1 \cup F^3, \ldots\big)\Big),$$

it is straightforward to calculate that

$$\begin{aligned}
&\pi_1(F^3 \cup F^4, F^1 \cup F^3, \ldots) - \pi_1(F^1 \cup F^2, F^1 \cup F^3, \ldots) \\
&\quad= (y-x)\left(\sum_{k=1}^{m_+} c_k(d_1+d_2+b_k) - \sum_{k=1}^{m_-} \bar{c}_k(d_1+d_2+\bar{b}_k) + \sum_{k=1}^{m_-} \bar{c}_k(d_1+\bar{b}_k) - \sum_{k=1}^{m_+} c_k(d_1+b_k)\right) \\
&\quad= (y-x)\big(c(y) - c(y-x)\big),
\end{aligned}$$

$$\begin{aligned}
&\pi_2(F^3 \cup F^4, F^2 \cup F^4, \ldots) - \pi_2(F^3 \cup F^4, F^1 \cup F^3, \ldots) \\
&\quad= x\left(-\sum_{k=1}^{m_+} c_k(d_1+d_2+b_k) + \sum_{k=1}^{m_-} \bar{c}_k(d_1+d_2+\bar{b}_k) + \sum_{k=1}^{m_+} c_k(d_2+b_k) - \sum_{k=1}^{m_-} \bar{c}_k(d_2+\bar{b}_k)\right) \\
&\quad= x\big(c(x) - c(y)\big),
\end{aligned}$$

$$\begin{aligned}
&\pi_1(F^1 \cup F^2, F^2 \cup F^4, \dots) - \pi_1(F^3 \cup F^4, F^2 \cup F^4, \dots) \\
&\quad = (y-x)\left(\sum_{k=1}^{m_+} c_k(d_1+d_2+b_k) - \sum_{k=1}^{m_-} \bar{c}_k(d_1+d_2+\bar{b}_k) + \sum_{k=1}^{m_-} \bar{c}_k(d_1+\bar{b}_k) - \sum_{k=1}^{m_+} c_k(d_1+b_k)\right) \\
&\quad = (y-x)\big(c(y) - c(y-x)\big),
\end{aligned}$$

$$\begin{aligned}
&\pi_2(F^1 \cup F^2, F^1 \cup F^3, \dots) - \pi_2(F^1 \cup F^2, F^2 \cup F^4, \dots) \\
&\quad = x\left(-\sum_{k=1}^{m_+} c_k(d_1+d_2+b_k) + \sum_{k=1}^{m_-} \bar{c}_k(d_1+d_2+\bar{b}_k) - \sum_{k=1}^{m_-} \bar{c}_k(d_2+\bar{b}_k) + \sum_{k=1}^{m_+} c_k(d_2+b_k)\right) \\
&\quad = x\left(c(x) - c(y)\right).
\end{aligned}$$

If $c(y-x) > c(y) > c(x)$, all private cost differences are negative and γ is an improvement cycle; if on the other hand $c(y-x) < c(y) < c(x)$, the 4-cycle in the other direction is an improvement cycle. Because every strategy combination is contained in γ we get the claimed result. □

4. A Characterization for Two-Player Games

We analyze implications of the Extended Monotonicity Lemma (Lemma 3.5) for two-player weighted congestion games. First, for ease of exposition, we restrict ourselves to the case $K = 2$, that is, we only regard those functions that can be written as an integral linear combination of two functions in $\mathcal{C}$ without offset. We define the following set of functions

$$\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C}) = \{c : \mathbb{R}_{\geq 0} \to \mathbb{R} : c(x) = a_1\, c_1(x) + a_2\, c_2(x), a_1, a_2 \in \mathbb{Z}, c_1, c_2 \in \mathcal{C}\} \subseteq \mathcal{L}_{\mathbb{Z}}(\mathcal{C}).$$

We remark that by setting all offsets b_k in (5) equal to zero, the construction in the proof of Lemma 3.5 only involves two players. Thus, we immediately obtain the following result.

Proposition 4.1. Let $\mathcal{C}$ be a set of continuous functions. If $\mathcal{C}$ is consistent for two-player games, then, $\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

We proceed investigating sets of functions $\mathcal{C}$ that guarantee that $\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

Lemma 4.2. Let $c_1, c_2 : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be two continuous, monotonic and non-constant functions. Then, the following are equivalent.

(1) For all $a_1, a_2 \in \mathbb{Z}$ the function $a_1\, c_1 + a_2\, c_2$ is monotonic on $\mathbb{R}_{\geq 0}$.
(2) There are $a, b \in \mathbb{R}$ such that $c_2(x) = a\, c_1(x) + b$ for all $x \geq 0$.

Proof. 2 ⇒ 1: Calculus.

1 ⇒ 2: Without loss of generality, we may assume that c_1 and c_2 are non-decreasing because multiplying functions with -1 has no impact on either statement of the lemma. As c_1 is non-constant and non-decreasing, there is $x_1 \geq 0$ with $c_1(x_1) = c_1(0)$ and $c_1(x) > c_1(0)$ for all $x > x_1$. Fix such $x > x_1$. For all $a_1, a_2 \in \mathbb{Z}$, the function $a_1\, c_1 + a_2\, c_2$ is monotonic. This implies that for every $y > x_1$ and every $\alpha \in \mathbb{Q}$ the expressions $\alpha\, c_1(x) + c_2(x) - \alpha\, c_1(0) - c_2(0)$ and $\alpha\, c_1(y) + c_2(y) - \alpha\, c_1(0) - c_2(0)$ have identical signs. Thus, for all $y > x_1$ and all $\alpha \in \mathbb{Q}$ at least one of the following two cases holds

(1) $\alpha \geq -\frac{c_2(x)-c_2(0)}{c_1(x)-c_1(0)}$ and $\alpha \geq -\frac{c_2(y)-c_2(0)}{c_1(y)-c_1(0)}$
(2) $\alpha \leq -\frac{c_2(x)-c_2(0)}{c_1(x)-c_1(0)}$ and $\alpha \leq -\frac{c_2(y)-c_2(0)}{c_1(y)-c_1(0)}$

This clearly implies

$$\frac{c_2(y) - c_2(0)}{c_1(y) - c_1(0)} = \frac{c_2(x) - c_2(0)}{c_1(x) - c_1(0)} \tag{6}$$

for all $y > x_1$, because otherwise any rational

$$\alpha \in \Big(\min\{-\frac{c_2(y) - c_2(0)}{c_1(y) - c_1(0)}, -\frac{c_2(x) - c_2(0)}{c_1(x) - c_1(0)}\}, \max\{-\frac{c_2(y) - c_2(0)}{c_1(y) - c_1(0)}, -\frac{c_2(x) - c_2(0)}{c_1(x) - c_1(0)}\}\Big)$$

would violate both constraints. From (6), we obtain

$$c_2(y) = \frac{c_2(x) - c_2(0)}{c_1(x) - c_1(0)} \cdot c_1(y) - \frac{c_2(x) - c_2(0)}{c_1(x) - c_1(0)} \cdot c_1(0) + c_2(0)$$

for all $y > x_1$. This shows the existence of $a, b \in \mathbb{R}$ with $c_2(x) = a\, c_1(x) + b$ for all $x > x_1$. Exchanging the roles of c_1 and c_2, we may also derive the existence of $a', b' \in \mathbb{R}$ such that $c_1(x) = a'\, c_2(x) + b'$ for all $x > x_2$, where x_2 is such that $c_2(x_2) = c_2(0)$ and $c_2(x) > c_2(0)$ for all $x > x_2$. This implies $x_1 = x_2$. Using the fact that c_1 and c_2 are continuous and constant on $[0, x_1]$ finishes the proof. □

We are now ready to prove our first main result.

Theorem 4.3. Let $\mathcal{C}$ be a set of continuous functions. Let $\mathcal{G}^2(\mathcal{C})$ be the set of two-player games such that cost functions are in $\mathcal{C}$. Then, the following conditions are equivalent.

(1) $\mathcal{C}$ is consistent for $\mathcal{G}^2(\mathcal{C})$.
(2) $\mathcal{C}$ is FIP-consistent for $\mathcal{G}^2(\mathcal{C})$.
(3) $\mathcal{C}$ contains only monotonic functions and for all non-constant $c_1, c_2 \in \mathcal{C}$, there are constants $a, b \in \mathbb{R}$ such that $c_1(x) = a\, c_2(x) + b$ for all $x \geq 0$.

Proof. *2* ⇒ *1* is trivial.

1 ⇒ *3*: Using Proposition 4.1 we get that $\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions. As $\mathcal{C} \subseteq \mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$, this implies in particular that $\mathcal{C}$ contains only monotonic functions. For all non-constant functions $c_1, c_2 \in \mathcal{C}$ and all $a_1, a_2 \in \mathbb{Z}$ the function $a_1\, c_1 + a_2\, c_2 \in \mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ is monotonic. Applying Lemma 4.2 then yields the result.

3 ⇒ *2*: Let $\mathcal{C}$ be as specified in *3*. Trivially, the claimed result holds if $\mathcal{C}$ contains only constant functions. If $\mathcal{C}$ contains a non-constant function c consider the set $\bar{\mathcal{C}} = \{a\, c(x) + b : a, b \in \mathbb{R}\} \supseteq \mathcal{C}$. We show that $\bar{\mathcal{C}}$ is consistent for $\mathcal{G}^2(\bar{\mathcal{C}})$. To this end, consider an arbitrary two-player game with costs in $\bar{\mathcal{C}}$ and demands $d_1 < d_2$. We distinguish the following three cases.

First case: $c(d_1) < c(d_2) < c(d_1 + d_2)$, or $c(d_1) > c(d_2) > c(d_1 + d_2)$. Since c is strictly monotonic with respect to the points d_1, d_2 and $d_1 + d_2$, there is a strictly monotonic function $\tilde{c}$ with $\tilde{c}(d_1) = c(d_1)$, $\tilde{c}(d_2) = c(d_2)$ and $\tilde{c}(d_1 + d_2) = c(d_1 + d_2)$. Consequently, we can replace every cost function $c \in \bar{\mathcal{C}} = \{a\, c(x) + b : a, b \in \mathbb{R}\}$ by a cost function $\tilde{c} \in \tilde{\mathcal{C}} = \{a\, \tilde{c}(x) + b : a, b \in \mathbb{R}\}$ without changing the players' private costs. As shown by Harks et al. [22], for any strictly monotonic function $\tilde{c}$, every weighted congestion game G with two players and cost functions in $\bar{\mathcal{C}} = \{a\, \tilde{c}(x) + b : a, b \in \mathbb{R}\}$ admits a potential function and, thus, has the FIP.

Second case: $c(d_1) = c(d_2)$. We set $\tilde{d}_1 = \tilde{d}_2 = 1$ and chose for every facility $f \in F$ a new cost function $\tilde{c}_f$ with $\tilde{c}_f(1) = c_f(d_1) = c_f(d_2)$ and $\tilde{c}_f(2) = c_f(d_1 + d_2)$. By construction, the unweighted congestion game with demands $\tilde{d}_1$, $\tilde{d}_2$ and costs $(\tilde{c}_f)_{f \in F}$ has the same private costs as the original game. Rosenthal [32] showed the existence of a potential function in all unweighted congestion games, thus, the game has the FIP.

Third case: $c(d_2) = c(d_1 + d_2)$. We have $\bar{c}(d_2) = \bar{c}(d_1 + d_2)$ for all $\bar{c} \in \bar{\mathcal{C}}$ and thus player 2 is always indifferent whether player 1 shares a facility with her or not. For the FIP and the existence of a PNE, we argue as follows: Consider the vector valued function $\phi : S \to \mathbb{R}$, $s \mapsto (\pi_2(s), \pi_1(s))$ which assigns to every strategy profile the vector which has the private cost of players 2 and 1 in first and second component respectively. We claim that ϕ decreases lexicographically along any improvement path. This is trivial for improvement moves of player 2. Since player 2 is indifferent whether player 1 shares with her a facility or not, every improvement move of player 1 does not affect the private cost of player 2 but decreases the private cost of player 1. This implies that the lexicographical order of $\phi(s)$ decreases along any improvement path, thus, every such path is finite. $\square$

5. A Characterization for the General Case

We now consider the case $n \geq 3$, that is, we consider weighted congestion games with at least three players. We will show that a set of continuous cost functions is consistent if and only if this set contains either only linear or only certain exponential functions. Our main tool for proving this result is to analyze implications of the Extended Monotonicity Lemma (Lemma 3.5) for three-player weighted congestion games. Formally, define

$$\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C}) = \{c : \mathbb{R}_{\geq 0} \to \mathbb{R} : c(x) = a_1\, c_1(x) + a_2\, c_1(x+\delta) : a_1, a_2 \in \mathbb{Z}, c_1 \in \mathcal{C}, \delta \in \mathbb{R}_{>0}\} \subseteq \mathcal{L}_{\mathbb{Z}}(\mathcal{C}).$$

Note that $\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C})$ involves a single offset $\delta > 0$, which requires only three players in the construction of the proof of the Extended Monotonicity Lemma. However, regarding three-player games in which the strategy available to the third player does not intersect with the strategies of the first two players we still get as a necessary condition that $\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ may only contain monotonic functions. We, thus, obtain the following result.

Proposition 5.1. Let $\mathcal{C}$ be a set of continuous functions. If $\mathcal{C}$ is consistent for three-player games, then both $\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ and $\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C})$ contain only monotonic functions.

We proceed characterizing the set of cost functions $\mathcal{C}$ for which $\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

Lemma 5.2. Let $\mathcal{C}$ be a set of continuous functions. Then, the following two are equivalent:

(1) $\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

(2) For every $c \in \mathcal{C}$ either $c(x) = a\, e^{\phi\, x} + b$ for some $a, b, \phi \in \mathbb{R}$, or $c(x) = a\, x + b$ for some $a, b \in \mathbb{R}$.

Proof. *2* $\Rightarrow$ *1*: Let $c \in \mathcal{C}$ be an exponential or an affine function. By simple calculus one can verify that every function $\tilde{c}(x) = a_1\, c(x) + a_2\, c(x+\delta)$ with $a_1, a_2 \in \mathbb{Z}$, $\delta \in \mathbb{R}_{>0}$ is exponential if c is exponential and affine if c is affine.

1 $\Rightarrow$ *2*: By contradiction, assume that $c \in \mathcal{C}$ is neither affine nor exponential. As $\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions, for all $\delta > 0$ and all $a_1, a_2 \in \mathbb{Z}$ the function $\tilde{c} : \mathbb{R}_{\geq 0} \to \mathbb{R}, x \mapsto a_1\, c(x) + a_2\, c(x+\delta)$ is monotonic. Referring to Lemma 4.2, this implies that for all $\delta > 0$ there are $a, b \in \mathbb{R}$ such that for all $x \geq 0$

$$c(x+\delta) = a\, c(x) + b. \tag{7}$$

As $c \in \mathcal{C}$ is neither affine nor exponential on $\mathbb{R}_{\geq 0}$, there must exist four points $0 \leq p_1 < p_2 < p_3 < p_4$ following neither an exponential nor an affine law, i.e. there are neither $\alpha, \beta, \phi \in \mathbb{R}$ such that $c(p_i) = \alpha\, e^{\phi p_i} + \beta$ for all $i \in \{1,2,3,4\}$ nor are there $\alpha, \beta \in \mathbb{R}$ such that $c(p_i) = \alpha\, p_i + \beta$ for all $i \in \{1,2,3,4\}$. As c is continuous, we may assume without loss of generality that p_1, p_2, p_3, p_4 are rational and we write them as $p_1 = 2m_1/(2k), \ldots, p_4 = 2m_4/(2k)$ for some $m_1, m_2, m_3, m_4, k \in \mathbb{N}$. For $\delta = 1/k$ we derive from (7) that there are $a, b \in \mathbb{R}$ such that for all $n \in \mathbb{N}$

$$c\big((n+1)/k\big) = a\, c\big(n/k\big) + b, \tag{8}$$

$$c\big((n+2)/k\big) = a\, c\big((n+1)/k\big) + b. \tag{9}$$

Subtracting (8) from (9) and rearranging terms, we obtain for all $n \in \mathbb{N}$

$$c\big((n+2)/k\big) - (a+1)\,c\big((n+1)/k\big) + a\,c\big(n/k\big)\big) = 0. \tag{10}$$

This defines a second order linear homogeneous recurrence relation on the sequence $c\big(n/k\big)_{n\in\mathbb{N}}$. Such recurrence relations can be solved with the method of characteristic equations, see [7, §3.2] for more details. The characteristic equation of the recurrence relation equals $x^2-(a+1)\,x+a = (x-1)(x-a)$. If $a \neq 1$, then the characteristic equation has two distinct roots and we obtain for even m that

$$c\big(m/k\big) = \beta \cdot 1^m + \alpha \cdot a^m = \beta + \alpha \cdot |a|^m = \alpha \cdot \exp(m\,\ln|a|) + \beta$$

for some constants $\alpha, \beta \in \mathbb{R}$. If on the other hand $a = 1$, we can evaluate $c\big(m/k\big)$ as

$$c\big(m/k\big) = \beta \cdot 1^m + \alpha m \cdot 1^m = \alpha \cdot m + \beta$$

for some constants $\alpha, \beta \in \mathbb{R}$. □

We are now ready to state our second main theorem.

Theorem 5.3. Let $\mathcal{C}$ be a set of continuous functions. Then, the following three are equivalent:

(1) $\mathcal{C}$ is consistent.
(2) $\mathcal{C}$ is FIP-consistent.
(3) $\mathcal{C}$ contains only affine functions *or* $\mathcal{C}$ contains only functions of type $c(x) = a_c\,e^{\phi\,x} + b_c$ where $a_c, b_c \in \mathbb{R}$ may depend on c while $\phi \in \mathbb{R}$ is independent of c.

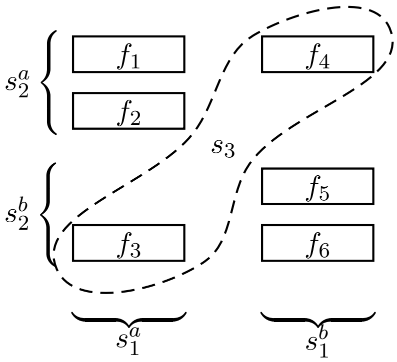
Proof. *2* $\Rightarrow$ *1* is trivial.

3 $\Rightarrow$ *2* follows because every weighted congestion games with such cost functions possesses a weighted potential, see [16, 22, 30].

1 $\Rightarrow$ *3*: By Proposition 5.1 both $\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ and $\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C})$ may only contain monotonic functions. Applying Lemma 5.2 we obtain that every $c \in \mathcal{C}$ is either affine or exponential. In addition, as shown in Lemma 4.2 for each two non-constant functions $c_1, c_2 \in \mathcal{C}$ there are $a, b \in \mathbb{R}$ such that $c_2(x) = a\,c_1(x) + b$ for all $x \geq 0$. Both results together imply 3. □

We conclude this section by giving an example that illustrates the main ideas presented so far. Recall, that Theorem 5.3 establishes that for each continuous, non-affine and non-exponential cost function c, there is a weighted congestion game G with uniform cost function c on all facilities that does not admit a PNE. In the following example, we show how such game for $c(x) = x^3$ is constructed.

Example 5.4. As the function $c(x) = x^3$ is neither affine nor exponential, there are $a_1, a_2 \in \mathbb{Z}$ and $\delta \in \mathbb{R}_{>0}$ such that $\tilde{c}(x) = a_1\,c(x) + a_2\,c(x+\delta)$ has a strict local extremum. In fact, we can choose $a_1 = 2, a_2 = -1$ and $\delta = 1$, that is, the function $\tilde{c}(x) = 2c(x) - c(x+1) = 2x^3 - (x+1)^3$ has a strict local minimum at $x_0 = 1 + \sqrt{2}$. In particular, we can choose $d_1 = 1$ and $d_2 = 2$ such that $\tilde{c}(d_1) = -6 > \tilde{c}(d_2) = -11 < \tilde{c}(d_1 + d_2) = -10$. The weighted congestion game without PNE is now constructed as follows: We introduce $2(|a_1| + |a_2|)$ facilities $f_1, \ldots, f_6$ and the following strategies $s_1^a = \{f_1, f_2, f_3\}$, $s_1^b = \{f_4, f_5, f_6\}$, $s_2^a = \{f_1, f_2, f_4\}$, $s_2^b = \{f_3, f_5, f_6\}$, and $s_3 = \{f_3, f_4\}$. We then set $S_1 = \{s_1^a, s_1^b\}$, $S_2 = \{s_2^a, s_2^b\}$, and $S_3 = \{s_3\}$, see Figure 2 for an illustration of the strategies. The so defined game has four strategy profiles, namely (s_1^a, s_2^a, s_3), (s_1^a, s_2^b, s_3), (s_1^b, s_2^a, s_3), (s_1^b, s_2^b, s_3). As Player 3 is an offset player, she has a single strategy only, thus, the players' private costs depend only on the choice of players 1 and 2. We derive that the 4-cycle γ shown in Figure 2 is a best-reply cycle in G. As there are no strategy profiles outside γ we conclude that G has no PNE.



$$\begin{tikzcd}
\pi(x_1^a, x_2^a, x_3) = (62, 2\cdot 81, 35) \arrow[d] & \pi(x_1^b, x_2^a, x_3) = (66, 2\cdot 80, 65) \arrow[l] \\
\pi(x_1^a, x_2^b, x_3) = (66, 2\cdot 80, 65) \arrow[r] & \pi(x_1^b, x_2^b, x_3) = (62, 2\cdot 81, 35) \arrow[u]
\end{tikzcd}$$

γ

Figure 2. (a) The players' strategies and (b) the improvement cycle γ of the game constructed in Example 5.4 that does not admit a PNE.

6. Weighted Network Congestion Games

In this section, we discuss the implications of our characterizations to the important subclass of weighted network congestion games. In these games, the facilities correspond to edges of a directed or undirected graph. Every player is associated with a positive demand that she wants to route from her origin to her destination on a path of minimum cost. We consider directed and undirected networks separately, starting with directed networks.

6.1. Directed Networks

We first give a version of the Extended Monotonicity Lemma for directed networks with two players and strictly positive costs.

Lemma 6.1 (Extended Monotonicity Lemma for Two-Player Games on Directed Networks)**.** Let $\mathcal{C}$ be a set of strictly positive and continuous functions. If $\mathcal{C}$ is consistent for two-player directed network congestion games, then $\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

Proof. Because singleton congestion games are a subclass of directed network congestion games by Corollary 3.4 every set $\mathcal{C}$ of consistent functions contains only monotonic functions. For a contradiction, assume that there are $a_1, a_2 \in \mathbb{Z}$ and monotonic functions $c_1, c_2 \in \mathcal{C}$ such that the function $c : \mathbb{R}_{\geq 0} \to \mathbb{R}$ defined as $c(x) = a_1 c_1(x) + a_2 c_2(x)$ is not monotonic. By Lemma 3.2 there are $x, y \in \mathbb{R}_{>0}$ with $y > x$ such that either $c(y-x) < c(y) < c(x)$ or $c(y-x) > c(y) > c(x)$. We choose the demands equal to $d_1 = y - x$ and $d_2 = x$. Note that c is monotonic if and only if $-c$ is monotonic, thus we may assume w.l.o.g. that $a_2 > 0$. In order to define the players' strategies we distinguish the following two cases.

First case: $a_1 < 0$: We use a construction similar to the proof of Lemma 3.5. To define the players' strategy spaces, consider the left network of Figure 3. The two players are represented by the two source-terminal pairs $(s_i, t_i), i = 1, 2$. The set of strategies available to player i equals the set of directed (s_i, t_i)-paths. The dashed edges in Figure 3 correspond to directed

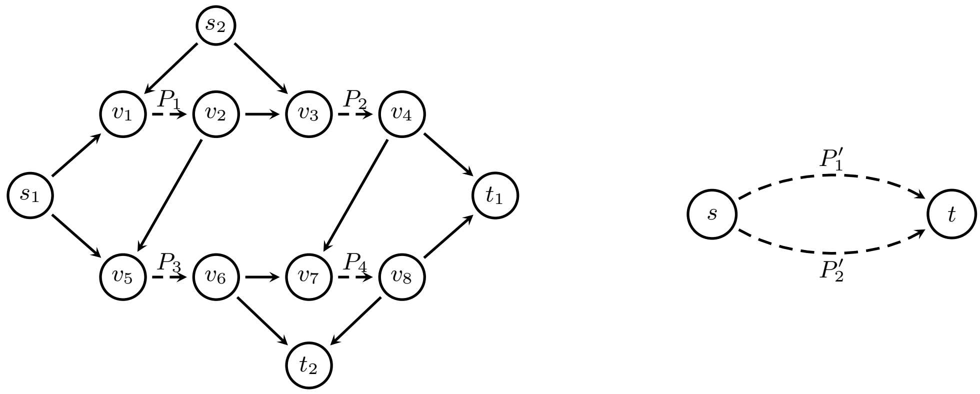


Figure 3. Directed network congestion games used in the proof of the Extended Monotonicity Lemma for Two-Player Directed Networks (Lemma 6.1).

paths $P_1, \dots, P_4$, which we choose as follows: the directed path P_1 from v_1 to v_2 contains $|a_1|$ edges with cost function c_1, the directed path P_2 from v_3 to v_4 contains a_2 edges with cost function c_2, the directed path P_3 from v_5 to v_6 contains a_2 edges with cost function c_2, and the directed path P_4 from v_7 to v_8 contains $|a_1|$ edges with cost function c_1. All other edges have an arbitrary cost function in $\mathcal{C}$, say c_1. Because all costs are strictly positive, for player 1 all strategies except the upper path $P_u = \{s_1 \to v_1, P_1, v_2 \to v_3, P_2, v_4 \to t_1\}$ and the lower path $P_d = \{s_1 \to v_5, P_3, v_6 \to v_7, P_4, v_8 \to t_1\}$ are strictly dominated in the sense that they have strictly higher costs than either P_u or P_d regardless of the strategy played by player 2. For player 2, all strategies except the left path $P_l = \{s_2 \to v_1, P_1, v_2 \to v_5, P_3, v_6 \to t_2\}$ and the right path $P_r = \{s_2 \to v_3, P_2, v_4 \to v_7, P_4, v_8 \to t_2\}$ are strictly dominated. We consider the 4-cycle $\gamma = \big((P_u, P_l), (P_d, P_l), (P_d, P_r), (P_u, P_r), (P_u, P_l)\big)$, and calculate that

$$\begin{aligned}\pi_1(P_d, P_l) - \pi_1(P_u, P_l) &= (y-x)\big(c_1(y-x) + a_2c_2(y) + c_1(y-x) - a_1c_1(y-x) + c_1(y-x) \\ &\qquad - c_1(y-x) + a_1c_1(y) - c_1(y-x) - a_2c_2(y-x) - c_1(y-x)\big) \\ &= (y-x)\big(a_1c_1(y) + a_2c_2(y) - a_1c_1(y-x) - a_2c_2(y-x)\big) \\ &= (y-x)\big(c(y) - c(y-x)\big)\end{aligned}$$

In the same fashion, we obtain $\pi_2(P_d, P_r) - \pi_2(P_d, P_l) = x\big(c(x) - c(y)\big)$, $\pi_1(P_u, P_r) - \pi_1(P_d, P_r) = (y-x)\big(c(y) - c(y-x)\big)$, and $\pi_2(P_u, P_l) - \pi_2(P_u, P_r) = x\big(c(x) - c(y)\big)$. If $c(y-x) > c(y) > c(x)$, then γ is an improvement cycle which gives that none of the strategy profiles contained in γ is a PNE. If on the other hand $c(y-x) < c(y) < c(x)$, we can reverse the direction of γ and get an improvement cycle. Because every strategy profile that uses only non-dominated strategies is contained in γ, the constructed directed network congestion game does not admit a PNE.

Second case: $a_1 > 0$: Consider the right network shown in Figure 3. Here, both players want to route from s to t, that is, $S_1 = S_2 = \{P_1', P_2'\}$. The directed paths P_1' and P_2' each contain a_1 edges with cost function c_1 and a_2 edges with cost function c_2. If $c(y-x) < c(y) < c(x)$, player 1 prefers to be alone on an (s,t)-path while player 2 wants to share the path with player 1. If $c(y-x) > c(y) > c(x)$, the argumentation works the other way round. We conclude that the game does not admit a PNE. □

Together with Lemma 4.2 and Theorem 4.3, we obtain the following characterization of consistency for two-player network congestion games on directed networks.

Theorem 6.2. Let $\mathcal{C}$ be a set of strictly positive and continuous functions and let $\mathcal{G}^2_{\mathrm{dn}}(\mathcal{C})$ be the set of two-player directed network games such that cost functions are in $\mathcal{C}$. Then, the following conditions are equivalent.

(1) $\mathcal{C}$ is consistent for $\mathcal{G}^2_{\mathrm{dn}}(\mathcal{C})$.
(2) $\mathcal{C}$ is FIP-consistent for $\mathcal{G}^2_{\mathrm{dn}}(\mathcal{C})$.
(3) $\mathcal{C}$ contains only monotonic functions and for all non-constant $c_1, c_2 \in \mathcal{C}$, there are constants $a, b \in \mathbb{R}$ with $c_1(x) = a\, c_2(x) + b$ for all $x \in \mathbb{R}_{\geq 0}$.

Using similar ideas as in the case of two players, we can also prove a version of the Extended Monotonicity Lemma for directed network games with three or more players.

Lemma 6.3 (Extended Monotonicity Lemma for Directed Networks)**.** Let $\mathcal{C}$ be a set of strictly positive and continuous functions. If $\mathcal{C}$ is consistent for three-player directed network congestion games, then $\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

Proof. Assume by contradiction that there are $a_1, a_2 \in \mathbb{Z}$, $\delta \in \mathbb{R}_{>0}$ and a monotonic function $c_1 \in \mathcal{C}$ such that the function $c : \mathbb{R}_{\geq 0} \to \mathbb{R}$ defined as $c(x) = a_1 c_1(x) + a_2 c_1(x+\delta)$ is not monotonic. We may again assume w.l.o.g. that c_1 is monotonic and that $a_2 > 0$. Note that because c_1 is monotonic, this implies $a_1 < 0$.

Consider the network in Figure 4 where again the directed paths P_1 and P_4 contain $|a_1|$ edges each, and the the directed paths P_2 and P_3 contains a_2 edges each. In addition to the players $i = 1, 2$ corresponding to the pairs $(s_i, t_i), i = 1, 2$ we now have a third player corresponding to the pair (s_3, t_3) with a single strategy $P_Q = \{P_3, Q, P_2\}$ and demand $d_3 = \delta$. Moreover, we set $d_1 = y - x$, $d_2 = x$ for $x, y \in \mathbb{R}_{>0}$ with $y > x$ such that either $c(y-x) < c(y) < c(x)$ or $c(y-x) > c(y) > c(x)$ holds (by Lemma 3.2 such values exist). We design the directed path Q from v_6 to v_3 so as to contain a sufficiently large number of edges, such that for players 1 and 2 all (s_i, t_i)-paths not containing Q are strictly less costly than every path that contains Q. As every (s_i, t_i)-path that does not contain Q has costs less than $2(a_2 - a_1 + 6)\, c_1(y+\delta)$ and every edge in Q has cost at least $c_1(\delta)$, it is sufficient to let Q contain $2(a_2 - a_1 + 6)\lceil \frac{c_1(y+\delta)}{c_1(\delta)} \rceil + 1$ edges. By construction of Q, for player 1, all strategies except the upper path $P_u = \{s_1 \to v_1, P_1, v_2 \to v_3, P_2, t_3 \to t_1\}$ and the lower path $P_d = \{s_1 \to s_3, P_3, v_6 \to v_7, P_4, v_8 \to t_1\}$ are strictly dominated in the sense that they have strictly higher costs than either P_u or P_d regardless of the strategies played by players 2 and 3. For player 2, all strategies except the left path $P_l = \{s_2 \to v_1, P_1, v_2 \to s_3, P_3, v_6 \to t_2\}$ and the right path $P_r = \{s_2 \to v_3, P_2, t_3 \to v_7, P_4, v_8 \to t_2\}$ are strictly dominated. With the same calculations as in Lemma 6.1 one can show that the 4-cycle $\gamma = \big((P_u, P_l, P_Q), (P_d, P_l, P_Q), (P_d, P_r, P_Q), (P_u, P_r, P_Q), (P_u, P_l, P_Q)\big)$ is an improvement cycle when traversed in the right direction. Because every strategy profile that uses only non-dominated strategies is contained in γ, we conclude that the thus constructed network congestion game does not admit a PNE. ☐

Using Lemma 5.2, we obtain the following characterization of cost functions that are consistent for weighted directed network congestion games.

Theorem 6.4. Let $\mathcal{C}$ be a set of strictly positive and continuous functions and let $\mathcal{G}_{\mathrm{dn}}(\mathcal{C})$ be the set of directed network congestion games such that cost functions are in $\mathcal{C}$. Then, the following are equivalent:

(1) $\mathcal{C}$ is consistent for $\mathcal{G}_{\mathrm{dn}}(\mathcal{C})$.
(2) $\mathcal{C}$ is FIP-consistent for $\mathcal{G}_{\mathrm{dn}}(\mathcal{C})$.
(3) $\mathcal{C}$ contains only affine functions *or* $\mathcal{C}$ contains only functions of type $c(x) = a_c\, e^{\phi\, x} + b_c$, where $a_c, b_c \in \mathbb{R}$ may depend on c while $\phi \in \mathbb{R}$ is independent of c.

This characterization is even valid for three-player games.

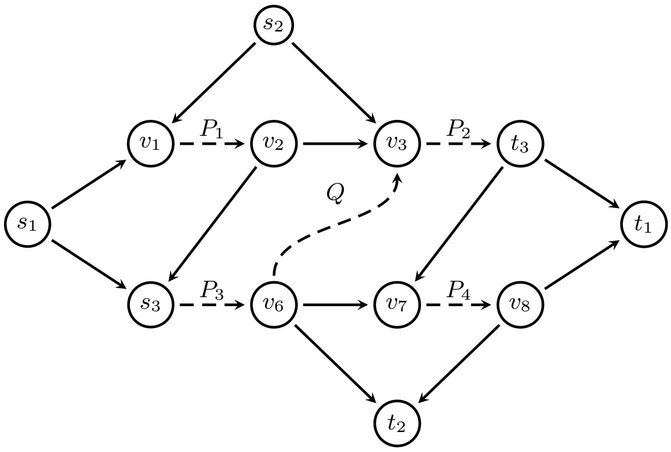


Figure 4. Multi-commodity directed network instance used in the proof of the Extended Monotonicity Lemma for Directed Networks (Lemma 6.3).

Remark 6.5. In games with negative costs the players strive to establish long paths. In this case, our construction does not work since e.g. player 2 prefers to take the detour $v_6 \to v_7 \to v_8 \to t_2$ instead of the edge $v_6 \to t_2$.

6.2. Undirected Networks

We first show that a version of the Extended Monotonicity Lemma holds also for two-player games on undirected networks. In such a game, we are given an undirected graph and for each player i two designated vertices s_i and t_i. Facilities correspond to the edges of the graph and the strategy set of each player i contains all simple s_i, t_i-paths. Each edge can be traversed in any direction and its cost depends on the aggregated flow.

Lemma 6.6 (Extended Monotonicity Lemma for Two-Player Games on Undirected Networks)**.** Let $\mathcal{C}$ be a set of strictly positive and continuous functions. If $\mathcal{C}$ is consistent for two-player undirected network congestion games, then $\mathcal{L}^2_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

Proof. For a contradiction, let $a_1, a_2 \in \mathbb{Z}$ and $c_1, c_2 \in \mathcal{C}$ be such that the function $c : \mathbb{R}_{\geq 0} \to \mathbb{R}$ defined as $c(x) = a_1\, c_1(x) + a_2\, c_2(x)$ is not monotonic and w.l.o.g. $a_2 > 0$. Moreover, let $x, y \in \mathbb{R}_{>0}$ with $y > x$ be such that either $c(y-x) < c(y) < c(x)$ or $c(y-x) > c(y) > c(x)$ holds. We set $d_1 = y - x$, $d_2 = x$ and distinguish the following two cases. If $a_1 < 0$ we consider the network in Figure 5 where the paths P_1 and P_4 each contain $|a_1|$ edges with cost function c_1 and the paths P_2 and P_3 each contain a_2 edges with cost function c_2. With similar calculations as in the proof of Lemma 6.1 one can verify that the 4-cycle $\gamma = \big((P_1 \cup P_2, P_1 \cup P_3), (P_3 \cup P_4, P_1 \cup P_3), (P_3 \cup P_4, P_2 \cup P_4), (P_1 \cup P_2, P_2 \cup P_4), (P_1 \cup P_2, P_1 \cup P_3)\big)$ is an improvement cycle if traversed in the right sense. If on the other hand $a_1 > 0$, we consider the undirected network shown in Figure 5 on the right and obtain the same contradiction as in Lemma 6.1. $\square$

Likewise, we obtain the following characterization for two-player games on undirected networks.

Theorem 6.7. Let $\mathcal{C}$ be a non-empty set of strictly positive and continuous functions and let $\mathcal{G}^2_{\text{un}}(\mathcal{C})$ be the set of two-player undirected network games such that cost functions are in $\mathcal{C}$. Then, the following conditions are equivalent.

(1) $\mathcal{C}$ is consistent for $\mathcal{G}^2_{\text{un}}(\mathcal{C})$.

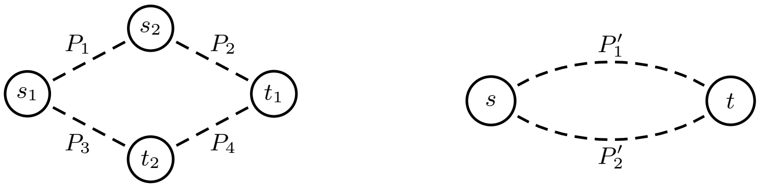


Figure 5. Undirected network congestion games used in the proof of Extended Monotonicity Lemma for Two-Player Undirected Networks (Lemma 6.6).

(2) $\mathcal{C}$ is FIP-consistent for $\mathcal{G}^2_{\text{un}}(\mathcal{C})$.
(3) $\mathcal{C}$ contains only monotonic functions and for all non-constant $c_1, c_2 \in \mathcal{C}$, there are constants $a, b \in \mathbb{R}$ with $c_1(x) = a\,c_2(x) + b$ for all $x \in \mathbb{R}_{\geq 0}$.

Turning to games with three players we are not able to characterize the set of consistent cost functions. However, we can still characterize consistency for games with at least four players.

Lemma 6.8 (Extended Monotonicity Lemma for Undirected Networks)**.** Let $\mathcal{C}$ be a set of strictly positive and continuous functions. If $\mathcal{C}$ is consistent for undirected network congestion games with at least four players, then $\mathcal{L}^3_{\mathbb{Z}}(\mathcal{C})$ contains only monotonic functions.

Proof. For a contradiction, suppose that there are $a_1, a_2 \in \mathbb{Z}$, $\delta \in \mathbb{R}_{>0}$ and a monotonic function $c_1 \in \mathcal{C}$ such that the function $c : \mathbb{R}_{\geq 0} \to \mathbb{R}$ defined as $c(x) = a_1 c_1(x) + a_2 c_1(x + \delta)$ is not monotonic, again w.l.o.g. c_1 is monotonic, $a_2 > 0$ and $a_1 < 0$.

Consider the network in Figure 6 where the paths P_1 and P_4 each contain $|a_1|$ edges and the the paths P_2 and P_3 each contain a_2 edges. The players $i = 1, 2$ correspond to the source sink pairs $(s_i, t_i), i = 1, 2$. Additionally, there are players associated with the source sink pairs (s_i, t_i), $i = 3, 4$ and demand $d_3 = d_4 = \delta$. Moreover, we set $d_1 = y - x$, $d_2 = x$ for $x, y \in \mathbb{R}_{>0}$ with $y > x$ such that either $c(y - x) < c(y) < c(x)$ or $c(y - x) > c(y) > c(x)$ holds (by Lemma 3.2 such values exist).

We endow every edge in the paths $Q_1, \dots, Q_8$ with cost function c_2 and make them sufficiently long such that players 3 and 4 always prefer to choose a strategy not containing any of these paths. Because the paths P_2 and P_3 have costs less than $a_2\,c_2(y + 2\delta)$ and every edge in Q_i, $i = 1, \dots, 8$ used by players 3 or 4 has cost at least $c_2(\delta)$, it suffices for all $i = 1, \dots, 8$ to let Q_i contain $a_2 \left\lceil \frac{c_2(y+2\delta)}{c_2(\delta)} \right\rceil + 1$ edges each. Then, for player 3, all strategies except P_2 are strictly dominated by P_2 and for player 4 all strategies except P_3 are strictly dominated by P_3. Given that players 3 and 4 will not use any of the Q_i-paths in equilibrium, we may assume that players 1 and 2 will not share any of the Q_i paths in equilibrium, w.l.o.g. player 1 always uses the paths $Q_1, \dots, Q_4$ instead of $Q_5, \dots, Q_8$ while player 2 always uses paths $Q_5, \dots, Q_8$ instead of $Q_1, \dots, Q_4$. With the same calculations as before one can show that there is an improvement cycle γ of the form $\gamma = \big((P_u, P_l, P_2, P_3), (P_d, P_l, P_2, P_3), (P_d, P_r, P_2, P_3), (P_u, P_r, P_2, P_3), (P_u, P_l, P_2, P_3)\big)$, where $P_u = P_1 \cup Q_1 \cup Q_2 \cup P_2$, $P_d = P_3 \cup Q_3 \cup Q_4 \cup P_4$, $P_l = Q_5 \cup P_1 \cup P_3 \cup Q_7$, and $P_r = Q_6 \cup P_2 \cup Q_4 \cup Q_8$. Because every strategy profile that uses only non-dominated strategies is contained in γ the thus constructed network congestion game does not admit a PNE. $\square$

Using the above Lemma, we obtain the following result.

Theorem 6.9. Let $\mathcal{C}$ be a set of strictly positive and continuous functions and let $\mathcal{G}_{\text{un}}(\mathcal{C})$ be the set of undirected network congestion games with at least four players and cost functions in $\mathcal{C}$. Then, the following are equivalent:

(1) $\mathcal{C}$ is consistent for $\mathcal{G}_{\text{un}}(\mathcal{C})$.
(2) $\mathcal{C}$ is FIP-consistent for $\mathcal{G}_{\text{un}}(\mathcal{C})$.

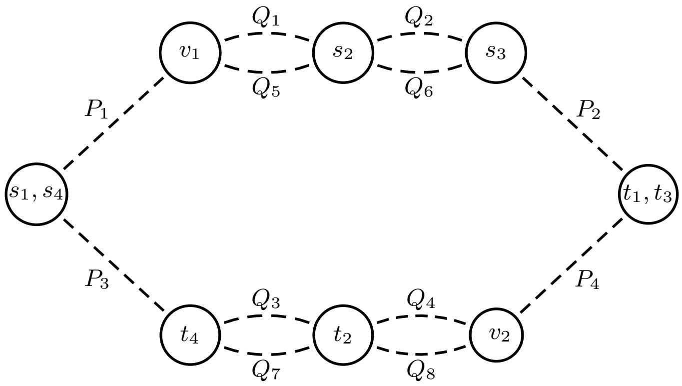


Figure 6. Undirected network congestion games used in the proof of the Extended Monotonicity Lemma for Undirected Networks (Lemma 6.8).

(3) $\mathcal{C}$ contains only affine functions *or* $\mathcal{C}$ contains only functions of type $c(x) = a_c\, e^{\phi\, x} + b_c$, where $a_c, b_c \in \mathbb{R}$ may depend on c while $\phi \in \mathbb{R}$ is independent of c.

For single-commodity network games (directed or undirected) we are not able to characterize consistency of cost functions. However, by introducing a super-source and a super-sink to the network constructions used it follows that the improvement cycles are preserved, thus, all characterizations for FIP-consistency obtained in this section continue to hold.

7. Weighted Singleton Congestion Games

In this section, we consider the case of singleton weighted congestion games. In this class of games, for every player i, every strategy $s_i \in S_i$ contains a single facility only. As mentioned in Corollary 3.4, the construction of the Monotonicity Lemma (Lemma 3.3) is even valid for singleton games, establishing that every set of continuous cost functions $\mathcal{C}$ that is consistent for singleton games may only contain monotonic functions. It is well known that singleton congestion games with weighted players and either only non-decreasing or only non-increasing cost functions admit a PNE, see [14, 15, 36]. Since the positive result for non-decreasing costs is established via a potential function, these games also possess the FIP. With similar arguments it is not difficult to establish the FIP also for the case of non-increasing costs.[1] To the best of our knowledge it was not known before, whether singleton weighted congestion games with both non-decreasing and non-increasing cost functions admit a PNE or even the FIP. Regarding the existence of PNE, for *two-player* games, we give a positive answer to this question.

Theorem 7.1. Let $\mathcal{C}$ be a set of continuous functions and let $\mathcal{G}^2_{\text{sgl}}(\mathcal{C})$ be the set of two-player games such that cost functions are in $\mathcal{C}$ and strategy spaces are sets of singletons. Then, $\mathcal{C}$ is consistent for $\mathcal{G}^2_{\text{sgl}}(\mathcal{C})$ if and only if $\mathcal{C}$ contains only monotonic functions.

Proof. The "only if"-part follows from Corollary 3.4. For the "if"-part let $\mathcal{M} = (N, F, S, (c_f)_{f \in F})$ be a congestion model with $|N| = 2$. W.l.o.g. we assume $d_1 \leq d_2$. We partition the set of facilities into sets F_- and F_+, where F_+ contains all facilities with non-decreasing cost functions (including all facilities with constant functions) and F_- all other facilities. W.l.o.g. we can assume that both player have access to all facilities in F_-, since we can replace the cost function of every facility that is contained in the strategy space of only one player by a constant function.

[1]Consider the function ϕ that assigns to each strategy profile the non-decreasingly sorted vector of the scaled players' private costs $(\pi_i/d_i)_{i \in N}$. Then, ϕ decreases lexicographically along any improvement path, establishing that every such path is finite

We initialize the players both playing g, where $g = \arg\min_{f \in F_-} c_f(d_1 + d_2)$. We distinguish two cases.

First case: Player 1 has an improving move from $(\{g\}, \{g\})$. In this case, we let player 1 move to one of her best replies $\{f_1\} \in S_1$. Using the special choice of g, we have $f_1 \in F_+$. If player 2 does not have an improving move from $(\{f_1\}, \{g\})$, we are done. So, let $\{f_2\}$ be a best reply of player 2 to $(\{f_1\}, \{g\})$. If $f_1 \neq f_2$, we claim that $(\{f_1\}, \{f_2\})$ is a PNE. To see this, note that if $f_2 \in F_+$, then player 2 switching from $\{g\}$ to $\{f_2\}$ does not make any of the facilities more attractive to player 1. If on the other hand, $f_2 \in F_-$, we get $\pi_1(\{f_2\}, \{f_2\}) \geq \pi_1(\{g\}, \{g\}) > \pi_1(\{f_1\}, \{f_2\}) = \pi_1(\{f_1\}, \{g\})$ by the choice of g, thus player 1 does not want to move to f_2 and we have reached an equilibrium. The only interesting case that remains is $\{f_1\} = \{f_2\}$. Again, if player 1 does not have an improving move, there is nothing left to show, so let $\{h_1\} \neq \{f_1\}$ be a best reply of player 1 to $(\{f_1\}, \{f_1\})$. Note that $h_1 \notin F_-$ because otherwise we get $\pi_2(\{f_1\}, \{h_1\})/d_2 \leq \pi_1(\{h_1\}, \{f_1\})/d_1 < \pi_1(\{f_1\}, \{f_1\})/d_1 = \pi_2(\{f_1\}, \{f_1\})/d_2$, where the first inequality follows since $d_2 \geq d_1$. This is a contradiction to the fact that $\{f_1\}$ was a best reply of player 2. As $\pi_2(\{h_1\}, \{f_1\}) \leq \pi_2(\{f_1\}, \{f_1\})$, player 2 does not want to deviate from $(\{h_1\}, \{f_2\})$. Also, player 1 will not deviate from $(\{h_1\}, \{f_2\})$ as $\{h_1\}$ was a best reply.

Second case: Player 1 has no improving move from $(\{g\}, \{g\})$. If also player 2 does not have an improving move from $(\{g\}, \{g\})$, we are done. Otherwise, let $\{f_2\} \in S_2$ be a best reply of player 2. Note that $\{f_2\} \notin S_1$ because otherwise $\{f_2\}$ would have been an improving move from $(\{g\}, \{g\})$ of player 1. If player 1 has no improving move from $(\{g\}, \{f_2\})$, we are done. Otherwise let $\{f_1\}$ be a best reply of player 1 to $(\{g\}, \{f_2\})$. Using that $f_1 \neq f_2$ and that $\pi_2(\{f_1\}, \{f_2\}) \leq \pi_2(\{g\}, \{g\})$, we have that $(\{f_1\}, \{f_2\})$ is a PNE. $\square$

Two-player singleton weighted congestion games with monotonic costs need not possess the FIP as shown in the following example.

Example 7.2. Consider the congestion model $\mathcal{M} = (N, F, S, (c_f)_{f \in F})$ with two players $N = \{1, 2\}$ who have access to all five facilities $F = \{g, f_1, f_2, f_3, f_4\}$. The cost functions of the facilities are shown in Table 1(a). Note that the cost function of facility g is strictly decreasing while all other cost functions are non-decreasing. The players' demands are given by $d_1 = 1$ and $d_2 = 2$. It is not hard to verify that the cycle γ defined as

$$\begin{aligned}\gamma = \big(&(\{g\}, \{g\}), (\{g\}, \{f_1\}), (\{f_1\}, \{f_1\}), (\{f_1\}, \{f_2\}), (\{f_3\}, \{f_2\}),\\ &(\{f_3\}, \{f_3\}), (\{f_4\}, \{f_3\}), (\{f_4\}, \{g\}), (\{g\}, \{g\})\big)\end{aligned}$$

is an improvement cycle.

We proceed showing that for singleton games with three players monotonicity of cost functions alone is not enough for the existence of a PNE. This is illustrated in the following example.

Example 7.3. Consider the congestion model $\mathcal{M} = (N, F, S, (c_f)_{f \in F})$ with $N = \{1, 2, 3\}$ and $F = \{f, g, h\}$. The used cost functions are given in Table 1(b). We claim that the weighted congestion game $G(\mathcal{M}) = (N, S, \pi)$ with $S_1 = \big\{\{g\}, \{h\}\big\}$, $S_2 = \big\{\{f\}, \{g\}\big\}$, $S_3 = \big\{\{f\}, \{h\}\big\}$ and $d_1 = 1$, $d_2 = 2$, $d_3 = 4$ does not admit a PNE. To see this, note that the best-reply graph γ shown in Figure 7 does not have a sink. Using that all strategy profiles are contained in γ the claimed result follows.

However, we are able to give a positive result for *symmetric* games in which the players have access to all facilities.

Theorem 7.4. Let $\mathcal{C}$ be a set of continuous functions and let $\mathcal{G}_{\text{sgl,sym}}(\mathcal{C})$ be the set of games such that cost functions are in $\mathcal{C}$ and strategy spaces are sets of singletons and equal for every player. Then, $\mathcal{C}$ is consistent for $\mathcal{G}_{\text{sgl,sym}}(\mathcal{C})$ if and only if $\mathcal{C}$ contains only monotonic functions.

facility	cost $c(x)$		
	$x=1$	$x=2$	$x=3$
g	10	5	3
f_1	2	2	9
f_2	8	8	8
f_3	1	7	7
f_4	6	6	6

facility	cost $c(x)$					
	$x=1$	$x=2$	$x=3$	$x=4$	$x=5$	$x=6$
f	0	0	2	3	3	3
g	5	1	1	1	0	0
h	2	2	2	2	4	4

Table 1. (a) Cost functions of the five facilities g, f_1, f_2, f_3, and f_4 facilities in the game of Example 7.2; (b) Cost functions of the three facilities f, g, and h in the game of Example 7.3.

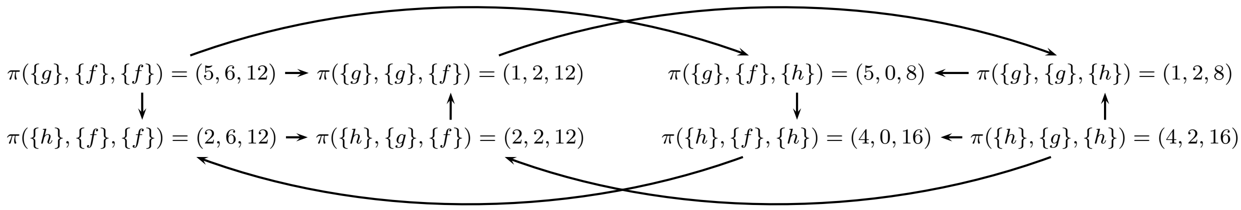


Figure 7. Best reply graph of the singleton weighted congestion game $G(\mathcal{M})$ constructed in Example 7.3. The vertical arcs correspond to best replies of player 1, the straight horizontal arcs to best replies of player 2 and the wide horizontal arcs to best replies of player 3. Since the graph does not have a sink, the game $G(\mathcal{M})$ does not possess a PNE.

Note that the only if part also follows from Corollary 3.4. In order to prove the if part, we give an algorithm that efficiently computes a PNE in such games. In the following, we denote by F_+ and F_- the set of facilities with non-decreasing and non-increasing costs, respectively. In order to obtain a partition of F, we introduce the convention, that facilities with constant cost functions are contained in F_+ only. The algorithm that we propose (Algorithm 1) initializes all players on the facility $g \in F_-$ that minimizes $c_g(\sum_{i\in N} d_i)$. Clearly, then no player has an incentive to switch to another facility $h \in F_-$. The key observation is that, as long as there is at least one player $i \in N$ that wants to switch to a facility $f \in F_+$, also the player with smallest demand does so. So we iteratively take the player with smallest weight on g and let her move to F_+. Then, we compute a sequence of best replies of the players on F_+ in order to assure that none of them has an incentive to deviate to another facility in F_+. Also, the players on F_- are placed on the facility minimizing $c_f(\sum_{i\in N: s_i \in F_-} d_i)$. Since we can prove that a player on F_+ never wants to move back to a facility in F_-, this process stops after a finite number of best-reply steps.

Lemma 7.5. Algorithm 1 computes a PNE.

Proof. Let us first remark that the computation of the partial PNE of players N_+ on F_+ in line 5 finishes after a finite sequence of best replies since the cost functions of the facilities in F_+ are non-decreasing, see [1, 23]. As at most n times such PNE is computed, the algorithm terminates after a finite number of best-reply steps.
Let z denote the outcome of the algorithm. Clearly, no player $j \in N_+$ can improve switching to another facility $f \in F_+$ since we always recompute a partial PNE in line 5. Also, no player $j \in N_-$ can improve unilaterally deviating to another facility $f \in F_-$ since $c_f(d_j) \geq c_f(\sum_{i\in N_-} d_i) \geq c_g(\sum_{i\in N_-} d_i)$. In addition, we know that player $k = \arg\min_{i\in N_-} d_i$ does not improve switching from facility g to another facility $f \in F_+$. In consequence, the same holds for every other player $j \in N_-$ since the cost for her on a facility $f \in F_+$ are not smaller. Finally, it is left to show that in z no player $j \in N_+$ has an interest to switch to some facility $f \in F_-$.

Algorithm 1 Computation of a PNE in symmetric singleton weighted congestion games

Input: Symmetric singleton weighted congestion game G
Output: PNE s of G

1: $N_- := N$, $N_+ := \emptyset$
2: Compute $g := \arg\min_{f\in F_-} c_f(\sum_{i\in N_-} d_i)$ and set $s_i := \{g\}$ for all $i \in N_-$
3: **if** $k = \arg\min_{i\in N_-} d_i$ *can improve switching to* $f \in F_+$ **then**
4: $\quad s_k := f$, $N_- := N_- \setminus \{k\}$, $N_+ := N_+ \cup \{k\}$
5: $\quad$ Compute a partial PNE $(t_i)_{i\in N_+}$ of N_+ on F_+ by best replies and set $(s_i)_{i\in N_+} := (t_i)_{i\in N_+}$
6: $\quad$ **Goto** line 2
7: **end if**
8: **return** s

For proving this, let i_t, $t = 1,\dots,T$, $T \in \mathbb{N}$ denote the player that switches from $g_t \in F_-$ to $f_t \in F_+$ in the t-th iteration of the algorithm and let $\tilde{z}^t$ and z^t denote the corresponding strategy profiles before and after the re-computation of the partial PNE on F_+ in line 5, respectively. We claim that

$$\min_{g\in F_-} c_g(\ell_g(z^t) + d_{i_t}) > \max_{f\in F_+:\ell_f(z^t)>0} c_f(\ell_f(z^t)) \text{ for all } t = 1,\dots,T, \tag{11}$$

where $\ell_g(z^t)$ and $\ell_f(z^t)$ denote the load on facility g respectively f in strategy profile z^t. For $t = 1$, the statement holds true, since player i_1 improves switching from F_- to F_+. Now, suppose (11) holds true for $t-1$. In the t-th iteration, player i_t changes her strategy from $g_t \in F_-$ to some facility $f_t \in F_+$. In consequence, $\min_{g\in F_-} c_g(\ell_g(z^t) + d_{i_t}) = c_{g_t}\big(\ell_{g_t}(z^t) + d_{i_t}\big) > c_{f_t}\big(\ell_{f_t}(\tilde{z}^t)\big)$. As the facilities in F_- have non-increasing cost functions, we obtain

$$\min_{g\in F_-} c_g\big(\ell_g(z^t) + d_{i_t}\big) \geq \min_{g\in F_-} c_g\big(\ell_g(z^{t-1}) + d_{i_{t-1}}\big).$$

By the induction hypothesis, this implies $\min_{g\in F_-} c_g\big(\ell_g(z^t)+d_{i_t}\big) > c_f(\ell_f(\tilde{z}^t))$ for all $f \in F_+\setminus\{f_t\}$ with $\ell_f(\tilde{z}^t) > 0$. Thus, we have established $\min_{g\in F_-} c_g\big(\ell_g(z^t)+d_{i_t}\big) > \max_{f\in F_+:\ell_f(\tilde{z}^t)>0} c_f(\ell_f(\tilde{z}^t))$. Since the maximum cost on F_+ cannot increase in the sequence of best-reply steps (c.f. [21]), we obtain $\min_{g\in F_-} c_g\big(\ell_g(z^t) + d_{i_t}\big) > \max_{f\in F_+:\ell_f(z^t)>0} c_f(\ell_f(z^t))$ as claimed.

Because the algorithm moves always the player with the currently smallest weight from F_- to F_+ (line 3) it holds that $d_{i_T} = \max_{i\in N_+} d_i$ which gives $\min_{g\in F_-} c_g\big(\ell_g(z)+d_i\big) \geq \min_{g\in F_-} c_g\big(\ell_g(z)+d_{i_T}\big) > \max_{f\in F_+} c_f(\ell_f(\tilde{z}))$ for all $i \in N_+$. Thus, no player $i \in N_+$ has an incentive to switch to a facility $g \in F_-$. □

While the above result implies that the set $\mathcal{C}$ of continuous and monotonic cost functions is consistent for symmetric singleton games, Example 7.2 implies that $\mathcal{C}$ is not FIP-consistent.

8. Conclusions

We obtained a characterization of the equilibrium existence problem in weighted congestion games with respect to the facilities' cost functions. The following issues have not been resolved. Our characterizations for network games require that cost functions are strictly positive. Moreover, for single-commodity games we were only able to characterize the FIP, not consistency. The single-commodity case, however, behaves completely different as every congestion game with positive and non-increasing costs admits a PNE in which all players use the socially optimal path (see also Anshelevich et al. [5] for a similar result in the context of network design games). Finally, it would be interesting to characterize consistency of cost functions for undirected networks with three players.

Acknowledgment. We thank Hans-Christian Kreusler for fruitful discussions related to the characterization of continuous and monotonic functions. We are very grateful to the associate editor and the two anonymous referees for their careful proofreading, constructive comments and their discovery of gaps in previous versions of this paper. The research of the second author was supported by the Deutsche Forschungsgemeinschaft within the research training group 'Methods for Discrete Structures' (GRK 1408)

References

[1] H. Ackermann, H. Röglin, and B. Vöcking, *Pure Nash equilibria in player-specific and weighted congestion games*, Theor. Comput. Sci. **410**, no. 17 (2009), pp. 1552–1563.

[2] H. Ackermann and A. Skopalik, *On the complexity of pure Nash equilibria in player-specific network congestion games*, in Proc. 3rd Internat. Workshop on Internet and Network Econom., X. Deng and F. Graham, eds., LNCS 4858, 2007, pp. 419 – 430.

[3] S. Aland, D. Dumrauf, M. Gairing, B. Monien, and F. Schoppmann, *Exact price of anarchy for polynomial congestion games.*, in Proc. 23rd Internat. Sympos. Theoretical Aspects of Comput. Sci., B. Durand and W. Thomas, eds., LNCS 3884, 2006, pp. 218–229.

[4] N. Andelman, M. Feldman, and Y. Mansour, *Strong price of anarchy*, Games Econom. Behav. **65**, no. 2 (2009), pp. 289–317.

[5] E. Anshelevich, A. Dasgupta, J. Kleinberg, É. Tardos, T. Wexler, and T. Roughgarden, *The price of stability for network design with fair cost allocation*, SIAM J. Comput. **38**, no. 4 (2008), pp. 1602–1623.

[6] B. Awerbuch, Y. Azar, and A. Epstein, *The price of routing unsplittable flow*, in Proc. of the thirty-seventh annual ACM Symposium on Theory of Computing, ACM Press, 2005, pp. 57–66.

[7] V. K. Balakrishnan, *Introductory Discrete Mathematics*, Courier Dover Publications, 1996.

[8] U. Bhaskar, L. Fleischer, D. Hoy, and C. Huang, *Equilibria of atomic flow games are not unique*, in Proc. of the 20th Symposium on Discrete Algorithms (SODA), 2009, pp. 748–757.

[9] K. Bhawalkar, M. Gairing, and T. Roughgarden, *Weighted congestion games: Price of anarchy, universal worst-case examples, and tightness*, in ESA (2), 2010, pp. 17–28.

[10] Bureau of Public Roads, *Traffic assignment manual*, U.S. Department of Commerce, Urban Planning Division, Washington, DC (1964).

[11] H. Chen and T. Roughgarden, *Network design with weighted players*, Theor. Comput. Syst. **45**, no. 2 (2009), pp. 302–324.

[12] G. Christodoulou and E. Koutsoupias, *The price of anarchy of finite congestion games*, in Proc. 37th Annual ACM Sympos. Theory Comput., 2005, pp. 67–73.

[13] J. Dunkel and A. Schulz, *On the complexity of pure-strategy Nash equilibria in congestion and local-effect games*, Math. Oper. Res. **33**, no. 4 (2008), pp. 851–868.

[14] E. Even-Dar, A. Kesselman, and Y. Mansour, *Convergence time to Nash equilibrium in load balancing*, ACM Trans. Algorithms **3**, no. 3 (2007), p. 32.

[15] D. Fotakis, S. Kontogiannis, E. Koutsoupias, M. Mavronicolas, and P. Spirakis, *The structure and complexity of Nash equilibria for a selfish routing game*, in Proc. 29th Internat. Colloquium on Automata, Languages and Programming, P. Widmayer, F. Triguero, R. Morales, M. Hennessy, S. Eidenbenz, and R. Conejo, eds., LNCS 2380, 2002, pp. 123 – 134.

[16] D. Fotakis, S. Kontogiannis, and P. Spirakis, *Selfish unsplittable flows*, Theor. Comput. Sci. **348**, no. 2-3 (2005), pp. 226–239.

[17] D. Fotakis, S. Kontogiannis, and P. Spirakis, *Atomic congestion games among coalitions*, in Proc. 33rd Internat. Colloquium on Automata, Languages and Programming, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., LNCS 4052, 2006, pp. 572–583.

[18] M. Gairing, B. Monien, and K. Tiemann, *Routing (un-) splittable flow in games with player-specific linear latency functions*, in Proc. 33rd Internat. Colloquium on Automata, Languages and

Programming, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., LNCS 4051, 2006, pp. 501–512.

[19] M. Goemans, V. Mirrokni, and A. Vetta, *Sink equilibria and convergence*, in Proc. 46th Annual IEEE Sympos. Foundations Comput. Sci., 2005, pp. 142–154.

[20] T. Harks and M. Klimm, *On the existence of pure Nash equilibria in weighted congestion games*, in Proc. 37rd Internat. Colloquium on Automata, Languages and Programming, S. Abramsky, C. Gavoille, C. Kirchner, F. M. auf der Heide, and P. Spirakis, eds., LNCS 6198, 2010, pp. 79 – 89.

[21] T. Harks, M. Klimm, and R. Möhring, *Strong Nash equilibria in games with the lexicographical improvement property*, in Proc. 5th Internat. Workshop on Internet and Network Econom., S. Leonardi, ed., LNCS 5929, 2009, pp. 463–470.

[22] T. Harks, M. Klimm, and R. Möhring, *Characterizing the existence of potential functions in weighted congestion games*, Theor. Comput. Syst. **49**, no. 1 (2011), pp. 46–70.

[23] S. Ieong, R. McGrew, E. Nudelman, Y. Shoham, and Q. Sun, *Fast and compact: A simple class of congestion games*, in Proc. 20th National Conf. Artificial Intelligence and the 17th Innovative Appl. Artificial Intelligence Conf., 2005, pp. 489–494.

[24] L. Libman and A. Orda, *Atomic resource sharing in noncooperative networks*, Telecommunication Systems **17**, no. 4 (2001), pp. 385–409.

[25] I. Milchtaich, *Congestion games with player-specific payoff functions*, Games Econom. Behav. **13**, no. 1 (1996), pp. 111 – 124.

[26] I. Milchtaich, *Topological conditions for uniqueness of equilibrium in networks*, Math. Oper. Res. **30**, no. 1 (2005), pp. 225–244.

[27] I. Milchtaich, *The equilibrium existence problem in finite network congestion games*, in Proc. 2nd Internat. Workshop on Internet and Network Econom., M. Mavronicolas and S. Kontogiannis, eds., LNCS 4286, 2006, pp. 87 – 98.

[28] D. Monderer and L. Shapley, *Potential games*, Games Econom. Behav. **14**, no. 1 (1996), pp. 124 – 143.

[29] A. Orda, R. Rom, and N. Shimkin, *Competitive routing in multi-user communication networks*, IEEE Transactions on Networking **1** (1993), pp. 510–521.

[30] P. Panagopoulou and P. Spirakis, *Algorithms for pure Nash equilibria in weighted congestion games*, ACM J. Exp. Algorithmics **11**, no. 2.7 (2006), pp. 1–19.

[31] O. Richman and N. Shimkin, *Topological uniqueness of the Nash equilibrium for selfish routing with atomic users*, Math. Oper. Res. **32**, no. 1 (2007), pp. 215–232.

[32] R. Rosenthal, *A class of games possessing pure-strategy Nash equilibria*, Internat. J. Game Theory **2**, no. 1 (1973), pp. 65–67.

[33] R. Rosenthal, *The network equilibrium problem in integers*, Networks **3** (1973), pp. 53 – 59.

[34] T. Roughgarden, *Intrinsic robustness of the price of anarchy*, in Proc. 41st Annual ACM Sympos.Theory Comput., M. Mitzenmacher, ed., 2009, pp. 513–522.

[35] T. Roughgarden and É. Tardos, *How bad is selfish routing?*, J. ACM **49**, no. 2 (2002), pp. 236–259.

[36] O. Rozenfeld and M. Tennenholtz, *Strong and correlated strong equilibria in monotone congestion games*, in Proc. 2nd Internat. Workshop on Internet and Network Econom., P. Spirakis, M. Mavronicolas, and S. Kontogiannis, eds., LNCS 4286, 2006, pp. 74 – 86.

[37] H. Yang and X. Zhang, *Existence of anonymous link tolls for system optimum on networks with mixed equilibrium behaviors*, Trans. Res. Part B **42**, no. 2 (2008), pp. 99–112.

Paper 7

Strong Equilibria in Games with the Lexicographical Improvement Property

Tobias Harks, Max Klimm and R. H. Möhring
Strong Nash Equilibria in Games with the Lexicographical Improvement Property
International Journal of Game Theory, accepted s.t. minor revision. Conference version in Proceedings of the 5th International Workshop On Internet And Network Economics (WINE), LNCS 5929, Springer, pp. 455-462, 2009
Full version to appear in International Journal of Game Theory

Abstract. We study a class of finite strategic games with the property that every deviation of a coalition of players that is profitable to each of its members strictly decreases the lexicographical order of a certain function defined on the set of strategy profiles. We call this property the *Lexicographical Improvement Property (LIP)* and show that in finite games it is equivalent to the existence of a generalized strong potential function. We use this characterization to derive existence, efficiency and fairness properties of strong equilibria. As our main result, we show that an important class of games that we call *bottleneck congestion games* has the LIP and thus the above mentioned properties. Turning to infinite games, the LIP does not imply the existence of a generalized strong potential and also the existence of strong equilibria does not follow. We introduce the slightly more general concept of the *pairwise* LIP and prove that whenever the pairwise LIP is satisfied for a continuous function, then there exists a strong equilibrium. As a consequence, we prove that splittable bottleneck congestion games with continuous facility cost functions possess a strong equilibrium.

1. Introduction

The theory of non-cooperative games is used to study situations that involve rational and selfish agents who are motivated by optimizing their own utilities rather than reaching some social optimum. In his seminal work [36] showed that every finite non-cooperative game has an equilibrium in mixed strategies. It is well known that mixed or correlated strategies have no meaningful physical interpretation for many strategic games arising in practice; see also the discussion by Osborne and Rubinstein about critics on mixed Nash equilibria [38, § 3.2]. For such games, one usually resorts to pure strategies and pure Nash equilibria (PNE) are the solution concept of choice. A PNE is a strategy profile from which no player has an incentive to unilaterally change her pure strategy. While the PNE concept excludes the possibility that a single player can unilaterally improve her utility, it does not necessarily imply that a PNE is stable against coordinated deviations of a group of players if their joint deviation is profitable for each of its members. So when coordinated actions are possible, the Nash equilibrium concept is not sufficient to analyze stable states of a game. To cope with the issue of coordination, we adopt the solution concept of a *strong equilibrium* (SE for short) proposed by [5]. In a strong equilibrium, no coalition (of any size) can deviate and strictly improve the utility of each of it members (while possibly lowering the utility of players outside the coalition). Clearly, every SE is a PNE, but the converse does not always hold. Thus, even though SE may rarely exist, they constitute a very robust and appealing stability concept.

One of the most successful approaches in establishing existence of PNE in finite games is the potential function approach introduced by [41] and formalized by [34]. One defines a real-valued function P on the set of strategy profiles of the game and shows that every improving move of a single player strictly reduces the value of P. Since the set of strategy profiles of such a (finite) game is finite, every sequence of improving moves reaches a PNE. In particular, every local minimum[1] of P is a PNE. [22] generalized this concept to *generalized strong potential functions*. Here, it is required that every improving move of a coalition (that is profitable to each of its members) strictly decreased the value of P. Clearly, the global minimum of a generalized strong potential is an SE, while the local minima of P correspond to (potentially non-strong) PNE.

In a recent line of research [16, 15, 2] lexicographic arguments have been used to prove the existence of (strong) equilibria. Here, it is argued that the strategy profile that minimizes the vector of the players' private costs with respect to the lexicographical order is a PNE or an SE. In this paper, we formalize and generalize this approach. We consider strategic games $G = (N, X, \pi)$, where N is the set of players, X the strategy space, and players experience non-negative private costs $\pi_i(x), i \in N$, for a strategy profile x. We say that G has the *lexicographical improvement property (LIP)* if there exists a vector-valued function $\phi : X \to \mathbb{R}^q_+, q \in \mathbb{N}$, such that every improving move (profitable deviation of an arbitrary coalition) from $x \in X$ strictly reduces $\phi(x)$ with respect to the lexicographical order. We say that G has the π-LIP if G satisfies the LIP with $\phi_i(x) = \pi_i(x), i \in N$. Clearly, requiring $q = 1$ in the definition of the LIP reduces to the case of a generalized strong potential.

The main focus of this paper is twofold. We first study desirable properties of arbitrary finite games having the π-LIP. These properties concern the existence, efficiency and fairness of SE. Second, we identify an important and quite general class of games, the *bottleneck congestion games*, for which we can prove the π-LIP and, hence, prove that these games possess SE with the above desirable properties.

Before we outline our results in more detail, let us give an informal definition of bottleneck congestion games. In a standard congestion game, there is a set of facilities, and the pure strategies of players are subsets of this set. Each facility f has a cost that is a function of its *load* usually defined as the number (or total weight) of players that select strategies containing f. The private cost of a player's strategy in a standard congestion game is given by the *sum* of the costs of the facilities in her strategy. In a bottleneck congestion game, the private cost function of a player is equal to the cost of the *most expensive* facility that she uses (L_∞-norm of the vector of players' costs of the strategy). Bottleneck congestion games occur in many real-world applications, e.g., communication networks. Referring to [6], [10], [24] and [39], the throughput of a stream of packets in a communication network is usually determined by the available bandwidth or the capacity of the weakest links. This aspect is captured more realistically by bottleneck congestion games in which the individual cost of a player is the maximum (instead of sum) of the delays

[1] Here, a local minimum is a strategy profile with the property that each other strategy profile that is reachable by a unilateral deviation has no smaller value of P.

in her strategy. Although they are a more realistic model for network routing than classical congestion congestion games, they have not received similar attention in the literature.

1.1. Our Results

We first develop a simple characterization of games having the LIP by means of the existence of a generalized strong potential function. The proof is constructive, that is, given a game G having the LIP for a function ϕ, we explicitly construct a generalized strong potential P. We further investigate games having the π-LIP with respect to efficiency and fairness of SE. Our characterization implies that there are SE satisfying various efficiency and fairness properties, e.g., Pareto efficiency and min-max fairness. Moreover, we derive tight bounds on the strong prices of stability and anarchy.

One of our main results shows that bottleneck congestion games have the π-LIP and, thus, possess SE with the above mentioned properties. Moreover, our characterization of games having the LIP implies that bottleneck congestion games have the strong finite improvement property. Note that for singleton congestion games, [15] and [16], have already proved existence of PNE by arguing that the vector of facility costs decreases lexicographically for every improving move. [2] used the same argument to even establish existence of SE in this case. Our work generalizes these results to arbitrary strategy spaces and more general facility cost functions. In contrast to most congestion games considered so far, we require only that the facility cost functions satisfy three properties: "non-negativity", "independence of irrelevant choices", and "monotonicity". Roughly speaking, the second and third condition assume that the cost of a facility solely depends on the set of players using that facility and that the cost decreases if some players leave that facility, respectively. Thus, this framework extends classical *load-based* models in which the cost of a facility depends on the number or total weight of players using it. Our assumptions are weaker than in the load-based models and even allow that the cost of a facility may depend on the *set* of players using it.

We then study *infinite* games, that is, games with infinite strategy spaces that can be described by compact subsets of $\mathbb{R}^p$, $p \in \mathbb{N}$. We slightly generalize the LIP by introducing the notion of a pairwise vector-valued potential function $\phi : X \to \mathbb{R}^q_+ \times \mathbb{R}^q_+, q \in \mathbb{N}$. Informally, G has the *pairwise lexicographical improvement property* if every coalitional improving move from $x \in X$ strictly reduces a certain lexicographical order of $\phi(x)$ (see Section 5 for the formal definition). We prove that continuity of ϕ in the definition of the pairwise LIP is sufficient for the existence of SE. We then introduce *splittable bottleneck congestion games*. A splittable bottleneck congestion game arises from a bottleneck congestion game G by allowing players to fractionally distribute a certain *demand* over the pure strategies of G. We prove that these games have the pairwise LIP when facility cost functions satisfy non-negativity, independence of irrelevant choices, and monotonicity. If additionally the facility cost functions are continuous we obtain the pairwise LIP for a continuous function ϕ and, thus, we obtain the existence of SE for splittable bottleneck congestion games. For *bounded* cost functions on the facilities (that may be discontinuous), we show that α-approximate SE exist for every $\alpha > 0$.

1.2. Further Related Work

The concept of a strong equilibrium (SE) was introduced by [5] and refined by [7] to coalition-proof Nash equilibria (CPNE). These are states that are stable against those deviations, that are themselves stable against further deviations by subsets of the original coalition. This implies that every SE is also a CPNE, but not conversely.

Congestion games were introduced by [41] and further studied by [34]. [22] studied the existence of SE in congestion games with monotone increasing cost functions. They showed that

SE need not exist in such games and gave a structural characterization of the strategy space for symmetric (and quasi-symmetric) congestion games that admit SE. Based on the previous work of [34], they also introduced the concept of a *generalized strong potential function*: a function on the set of strategy profiles that decreases for every profitable deviation of a coalition. [42] further explored the existence of (correlated) SE in congestion games with non-increasing cost functions.

A further generalization of congestion games has been proposed by [33], who allows for player-specific facility cost functions (for subsequent work on weighted congestion games with player-specific facility cost functions see also [31], [19] and [1]). Milchtaich proves existence of PNE under restrictions on the strategy space (singleton strategies). As shown by [45], the model of [26] is equivalent to that of Milchtaich, which is worth noting as [26] established the existence of SE in these games.

Several authors studied the existence and efficiency of PNE and SE in specific classes of congestion games. For example, [15] showed that job scheduling games (on unrelated machines) have a PNE by arguing that the load-lexicographically minimal schedule is a PNE. [16] considered a scheduling model in which the processing time of a machine may depend on the set of jobs scheduled on that machine. For this model, they proved existence of PNE analogous to the proof of Even-Dar et al. [2] considered scheduling games on unrelated machines and proved that the load-lexicographically minimal schedule is even an SE. They further studied differences between PNE and SE and derived bounds on the (strong) price of anarchy and stability, respectively. [9] recently studied the strong price of anarchy of SE in general congestion games.

Bottleneck congestion games with network structure have been considered by [6]. They studied existence of PNE in the unsplittable flow and in the splittable flow setting, respectively. They observed that standard techniques (such as Kakutani's fixed-point theorem) for proving existence of PNE do not apply to bottleneck routing games, as the players' private cost functions may be discontinuous. They proved existence of PNE by showing that bottleneck games are better reply secure, quasi-convex, and compact. Under these conditions, they recall Reny's existence theorem [40] for better reply secure games with possibly discontinuous private cost functions. Banner and Orda, however, do not study SE. Note that our proof of the existence of SE is direct and constructive. Bottleneck routing with *non-atomic* players and elastic demands has been studied by [10]. Among other results, they derived bounds on the price of anarchy. For subsequent work on the price of anarchy in bottleneck routing games with atomic and non-atomic players, we refer to the paper by [32].

After the publication of a preliminary version of this paper [21], there has been subsequent work on the computational complexity of SE and their worst-case inefficiency. [20] settled the complexity of computing SE for the unit-demand model. [46] studied bottleneck congestion games on networks with weighted demands and identified cases in which there are efficient algorithms computing SE; [12] investigated the worst-case inefficiency of strong equilibria in bottleneck congestion games with affine linear cost functions.

2. Preliminaries

We consider strategic games $G = (N, X, \pi)$, where $N = \{1, \dots, n\}$ is the non-empty and finite set of players, $X = \bigtimes_{i \in N} X_i$ is the non-empty strategy space, and $\pi : X \to \mathbb{R}^n_+$ is the combined *private cost* function that assigns a private cost vector $\pi(x)$ to each strategy profile $x \in X$. These games are cost minimization games and we assume additionally that the private cost functions are non-negative. A strategic game is called *finite* if X is finite. We use standard game theory notation; for a coalition $S \subseteq N$ we denote by $-S$ its complement and by $X_S = \bigtimes_{i \in S} X_i$ we denote the set of strategy profiles of players in S.

Definition 2.1 (Strong equilibrium (SE))**.** A strategy profile x is a strong equilibrium if there is no coalition $\emptyset \neq S \subseteq N$ that has an alternative strategy profile $y_S \in X_S$ such that $\pi_i(y_S, x_{-S}) - \pi_i(x) < 0$ for all $i \in S$.

A pair $\big(x, (y_S, x_{-S})\big) \in X \times X$ is called an *improving move* (or *profitable deviation*) of coalition S if $\pi_i(x_S, x_{-S}) - \pi_i(y_S, x_{-S}) > 0$ for all $i \in S$. We denote by $I(S)$ the set of improving moves of coalition $S \subseteq N$ in a strategic game $G = (N, X, \pi)$ and we set $I = \bigcup_{S \subseteq N} I(S)$. We call a sequence of strategy profiles $\gamma = (x^0, x^1, \dots)$ an *improvement path* if every pair $(x^k, x^{k+1}) \in I$ for all $k = 0, 1, 2, \dots$. One can interpret an improvement path as a path in the *improvement graph* $\mathcal{G}(G)$ derived from G, where every strategy profile $x \in X$ corresponds to a node in $\mathcal{G}(G)$ and two nodes x, x' are connected by a directed edge (x, x') if and only if $(x, x') \in I$. An important property of finite strategic games is the finite improvement property (FIP). This property requires that each improvement path of unilateral improvements is finite. Equivalently, we say that G has the *strong finite improvement property (SFIP)* if every improvement path is finite. Clearly, the SFIP implies the FIP, but not conversely. A necessary and sufficient condition for the SFIP is the existence of a generalized strong potential function, which we define below (see also [34] and [22]).

Definition 2.2 (Generalized strong potential game)**.** A strategic game $G = (N, X, \pi)$ is called a *generalized strong potential game* if there is a function $P : X \to \mathbb{R}$ such that $P(x) - P(y) > 0$ for all $(x, y) \in I$. P is called a *generalized strong potential* of G.

In this paper, we define an equivalent property, the *Lexicographical Improvement Property (LIP)*. For this purpose, we will first define the sorted lexicographical order.

Definition 2.3 (Sorted lexicographical order)**.** Let $a, b \in \mathbb{R}^q_+$ and denote by $\tilde{a}, \tilde{b} \in \mathbb{R}^q_+$ be the sorted vectors derived from a, b by permuting the entries in non-increasing order, that is, $\tilde{a}_1 \geq \cdots \geq \tilde{a}_q$ and $\tilde{b}_1 \geq \cdots \geq \tilde{b}_q$. Then, a is *strictly sorted lexicographically smaller* than b (written $a \prec b$) if there exists an index m such that $\tilde{a}_i = \tilde{b}_i$ for all $i < m$, and $\tilde{a}_m < \tilde{b}_m$. The vector a is *sorted lexicographically smaller* than b (written $a \preceq b$) if either $a \prec b$ or $\tilde{a} = \tilde{b}$.

The lexicographical improvement property of a strategic game requires that there is a vector-valued function $\phi : X \to \mathbb{R}^q_+$ that is strictly decreasing with respect to the sorted lexicographical order on $\mathbb{R}^q_+$ for every improvement step.

Definition 2.4 (Lexicographical improvement property, π-LIP)**.** A finite strategic game $G = (N, X, \pi)$ has the *lexicographical improvement property* (LIP) if there exist $q \in \mathbb{N}$ and a function $\phi : X \to \mathbb{R}^q_+$ such that $\phi(x) \succ \phi(y)$ for all $(x, y) \in I$. G has the *π-LIP* if G has the LIP for $\phi = \pi$.

If a game G has the LIP for a function ϕ, we will call ϕ a *generalized strong vector-valued potential* of G. Clearly, the function ϕ is a generalized strong potential if $q = 1$. The next proposition states that the LIP is equivalent to the existence of a generalized strong potential, regardless of q.

Proposition 2.5. Let $G = (N, X, \pi)$ be a finite strategic game. Then, the following statements are equivalent.

(1) G has a generalized strong vector-valued potential $\phi : X \to \mathbb{R}^q_+, q \in \mathbb{N}$.
(2) G has a generalized strong potential function $P : X \to \mathbb{R}_+$.

Proof. We only prove 1. $\Rightarrow$ 2. as the reverse direction is trivial. We will show that $P_M(x) = \sum_{i=1}^q \phi_i(x)^M$ is a generalized strong potential for M large enough. Let $S \subseteq N$ and $(x, (y_S, x_{-S})) \in I(S)$ be arbitrary. We will calculate $P_{M'}(x) - P_{M'}(y_S, x_{-S}) = \sum_{i=1}^q (\phi_i(x)^{M'} - \phi_i(y_S, x_{-S})^{M'})$ for some M'. To this end, let us denote by $\tilde{\phi}(x)$ and $\tilde{\phi}(y_S, x_{-S})$ the vectors that arise by sorting $\phi(x)$

and $\phi(y_S, x_{-S})$ in non-increasing order. As $\phi(y_S, x_{-S}) \prec \phi(x)$, there is an index $m \in \{1, \dots, q\}$ such that $\tilde{\phi}_i(x) = \tilde{\phi}_i(y_S, x_{-S})$ for all $i < m$ and $\tilde{\phi}_m(x) < \tilde{\phi}_m(y_S, x_{-S})$. We then obtain

$$\begin{aligned} P_{M'}(x) - P_{M'}(y_S, x_{-S}) &= \sum_{i=1}^{q} \phi_i(x)^{M'} - \sum_{i=1}^{q} \phi_i(y_S, x_{-S})^{M'} \\ &= \tilde{\phi}_m(x)^{M'} - \tilde{\phi}_m(y_S, x_{-S})^{M'} + \sum_{i=m+1}^{q} \tilde{\phi}_i(x)^{M'} - \sum_{i=m+1}^{q} \tilde{\phi}_i(y_S, x_{-S})^{M'} \\ &\geq \tilde{\phi}_m(x)^{M'} - \tilde{\phi}_m(y_S, x_{-S})^{M'} - (q-m)\tilde{\phi}_m(y_S, x_{-S})^{M'} \\ &\geq \tilde{\phi}_m(x)^{M'} - q\tilde{\phi}_m(y_S, x_{-S})^{M'}. \end{aligned} \tag{1}$$

Standard calculus shows that the expression on the right hand side of (1) is positive if

$$M' \;>\; \log(q) \,/\, \big(\log(\tilde{\phi}_m(x)) - \log(\tilde{\phi}_m(y_S, x_{-S}))\big) > 0.$$

Clearly, M' depends on $(x, y) \in I$, but as the number of improvement steps is finite, we may chose $M = \max_{(x,y) \in I} M'\big((x, y)\big)$ and obtain the claimed result. ☐

3. Efficiency and Fairness of SE in Games with the π-LIP

As the LIP implies the existence of SE, it is natural to investigate efficiency and fairness properties of these SE. We here consider strict Pareto efficiency, min-max fairness, strong price of anarchy, and strong price of stability.

3.1. Pareto Efficiency

Pareto efficiency is one of the fundamental concepts studied in the economics literature, see [30]. For a strategic game $G = (N, X, \pi)$, a strategy profile x is called *weakly Pareto efficient* if there is no $y \in X$ such that $\pi_i(y) < \pi_i(x)$ for all $i \in N$. A strategy profile x is *strictly Pareto efficient* if there is no $y \in X$ such that $\pi_i(y) \leq \pi_i(x)$ for all $i \in N$, where at least one inequality is strict. So strictly Pareto efficient strategy profiles are those for which every improvement of a coalition of players is to the expense of at least one player outside the coalition. Pareto efficiency has also been studied in the context of standard congestion games (with sum-objective); [22] give sufficient conditions on the strategy spaces of congestion games that guarantee the existence of an SE which is strictly Pareto efficient, [9] quantify the social welfare achieved in weakly Pareto efficient pure Nash equilibria.

Clearly, every SE is weakly Pareto optimal as it is resilient against a profitable deviation of the whole player set N. In games with the π-LIP this result can be strengthened in the sense that there is always an SE, which is even *strictly* Pareto efficient.

Theorem 3.1. Let G be a finite strategic game having the π-LIP. Then, there exists an SE that is strictly Pareto optimal.

Proof. The sorted lexicographic minimum x of π is an SE. To see that it also strictly Pareto efficient, assume by contradiction that there is $y \in X$ and a player i such that $\pi_i(y) < \pi_i(x)$ and $\pi_j(y) \leq \pi_j(x)$ for all $j \in N \setminus \{i\}$. Then, $y \prec x$, contradicting the minimality of x. ☐

3.2. Min-Max-Fairness

Min-max fairness is a central topic in resource allocation in communication networks, see [43] for an overview and pointers to the large body of research in this area. While strict Pareto efficiency requires that there is no alternative profile that improves the cost for a single player without strictly deteriorating the costs of the other players, the notion of min-max-fairness is stronger. A profile x is called *min-max fair* if for any other strategy profile y with $\pi_i(y) < \pi_i(x)$ for some $i \in N$, there exists either $j \in N \setminus \{i\}$ such that $\pi_j(x) \geq \pi_i(x)$ and $\pi_j(y) > \pi_j(x)$, or there exists $j \in N \setminus \{i\}$ such that $\pi_j(x) < \pi_i(x)$ and $\pi_j(y) \geq \pi_i(x)$. Note that in contrast to Pareto efficiency, an improvement that increases the cost of a player with smaller original cost is allowed (up to the threshold $\pi_i(x)$). It is easy to see that every min-max-fair strategy profile is a strictly Pareto efficient state, but not conversely.

Theorem 3.2. Let G be a finite strategic game having the π-LIP. Then, there exists an SE that is min-max fair.

Proof. We show that the strategy profile x minimizing π with respect to the sorted lexicographic order $\preceq$ is min-max fair. Assume by contradiction that there is another strategy profile y such that $\pi_i(y) < \pi_i(x)$ for some $i \in N$ and the following two statements hold:

(1) $\pi_j(y) \leq \pi_j(x)$ for all $j \in N \setminus \{i\}$ with $\pi_j(x) \geq \pi_i(x)$

(2) $\pi_j(y) < \pi_i(x)$ for all $j \in N \setminus \{i\}$ with $\pi_j(x) < \pi_i(x)$.

We can observe that every entry of $\pi(x)$, which is above $\pi_i(x)$ only decreases under y, while every entry strictly below $\pi_i(x)$ may only increase to a value strictly below the threshold $\pi_i(x)$. Because the value $\pi_i(x)$ strictly decreases under y, we obtain $\pi(y) \prec \pi(x)$, contradicting the minimality of x. □

3.3. Price of Stability and Price of Anarchy

To quantify the efficiency loss of selfish behavior with respect to a predefined social cost function, two notions have evolved. The *price of anarchy* has been introduced by [28] in the context of congestion games and is defined as the ratio of the cost of the worst pure Nash equilibrium and that of the social optimum. A more optimistic performance index termed the *price of stability* measures the ratio of the cost of the best pure Nash equilibrium and that of the social optimum [3, 4]. Both concepts have been studied extensively in the computer science and operations research literature, see [37, Part III] for a survey. More recently, they have also been studied in the economic literature, see e.g. [23, 35].

[2] propose to study also the worst case ratio of the cost of an SE and that of a social optimum, which they term the *strong price of anarchy*. Clearly, the strong price of anarchy is not larger than the price of anarchy. For some classes of games this inequality is strict, see e.g. the results of [11] and [18] on the price of anarchy and strong price of anarchy of scheduling games on related machines, respectively. [2] also define the *strong price of stability* in the obvious way as the ratio of the cost of the best SE and that of the social optimum. Formally, given a game $G = (N, X, \pi)$ and a social cost function $C : X \to \mathbb{R}_+$, whose minimum is attained in a strategy profile $y \in X$, let $X^{\mathrm{SE}} \subseteq X$ denote the set of strong equilibria. Then, the strong price of anarchy for G with respect to C is defined as $\sup_{x \in X^{\mathrm{SE}}} C(x)/C(y)$ and the strong price of stability for G with respect to C is defined as $\inf_{x \in X^{\mathrm{SE}}} C(x)/C(y)$. We will consider the following natural social cost functions: the sum-objective or L_1-norm defined as $L_1(x) = \sum_{i \in N} \pi_i(x)$, the L_p-objective or L_p-norm, $p \in \mathbb{N}$, defined as $L_p(x) = (\sum_{i \in N} \pi_i(x)^p)^{1/p}$, and the min-max objective or L_∞-norm defined as $L_\infty(x) = \max_{i \in N}\{\pi_i(x)\}$.

Theorem 3.3. Let G be a finite strategic game with the π-LIP. Then, the strong price of stability w.r.t. L_∞ is 1, and for any $p \in \mathbb{N}$, the strong price of stability w.r.t. L_p is less or equal to $n^{1/p}$.

Proof. To see that the strong price of stability w.r.t. L_∞ is 1, note that a lexicographic minimum x^* of π is an SE. By construction, x^* minimizes L_∞.

For the proof of the result concerning L_p we first show that for arbitrary $p, q \in \mathbb{N}$ with $p < q$ and $x \in \mathbb{R}^n_+$ we have $L_p(x) \leq n^{1/p-1/q} L_q(x)$ and $L_p(x) \leq n^{1/p} L_\infty(x)$. To see the first inequality,

	l	r
u	$(k-\epsilon, k-\epsilon, k-\epsilon, \dots, k-\epsilon)$	$(k, k, k, \dots, k)$
d	$(k, 0, 0, \dots, 0)$	$(k, \epsilon, k, \dots, k)$

	l	r
u	$(0, 0)$	$(0, k)$
d	(k, k)	$(0, k)$

Figure 1. (a) Private costs received by the players for strategy profiles $X_1 \times X_2$ of the game considered in Example 3.4. (b) A game with unbounded price of anarchy w.r.t. any L_p-norm as considered in Example 3.5.

set $a = \frac{q}{p} > 1$ and $b > 0$ be such that $\frac{1}{a} + \frac{1}{b} = 1$. By Hölder's inequality we have

$$L_p(x) = \Big(\sum_{i=1}^n x_i^p\Big)^{1/p} \leq \left(\Big(\sum_{i=1}^n x_i^{p \cdot a}\Big)^{1/a}\Big(\sum_{i=1}^n 1\Big)^{1/b}\right)^{1/p} = n^{1/(pb)} L_q(x) = n^{1/p-1/q} L_q(x).$$

For the L_∞-norm we have $a = \infty$ and $b = 1$, thus, we obtain $L_p(x) \leq n^{1/p} L_\infty$.

Next, let x^* be a lexicographic minimum of π. Fix $p \in N$ and let y be a strategy profile minimizing L_p. We derive the inequalities $L_p(x^*) \leq n^{1/p} L_\infty(x^*) \leq n^{1/p} L_\infty(y) \leq n^{1/p} L_p(y)$, where we use for the second inequality that x^* minimizes L_∞ and for the third inequality that the L_p-norm is decreasing in p. □

We now provide an example of a class of games with the π-LIP whose parameters can be chosen in such a way that the price of stability w.r.t. L_p is arbitrarily close to $n^{1/p}$, implying that the result of Theorem 3.3 is tight.

Example 3.4 (Price of stability)**.** Consider the game $G = (N, X, \pi)$ with $N = \{1, \dots, n\}$, $X_1 = \{u, d\}$, $X_2 = \{l, r\}$ and $X_i = \{z\}$ for $3 \leq i \leq n$. For $k > \epsilon$, the private costs are shown in Fig. 1 (left). It is straightforward to check that this game has the π-LIP. The unique SE is the strategy profile $(u, l, z, \dots, z)$ realizing a private cost vector of $(k-\epsilon, \dots, k-\epsilon)$. For any $p \in \mathbb{N}$, there is $\epsilon > 0$ such that $L_p(\cdot)$ is maximized in strategy profile $(d, l, z, \dots, z)$ realizing a cost vector of $(k, 0, \dots, 0)$. Hence the price of stability approaches $n^{1/p}$.

So far, our results concern the price of stability only. The next example shows that games with the π-LIP may have an unbounded price of anarchy.

Example 3.5 (Unbounded price of anarchy)**.** Consider the game $G = (N, X, \pi)$ with $N = \{1, 2\}$, $X_1 = \{u, d\}$, $X_2 = \{l, r\}$ and private costs given in Fig. 1 (right) for any $k > 0$. It is straightforward to check that this game has the π-LIP and that both (u, l) and (d, r) are SE. Hence, the price of anarchy w.r.t. any L_p norm is unbounded from above.

4. Bottleneck Congestion Games

We now present a rich class of finite games satisfying the π-LIP. We call these games *bottleneck congestion games*. They are natural generalizations of variants of congestion games. In contrast to standard congestion games, we focus on *bottleneck-objectives*, that is, the cost of a player only depends on the highest cost of the facilities she uses. For the sake of a clean mathematical definition, we introduce the general notion of a congestion model.

Definition 4.1 (Congestion model)**.** A tuple $\mathcal{M} = (N, F, X, (c_f)_{f \in F})$ is called a *congestion model* if $N = \{1, \dots, n\}$ is a non-empty, finite set of players, $F = \{1, \dots, m\}$ is a non-empty set of facilities, and $X = \bigtimes_{i \in N} X_i$ is the set of strategies. For each player $i \in N$, her collection of pure strategies X_i is a non-empty set of subsets of F. Given a strategy profile x, we define $\mathcal{N}_f(x) = \{i \in N : f \in x_i\}$ for all $f \in F$. Every facility $f \in F$ has a cost function $c_f : \bigtimes_{i \in N} X_i \to \mathbb{R}_+$ satisfying

***Non-negativity*:** $c_f(x) \geq 0$ for all $x \in X$,
***Independence of Irrelevant Choices*:** $c_f(x) = c_f(y)$ for all $x, y \in X$ with $\mathcal{N}_f(x) = \mathcal{N}_f(y)$,
***Monotonicity*:** $c_f(x) \leq c_f(y)$ for all $x, y \in X$ with $\mathcal{N}_f(x) \subseteq \mathcal{N}_f(y)$.

Given a congestion model, we now define bottleneck congestion games.

Definition 4.2 (Bottleneck congestion game)**.** Let $\mathcal{M} = (N, F, X, (c_f)_{f\in F})$ be a congestion model. The corresponding *bottleneck congestion game* is the strategic game $G(\mathcal{M}) = (N, X, \pi)$ in which π is defined as $\pi = \bigtimes_{i\in N} \pi_i$ and $\pi_i(x) = \max_{f\in x_i} c_f\big(x\big)$.

A bottleneck congestion game with $|x_i| = 1$ for all $x_i \in X_i$ and $i \in N$ will be called a *singleton* bottleneck congestion game. Note that for singleton strategies, congestion games with bottleneck objective and congestion games with sum-objective coincide.

Our assumptions on the cost functions are weaker than the load-based models often used in the congestion games literature, e.g., [6]. In our approach, we only require that the cost function $c_f(x)$ of facility f for strategy profile x depends on the set of players using f in x and that costs are increasing with larger sets. Note that this may cover, e.g., dependencies on the identities of players using f. Our condition "Independence of Irrelevant Choices" is also weaker than the one frequently used in the literature. In [25, 26, 27], the definition of "Independence of Irrelevant Choices" requires that the strategy spaces are symmetric and, given a strategy profile $x = (x_1, \ldots, x_n)$, the utility of a player i depends only on her own choice x_i and the cardinality of the set of other players who also choose x_i. On the one hand, our model is more general as it does neither require symmetry of strategies, nor that the utility of player i only depends on the set-cardinality of other players who also choose x_i. On the other hand, the model of Konishi et al allows for player-specific facility cost functions, which our model does not. For the relationship between games considered by [25, 26, 27] and congestion games, see the discussion in [45].

Before we prove that bottleneck congestion games have the π-LIP and thus possess an SE with the efficiency and fairness properties shown in the last section, we give a series of examples of games that fit into the rich class of bottleneck congestion games and show how they are related to the literature.

Scheduling Games. Scheduling games model situations where each player controls a task that needs to be processed by one machine out of a finite number of available machines, see [44] for a survey. In each strategy profile each player $i \in N$ selects a single machine $x_i \in X_i$ where her job is processed. In the most general machine model of unrelated machines each job is associated with a machine-dependent weight $w_{i,f} \in \mathbb{R}_+$. Scheduling games are singleton bottleneck congestion games where the cost function of machine f is defined as $c_f(x) = \sum_{i\in N: x_i = f} w_{i,f}$. This function satisfies non-negativity, independence of irrelevant choices and monotonicity. The existence of SE in scheduling games has been established before by [2] by arguing that the lexicographically minimal schedule is a strong equilibrium. They also showed that the strong price of stability w.r.t. L_∞ is 1. Note that our general framework of bottleneck congestion games allows more complex cost structures on the machines than in these classical load-based models. One such example are dependencies between the weights of jobs on the same machine.

Resource Allocation in Wireless Networks. Interference games are motivated by resource allocation problems in wireless networks. Consider a set of n terminals that want to connect to one out of m available base stations. Terminals assigned to the same base station impose interferences among each other as they use the same frequency band. We model the interference relations by an undirected interference graph $\mathcal{D} = (V, E)$, where $V = \{1, \ldots, n\}$ is the set of vertices/terminals and an edge $e = (v, w)$ between terminals v, w has a non-negative weight $w_e \geq 0$ representing the level of pair-wise interference. We assume that the service quality of a base station j is proportional to the total interference $w(j)$, which is defined as $w(j) = \sum_{(v,w)\in E: x_v = x_w = j} w_{(v,w)}$.

We now obtain an interference game as follows. The nodes of the graph are the players, the set of strategies is given by $X_i = \{1, \dots, m\}, i = 1, \dots, n$, that is, the set of base stations, and the private cost function for every player is defined as $\pi_i(x) = w(x_i),\ i = 1, \dots, n$. Interference games fit into the framework of singleton bottleneck congestion games with m facilities.

Note that in interference games, we crucially exploit the property that facility cost functions depend on the *set* of players using the facility, that is, their *identity* determines the resulting cost. While the existence of an SE in all interference games follows from our main theorem, most previous game-theoretic works addressing wireless networks only considered Nash equilibria, see for instance [29] and [14].

Bottleneck Routing in Networks. A special case of bottleneck congestion games are bottleneck routing games. Here, the set of facilities are the edges of a directed or undirected graph $\mathcal{D} = (V, A)$. Every edge $a \in A$ has a load dependent cost function c_a. Every player is associated with a pair of vertices (s_i, t_i) and a fixed demand $d_i > 0$ that she wishes to send along the chosen path in $\mathcal{D}$ connecting s_i to t_i. The private cost for every player is the maximum arc cost along the path, which is a common assumption for data routing in computer networks, see [24, 6, 10, 39]. The existence of PNE in bottleneck routing games has been studied before by [6]. They, however, did not study the existence of SE. To the best of our knowledge, our main result (Theorem 4.3) establishes also for the first time that bottleneck routing games have the FIP; [6] only proved that best-response dynamics converge.

4.1. Existence of SE

We are now ready to state our main result for bottleneck congestion games, providing a large class of games that satisfies the π-LIP.

Theorem 4.3. Every bottleneck congestion game has the π-LIP.

Proof. For an arbitrary improving move $(x, (y_S, x_{-S})) \in I$, let $j \in S$ be a member of the coalition with highest cost before the improvement step, i.e., $j \in \arg\max_{i \in S} \pi_i(x)$. We set $N^+ = \{i \in -S : \pi_i(x) \geq \pi_j(x)\}$ and claim that $\pi_i(x) \geq \pi_i(y_S, x_{-S})$ for all $i \in N^+$. To see this, suppose there is $i \in N^+$ such that $\pi_i(x) < \pi_i(y_S, x_{-S})$. The independence of irrelevant choices and the monotonicity of the cost functions imply that there is a member $k \in S$ of the coalition with $y_k \cap x_i \neq \emptyset$. We obtain

$$\pi_j(x) \geq \pi_k(x) > \pi_k(y_S, x_{-S}) \geq \pi_i(y_S, x_{-S}) > \pi_i(x),$$

which contradicts $i \in N^+$. Next, we define $N^- = \{i \in -S : \pi_i(x) < \pi_j(x)\}$ and claim that $\pi_i(y_S, x_{-S}) < \pi_j(x)$ for all $i \in N^-$. To see this, suppose there is $i \in N^-$ such that $\pi_i(y_S, x_{-S}) \geq \pi_j(x)$. Because $\pi_j(x) \geq \pi_i(x)$, the independence of irrelevant choices and the monotonicity of the cost functions, there is a member $k \in S$ of the coalition with $y_k \cap x_i \neq \emptyset$ giving rise to

$$\pi_j(x) \geq \pi_k(x) > \pi_k(y_S, x_{-S}) \geq \pi_i(y_S, x_{-S}) \geq \pi_j(x),$$

which is a contradiction. Note that $N = N^+ \cup N^- \cup S$ and that we have shown $\pi_i(x) \geq \pi_i(y_S, x_{-s})$ for all $i \in N^+$ and $\pi_i(y_S, x_{-S}) < \pi_j(x)$ for all $i \in N^-$. As the private cost of the players with cost larger than $\pi_j(x)$ does not increase, the private cost of player j strictly decreases, and the private costs of all other players may only increase up to a value strictly smaller than $\pi_j(x)$, we have $\pi(x) \succ \pi(y_S, x_{-S})$ as claimed. □

As a corollary of Theorem 4.3 we obtain that bottleneck congestion games possess SE with the efficiency and fairness properties shown in Section 3. Note that our existence result holds for *arbitrary* strategy spaces. This contrasts a result of [22] who have shown that for standard congestion games (with sum-objective) a certain combinatorial property of the players' strategy spaces (called *good configuration*) is necessary and sufficient for the existence of SE.

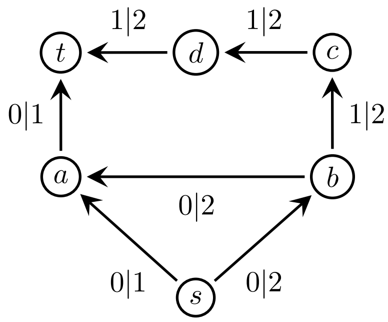


Figure 2. Bottleneck routing game with multiple SE.

In bottleneck congestion games, the vector-valued potential function need not be unique. In fact, one can prove with similar arguments as in the proof of Theorem 4.3 that the function $\psi : X \to \mathbb{R}_+^{mn}$ defined as $\psi_{i,f}(x) = c_f(x)$, if $f \in x_i$, and $\psi_{i,f} = 0$, otherwise, decreases lexicographically along any improvement path. Moreover, if cost functions are strictly monotonic, one can show along the same lines, that also the function $\upsilon : X \to \mathbb{R}^m$ defined as $\upsilon(x) = (c_f(x))_{f \in F}$ has this property. Interestingly, the lexicographical minima of the functions π, ψ, and υ need not coincide, as illustrated in the following example.

Example 4.4. Consider the symmetric bottleneck routing game with two players $N = \{1, 2\}$ depicted in Fig. 2. Here, edges correspond to facilities; the cost of each edge depends only on the number of players using it and is given explicitly for the two possible values. The strategy set X_i of each player $i \in N$ comprises all paths from s to t, that are $P_1 = \{(sa), (at)\}, P_2 = \{(sb), (bc), (cd), (dt)\}$ and $P_3 = \{(sb), (ba), (at)\}$. There are three types of SE. In the first type, one player plays P_1 and the other player plays P_2. Here, the player on P_1 experiences a cost of 0 while the player on P_2 experiences a cost of 1. It is easy to see, that (upon permutation of the two players) this strategy profile is the unique lexicographical minimum of π. In the second type of SE one player chooses P_1 while the other player chooses P_3. Here, both players experience a cost of 1, thus this SE is not strictly Pareto efficient. It is easy to see that this equilibrium minimizes lexicographically both ψ and ν. There is a third SE where both players choose P_1. This profile minimizes none of the functions π, ψ, and υ. These different SE have also different efficiency properties. While the lexicographical minimum x^π of π is strictly Pareto efficient and min-max fair (as show in Theorems 3.1 and 3.2), the lexicographical minimum x^υ of υ has the property that it is strictly Pareto efficient with respect to using the resources, i.e., there is no strategy profile $y \in X$ such that $c_f(y) \leq c_f(x^\upsilon)$ for all $f \in F$ where at least one of these inequalities is strict.

5. Infinite Strategic Games

We now consider *infinite* strategic games in which the players' strategy sets are topological spaces and the private cost functions are defined on the product topology. Formally, an infinite game is a tuple $G = (N, X, \pi)$, where $N = \{1, \ldots, n\}$ is a set of players, and $X = X_1 \times \cdots \times X_n$ is the set of pure strategies, where we assume that $X_i \subseteq \mathbb{R}^{n_i}$, $n_i \in \mathbb{N}$, $i \in N$ are compact sets. The cost function for player i is defined by a non-negative real-valued function $\pi_i : X \to \mathbb{R}_+$, $i \in N$. Turning from finite games to infinite games, it becomes more complicated to characterize structural properties of games having the LIP. First, for an infinite game (as described above), Proposition 2.5 is no longer valid, that is, infinite games with the LIP need not possess a generalized strong potential.[2] Also the existence of an SE does not immediately follow. The global minimum of the function ϕ associated with the LIP need not exist as the strategy space is not finite. We will show that continuity of ϕ is sufficient for the existence of an SE. However, this assumption may be too strong for many classes of games. For instance, the splittable version of bottleneck congestion games (formally defined in Section 5.1) has the π-LIP but the function π may be discontinuous in general.

To obtain existence results for SE also for splittable bottleneck congestion games, we slightly generalize the lexicographical improvement property. Let $G = (N, X, \pi)$ be an infinite game and let $\phi : X \to \mathbb{R}^q_+ \times \mathbb{R}^q_+$ be a function that associates with each strategy profile a pair $\phi(x) = (\phi^{(1)}(x), \phi^{(2)}(x))$. For two indices $i, j \in \{1, \dots, q\}$ and two strategy profiles $x, y \in X$, let $\phi_i(x) \leq \phi_j(y)$ if and only if $\phi^{(1)}_i(x) < \phi^{(1)}_j(y)$ or $\phi^{(1)}_i(x) = \phi^{(1)}_j(y)$ and $\phi^{(2)}_i(x) \leq \phi^{(2)}_j(y)$. Let $\phi_i(x) < \phi_j(y)$ if and only if $\phi_i(x) \leq \phi_j(y)$ and $\phi_i(x) \neq \phi_j(y)$. Moreover, let $\preceq$ denote the sorted lexicographical order, where $\phi_i(x)$ is sorted according to $\leq$. Then, we say that ϕ is a *pairwise strong vector-valued potential* if $\phi(y) \prec \phi(x)$ for all $(x, y) \in I$. G has the *pairwise lexicographical improvement property* if it admits a pairwise strong vector-valued potential.

Clearly, every game with the LIP has also the pairwise LIP, as we may simply set the second component of the pairwise strong vector-valued potential equal to the first component (or, alternatively, equal to zero). We proceed showing that every game with a continuous pairwise strong vector-valued potential admits an SE.

Theorem 5.1. Every infinite game having a continuous pairwise strong vector-valued potential ϕ possesses an SE.

Proof. By assumption, there exists $q \in \mathbb{N}$ and a function $\phi : X \to \mathbb{R}^q_+ \times \mathbb{R}^q_+$ such that $\phi(y_S, x_{-S}) \prec \phi(x)$ for all $(x, (y_S, x_{-S})) \in I$.

To get the desired result, we will show by complete induction over $q \in \mathbb{N}$ that for each $q \in \mathbb{N}$, each compact $X \neq \emptyset$ and each continuous function $\phi : X \to \mathbb{R}^q_+ \times \mathbb{R}^q_+$ there is strategy profile $x_{\min} \in X$ with $x_{\min} \preceq x$ for all $x \in X$.

For the base case $q = 1$, let $Y = \{x \in X : \phi^{(1)}(x) = \min_{x \in X} \phi^{(1)}(x)\}$ be the subset of X where the first component $\phi^{(1)}$ is minimized. Note that Y is non-empty and compact as ϕ is continuous and X is compact. Next, let $Y' = \{x \in Y : \phi^{(2)}(x) = \min_{x \in Y} \phi^{(2)}(x)\}$. With the same arguments, $Y' \neq \emptyset$ and by construction, Y' contains all vectors that minimize ϕ.

For the induction step, suppose that for fixed $k \in \mathbb{N}$ the statement holds true for all functions $\phi' : X' \to \mathbb{R}^q_+ \times \mathbb{R}^q_+$ with $q \leq k-1$ and consider an arbitrary compact X and an arbitrary function $\phi : X \to \mathbb{R}^k_+ \times \mathbb{R}^k_+$. In order to construct a lexicographical minimum of ϕ, we set $K = \{1, \dots, k\}$ and solve the minimization problem

$$\min_{x \in X} \max_{i \in K} \phi^{(1)}_i(x) \tag{2}$$

of minimizing the maximum value within the first component of ϕ. Let α be the optimal value of (2). Intuitively, α is the maximum value in the first component of all candidate lexicographically minimal vectors of ϕ. It follows that for each candidate lexicographically minimal vector $x \in X$, there will be a subset $J \subseteq K$ of indices such that $\phi^{(1)}_j(x) = \alpha$ for all $j \in J$. For arbitrary $\emptyset \neq J \subseteq K$, we set

$$Y^J = \{x \in X : \phi^{(1)}_i(x) \leq \alpha \ \forall i \in K \setminus J,\ \phi^{(1)}_j(x) = \alpha \ \forall j \in J\}$$

[2] This observation resembles Debreu's result showing that the lexicographical ordering on an uncountable subset of $\mathbb{R}^2$ cannot be represented by a real-valued function [13].

and solve the minimization problem

$$\alpha^J = \min_{x \in Y^J} \max_{j \in J} \phi_j^{(2)}(x), \tag{3}$$

where we set $\alpha^J = \infty$, if $Y^J = \emptyset$. For each $\emptyset \neq J \subseteq K$ with $\alpha^J < \infty$ and $j \in J$, consider the compact set $Y^{J,j} = \{x \in X : \phi_i^{(1)}(x) \leq \alpha \; \forall i \in K \setminus J, \phi_i^{(1)}(x) = \alpha \; \forall i \in J, \phi_j^{(2)}(x) = \alpha^J\}$ and the function $\phi^{J,j} : Y^{J,j} \to \mathbb{R}_+^{k-1}$ that arises from ϕ by deleting the j-th index, i.e., $\phi_i^{J,j}(y) = (\phi_i^{(1)}(y), \phi_i^{(2)}(y))$ for all $i < j$ and $\phi_i^{J,j}(y) = (\phi_{i+1}^{(1)}(y), \phi_{i+1}^{(2)}(y))$ for all $i \in \{j, \dots, k-1\}$. Clearly, for all $J \in 2^K \setminus \emptyset$ with $\alpha^J < \infty$ and $j \in J$, the function $\phi^{J,j}$ is continuous and its domain $Y^{J,j}$ is compact. Note that additionally for all $J \in 2^K \setminus \emptyset$ with $\alpha^J < \infty$, there is at least one $j \in J$ with $Y^{J,j} \neq \emptyset$. Next, for all $J \in 2^K \setminus \emptyset$ and $j \in J$ with $Y^{J,j} \neq \emptyset$ we apply the induction hypothesis and obtain finitely many vectors $y_{\min}^{J,j}$ minimizing $\phi^{J,j}$ on $Y^{J,j}$. We claim that the lexicographically minimal vector among the vectors $\big((\alpha, \alpha^J), y_{\min}^{J,j}\big) \in \mathbb{R}_+^k \times \mathbb{R}_+^k$ for each such pair J, j is also a lexicographical minimum of the original function ϕ on X. For a contradiction, suppose that there is a vector $z \in X$ with $\phi(z) \prec \big((\alpha, \alpha^J), y_{\min}^{J,j}\big)$ for all such J, j. Note that there is a non-empty set $\emptyset \neq J' \subseteq K$ such that $\phi_i^{(1)}(z) = \alpha$ for all $i \in J'$ as otherwise we obtain a contradiction to the fact that α is the optimal value of (2). Moreover, because α^J is the optimal value of (3), for at least one index $j \in J'$, we have $\phi_{j'}^{(2)}(z) = \alpha^{J'}$. Using this fact together with the induction hypothesis that $y_{\min}^{J',j'}$ is minimal among the vectors with $\phi_j^{(1)}(z) = \alpha$ for all $j \in J'$ and $\phi_{j'}^{(2)} = \alpha^{J'}$ gives the contradiction.

□

5.1. Splittable Bottleneck Congestion Games

In this section, we introduce the splittable counterpart of bottleneck congestion games. We start with a congestion model $\mathcal{M} = (N, F, X, (c_f)_{f \in F})$ with $X_i = \{x_{i,1}, \dots, x_{i,n_i}\}$, $n_i \in \mathbb{N}$, $i \in N$, where as usual every $x_{i,j}$ is a subset of facilities of F. From $\mathcal{M}$ we derive a corresponding *splittable congestion model* $\mathcal{M}_s = (N, F, X, d, \Delta, (c_f)_{f \in F})$, where $d \in \mathbb{R}_+^n$, $\Delta = \Delta_1 \times \dots \times \Delta_n$, and

$$\Delta_i = \Big\{\xi_i = (\xi_{i,1}, \dots, \xi_{i,n_i}) : \; \xi_{i,k} \geq 0 \; \forall k \in \{1, \dots, n_i\}, \; \sum_{k=1}^{n_i} \xi_{i,k} = d_i\Big\}.$$

The strategy profile $\xi_i = (\xi_{i,1}, \dots, \xi_{i,n_i})$ of player i can be interpreted as a distribution of non-negative *intensities* over the elements in X_i satisfying $\sum_{k=1}^{n_i} \xi_{i,k} = d_i$ for $d_i \in \mathbb{R}_+, i \in N$. Clearly, Δ_i is a compact subset of $\mathbb{R}_+^{n_i}$ for all $i \in N$. For a profile $\xi = (\xi_1, \dots, \xi_n)$, we define $\xi_{i,f} = \sum_{k \in \{1,\dots,n_i : f \in x_{i,k}\}} \xi_{i,k}$ as the total intensity put on facility f by player i; the set of used facilities of player i is defined as $F_i(\xi) = \{f \in F : \xi_{i,f} > 0\}$. We assume that for all $f \in F$ the cost function $c_f : \Delta \to \mathbb{R}_+$ satisfies the assumptions

***Non-negativity*:** $c_f(\xi) \geq 0$ for all $\xi \in \Delta$,
***Independence of Irrelevant Choices*:**
$c_f(\xi) = c_f(\xi')$ for all $\xi, \xi' \in \Delta$ with $\xi_{i,f} = \xi'_{i,f}$ for all $i \in N$,
***Monotonicity*:** $c_f(\xi) \leq c_f(\xi')$ for all $\xi, \xi' \in \Delta$ with $\xi_{i,f} \leq \xi'_{i,f}$ for all $i \in N$,
***Continuity*:** $c_f(\xi)$ is continuous in $\xi_{i,f}$ for all $i \in N, f \in F$.

Up to continuity, we basically impose the same assumptions as in the case of finite bottleneck congestion games.

Definition 5.2 (Splittable bottleneck congestion game)**.** For the splittable congestion model $\mathcal{M}_s = (N, F, X, d, \Delta, (c_f)_{f \in F})$, we define the corresponding *splittable bottleneck congestion game* as the infinite strategic game $G(\mathcal{M}_s) = (N, \Delta, \pi)$, where π is defined as $\pi = \bigtimes_{i\in N} \pi_i$ and $\pi_i(\xi) = \max_{f\in F_i(\xi)} c_f(\xi)$.

Before we show that splittable bottleneck congestion games always have an SE, we give examples of games that fit into the model.

Bottleneck routing games with splittable demands. The facilities correspond to the edges of a directed graph $\mathcal{D} = (V, A)$. Each player i is associated with a source-sink pair $(s_i, t_i) \in V \times V$ and a positive demand d_i that she wishes to route from s_i to t_i. The private cost of each player equals the maximum cost over all facilities she uses with positive demand. The fundamental difference to non-splittable bottleneck congestion games is that each player i is allowed to distribute her demand among all paths connecting s_i and t_i, thus, bottleneck routing games with splittable demands serve as a model of multi-path routing protocols in telecommunication networks, see [6]. They, however, study only existence of PNE. In addition to being more general, our result gives also an alternative and constructive proof for the existence of PNE in bottleneck routing games with splittable demands compared to the involved proof by [6].

Scheduling of malleable jobs. In the scheduling literature jobs are called *malleable* if they can be distributed among multiple machines [17, 8]. In a scheduling game with malleable jobs, each player i controls a job with weight w_i that she distributes over an arbitrary *subset* of allowable machines. The private cost is determined by the makespan, which is a non-decreasing function of the total load of the machine that finishes latest among the chosen machines. To the best of our knowledge, our work investigates for the first time the existence of equilibria (PNE or SE) in such games.

5.2. Existence of SE

As mentioned earlier, using similar arguments as in the proof of Theorem 4.3 one can prove that splittable bottleneck congestion games have the π-LIP. However, the function π may be discontinuous even if cost functions are continuous. To see this, consider the bottleneck congestion game with one player having access to two facilities $X_1 = \{\{f_1\}, \{f_2\}\}$ over which she has to assign a demand of size 1. The facility f_1 has a cost function equal to the load, while facility f_2 has a constant cost function equal to 2. Let $\xi_{1,2}(\epsilon) = \epsilon > 0$ be assigned to facility f_2 and the remaining demand $\xi_{1,1}(\epsilon) = 1 - \epsilon$ be assigned to f_1. Then, for *any* $\epsilon > 0$ we have $\pi(\xi(\epsilon)) = 2$, while $\pi(\xi(0)) = 1$.

To resolve this difficulty, we define the load of facility f under strategy profile ξ as $\ell_f(\xi) = \sum_{i\in N} \xi_{i,f}$ and show that $\nu : \Delta \to \mathbb{R}^m_+ \times \mathbb{R}^m_+, \xi \mapsto (c_f(\xi), \ell_f(\xi))_{f\in F}$ is a continuous pairwise strong vector-valued potential.

Theorem 5.3. Every splittable bottleneck congestion game possesses an SE.

Proof. We show that the function $\nu : \Delta \to \mathbb{R}^m \times \mathbb{R}^m, \xi \mapsto (c_f(\ell_f(\xi), \ell_f(\xi))_{f\in F}$ is a pairwise strong vector-valued potential. Because ν is continuous, Theorem 5.1 gives then the desired result. Let $S \subseteq N$ be an arbitrary coalition and let $(\xi, (\xi'_S, \xi_{-S})) \in I(S)$ be an arbitrary improving move of coalition S. Choose a deviating player $j \in \arg\max_{i\in S} \pi_i(\xi)$ with highest cost before the improving move and one of the facilities $g \in \arg\max_{f\in F_j(\xi)} c_f(\xi)$ at which $\pi_j(\xi)$ is attained. Decompose F into F^+ and F^- defined as $F^+ = \{f \in F : c_f(\xi) \geq c_g(\xi)\}$ and $F^- = \{f \in F : c_f(\xi) < c_g(\xi)\}$.

We first claim that $c_f(\xi'_S, \xi_{-S}) \leq c_f(\xi)$ for all $f \in F^+$. Assume by contradiction that there is $f \in F^+$ with $c_f(\xi'_S, \xi_{-S}) > c_f(\xi)$. By the independence of irrelevant choices and the monotonicity of the cost functions, this implies that there is a player $k \in S$ with $\xi'_{k,f} > 0$. We obtain $\pi_k(\xi'_S, \xi_{-S}) \geq c_f(\xi'_S, \xi_{-S}) > c_f(\xi) \geq c_g(\xi) = \pi_j(\xi) \geq \pi_k(\xi)$, which contradicts that k must improve.

Next we show that $\ell_f(\xi'_S, \xi_{-S}) \leq \ell_f(\xi)$ for all $f \in F^+$ with $c_f(\xi'_S, \xi_{-S}) = c_f(\xi)$. For a contradiction, assume that there is $f \in F^+$ with $\ell_f(\xi'_S, \xi_{-S}) > \ell_f(\xi)$ and $c_f(\xi'_S, \xi_{-S}) = c_f(\xi)$. Again this implies the existence of a player $k \in S$ with $\xi'_{k,f} > 0$. Using $c_f(\xi'_S, \xi_{-S}) = c_f(\xi)$, we obtain the same contradiction to the fact that k improves as before.

Finally, we claim that $c_f(\xi'_S, \xi_{-S}) < c_g(\xi)$ for all $f \in F^-$. To see this, assume that there is $f \in F^-$ with $c_f(\xi'_S, \xi_{-S}) \geq c_g(\xi)$. This again implies that there is a player $k \in S$ with $\xi'_{k,f} > 0$, thus, $\pi_k(\xi'_S, \xi_{-S}) \geq c_f(\xi'_S, \xi_{-S}) \geq c_g(\xi) = \pi_j(\xi) \geq \pi_k(\xi)$, and player k did not improve, contradiction!

To finish the proof, we show that $(c_g(\xi'_S, \xi_{-S}), \ell_g(\xi'_S, \xi_{-S})) < (c_g(\xi), \ell_g(\xi))$. We distinguish two cases. If $\xi'_{j,g} > 0$, we obtain $c_g(\xi'_S, \xi_{-S}) < c_g(\xi)$ using the fact that player j improves. For the second case, let $\xi'_{j,g} = 0$ and assume by contradiction that $(c_g(\xi'_S, \xi_{-S}), \ell_g(\xi'_S, \xi_{-S})) \geq (c_g(\xi), \ell_g(\xi))$. If $c_g(\xi'_S, \xi_{-S}) > c_g(\xi)$, we immediately derive the existence of a player $k \in S$ with $\xi'_{k,g} > 0$. On the other hand, if $c_g(\xi'_S, \xi_{-S}) = c_g(\xi)$ and $\ell_g(\xi'_S, \xi_{-S}) \geq \ell_g(\xi)$, we obtain the existence of $k \in S$ with $\xi'_{k,g} > 0$ using that $\xi'_{j,g} = 0$. In both cases, we calculate $\pi_k(\xi'_S, \xi_{-S}) \geq c_g(\xi'_S, \xi_{-S}) \geq c_g(\xi) = \pi_j(\xi) \geq \pi_k(\xi)$, a contradiction to the fact that k improves.

□

5.3. Existence of Approximate SE

In this section, we relax the continuity assumption on the facility cost functions by assuming that they are only bounded from above. We will prove that bottleneck congestion games with bounded cost functions possess an α-approximate SE for every $\alpha > 0$. An *α-approximate* strong equilibrium is stable only against (coalitional) improving moves that decrease the private cost of every moving player by at least $\alpha > 0$. More formally, we denote by $I^\alpha(S) \subset X \times X$ the set of tuples $(x, (y_S, x_{-S}))$ of α-improving moves for $S \subseteq N$ and define by I^α their union. Then a strategy profile x is an *α-approximate strong equilibrium* if no coalition $\emptyset \neq S \subseteq N$ has an alternative strategy profile y_S such that $\pi_i(x) - \pi_i(y_S, x_{-S}) > \alpha$, for all $i \in S$. We call a function $P : X \to \mathbb{R}$ an *α-generalized strong potential* if $(x, y) \in I^\alpha$ implies $P(x) > P(y)$.

Theorem 5.4. Every splittable bottleneck congestion game with bounded cost functions possesses an α-approximate SE for every $\alpha > 0$.

We prove the theorem by stating a useful lemma.

Lemma 5.5. Let the function $\psi : \Delta \to \mathbb{R}^{mn}_+$ be defined as

$$\psi_{i,f}(\xi) = \begin{cases} c_f(\xi), & \text{if } f \in F_i(\xi) \\ 0, & \text{else} \end{cases} \qquad \text{for all } i \in N, f \in F.$$

Moreover, let $\alpha > 0$ and define $P_M(\xi) = \sum_{f \in F, i \in N} \psi_{i,f}(\xi)^M$, where $M \geq (2\,\psi_{\max}/\alpha + 1)\,\log(n\,m)$ and $\psi_{\max} = \sup_{\xi \in \Delta, f \in F} c_f(\xi)$. Then, P_M is an α-generalized strong potential satisfying $P_M(\xi) - P_M(\xi') \geq (\alpha/2)^M$ for all $(\xi, \xi') \in I^\alpha$.

Proof. We must show that $P_M(\xi) - P_M(\xi'_S, \xi_{-S}) \geq (\alpha/2)^M$ for an arbitrary α-improving move $(\xi, (\xi'_S, \xi_{-S})) \in I^\alpha$. Let $j \in \arg\max_{i \in S} \pi_i(\xi'_S, \xi_{-S})$. We define $\Psi^+ = \{(i, f) \in -S \times F : \psi_{i,f}(\xi) \geq \pi_j(\xi'_S, \xi_{-S})\}$ and $\Psi^- = \{(i, f) \in -S \times F : \psi_{i,f}(\xi) < \pi_j(\xi'_S, \xi_{-S})\}$. We claim that

$$\psi_{i,f}(\xi'_S, \xi_{-S}) \leq \psi_{i,f}(\xi) \qquad \text{for all } (i, f) \in \Psi^+, \tag{4}$$

$$\psi_{i,f}(\xi'_S, \xi_{-S}) \leq \pi_j(\xi'_S, \xi_{-S}) \qquad \text{for all } (i, f) \in \Psi^-. \tag{5}$$

To prove (4), suppose there is $(i, g) \in \Psi^+$ such that $\psi_{i,g}(\xi) < \psi_{i,g}(\xi'_S, \xi_{-S})$. Because of the independence of irrelevant choices and the monotonicity of cost functions there exists $k \in S$ with

$g \in F_k(\xi'_S, \xi_S)$ implying

$$\pi_j(\xi'_S, \xi_{-S}) \leq \psi_{i,g}(\xi) < \psi_{i,g}(\xi'_S, \xi_{-S}) \leq \pi_k(\xi'_S, \xi_{-S}) \leq \pi_j(\xi'_S, \xi_{-S}),$$

which is a contradiction. For proving (5), suppose there is $(i, g) \in \Psi^-$ such that $\psi_{i,g}(\xi'_S, \xi_{-S}) > \pi_j(\xi'_S, \xi_{-S})$. Again, independence of irrelevant choices and monotonicity of cost functions implies that there is $k \in S$ with $g \in F_k(\xi'_S, \xi_{-S})$ giving rise to

$$\pi_k(\xi'_S, \xi_{-S}) \geq \psi_{i,g}(\xi'_S, \xi_{-S}) > \pi_j(\xi'_S, \xi_{-S}) \geq \pi_k(\xi'_S, \xi_{-S}),$$

which is a contradiction. To finish the proof, we observe that $N \times F = \Psi^+ \cup \Psi^- \cup (S \times F)$. Then,

$$\begin{aligned} P_M(\xi) - P_M(\xi'_S, \xi_{-S}) &= \sum_{(i,f) \in \Psi^+ \cup \Psi^- \cup (S \times F)} \psi_{i,f}(\xi)^M - \psi_{i,f}(\xi'_S, \xi_{-S})^M \\ &\geq \sum_{(i,f) \in \Psi^- \cup (S \times F)} \psi_{i,f}(\xi)^M - \psi_{i,f}(\xi'_S, \xi_{-S})^M. \end{aligned}$$

The inequality follows from the first claim. We further derive

$$\begin{aligned} &\sum_{(i,f) \in \Psi^- \cup (S \times F)} \psi_{i,f}(\xi)^M - \psi_{i,f}(\xi'_S, \xi_{-S})^M \\ &\quad \geq \sum_{f \in (S \times F)} \psi_{i,f}(\xi)^{M(\alpha)} - \sum_{(i,f) \in \Psi^- \cup (S \times F)} \psi_{i,f}(\xi'_S, \xi_{-S})^M \\ &\quad \geq (\pi_j(\xi'_S, \xi_{-S}) + \alpha)^M - n\, m\, \pi_j(\xi'_S, \xi_{-S})^M, \end{aligned}$$

where the first inequality follows from the non-negativity of ψ. The second inequality follows from $\pi_j(\xi) \geq \pi_j(\xi'_S, \xi_{-S}) + \alpha$ and the second claim. To this end, we obtain

$$P_M(\xi) - P_M(\xi'_S, \xi_{-S}) \geq (\alpha/2)^M + \big(\pi_j(\xi'_S, \xi_{-S}) + \alpha/2\big)^M - n\, m\, \pi_j(\xi'_S, \xi_{-S})^M \geq (\alpha/2)^M,$$

where the last inequality follows from the choice of M. □

Proof of Theorem 5.4. Fix $\alpha > 0$. Since Δ is compact and P_M (as defined in Lemma 5.5) is bounded, there is a strategy profile z satisfying $P_M(z) \leq \inf_{\xi \in \Delta} P_M(\xi) - \epsilon$ with $0 < \epsilon < (\alpha/2)^M$. We claim that z is an α-approximate SE. Suppose not. Then by Lemma 5.5 there exists a profitable deviation $(z, (\xi'_S, z_{-S})) \in I^\alpha(S)$ with $P_M(z) - P_M(\nu_S, z_{-S}) \geq (\alpha/2)^M > \epsilon$, which contradicts the approximation guarantee of z.

□

Acknowledgements. We thank Michal Feldman for giving an inspiring talk in Berlin about SE and Leah Epstein for pointing out an error in an earlier version of this paper. We are very grateful to the two anonymous referees for their numerous suggestions that helped to improve the presentation of the paper.

References

[1] H. Ackermann, H. Röglin, and B. Vöcking, *Pure Nash equilibria in player-specific and weighted congestion games*, Theor. Comput. Sci. **410**, no. 17 (2009), pp. 1552–1563.

[2] N. Andelman, M. Feldman, and Y. Mansour, *Strong price of anarchy*, Games Econom. Behav. **65**, no. 2 (2009), pp. 289–317.

[3] E. Anshelevich, A. Dasgupta, J. Kleinberg, É. Tardos, T. Wexler, and T. Roughgarden, *The price of stability for network design with fair cost allocation*, in Proc. 45th Annual IEEE Sympos. Foundations Comput. Sci., 2004, pp. 295–304.

[4] E. Anshelevich, A. Dasgupta, J. Kleinberg, É. Tardos, T. Wexler, and T. Roughgarden, *The price of stability for network design with fair cost allocation*, SIAM J. Comput. **38**, no. 4 (2008), pp. 1602–1623.

[5] R. Aumann, *Acceptable points in general cooperative n-person games*, in Contributions to the Theory of Games, 1959.

[6] R. Banner and A. Orda, *Bottleneck routing games in communication networks*, IEEE Journal on Selected Areas in Communications **25**, no. 6 (2007), pp. 1173–1179.

[7] D. Bernheim, B. Peleg, and M. Whinston, *Coalition-proof Nash equilibria i. Concepts*, J. Econom. Theory **42**, no. 1 (1987), pp. 1–12.

[8] T. Carroll and D. Grosu, *Incentive compatible online scheduling of malleable parallel jobs with individual deadlines*, in Proc. 39th Internat. Conf. on Parallel Processing, 2010, pp. 516–524.

[9] S. Chien and A. Sinclair, *Strong and Pareto price of anarchy in congestion games*, in Proc. 36th Internat. Colloquium on Automata, Languages and Programming, S. Albers, A. Marchetti-Spaccamela, Y. Matias, S. Nikoletseas, and W. Thomas, eds., LNCS 5555, 2009.

[10] R. Cole, Y. Dodis, and T. Roughgarden, *Bottleneck links, variable demand, and the tragedy of the commons*, in Proc. 17th Annual ACM-SIAM Sympos. on Discrete Algorithms, 2006, pp. 668–677.

[11] A. Czumaj and B. Vöcking, *Tight bounds for worst-case equilibria*, Transactions on Algorithms **3**, no. 1 (2007).

[12] B. de Keijzer, G. Schäfer, and O. Telelis, *On the inefficiency of equilibria in linear bottleneck congestion games*, in Proc. 3rd Internat. Sympos. Algorithmic Game Theory, 2010, pp. 335–346.

[13] G. Debreu, *Representation of a preference ordering by a numerical function*, in Decision Processes, R. Thrall, C. Coombs, and R. Davies, eds., Wiley, New York, 1954, pp. 159–175.

[14] R. Etkin, A. Parekh, and D. Tse, *Spectrum sharing for unlicensed bands*, IEEE J. Sel. Area Comm. **25**, no. 3 (2007), pp. 517–528.

[15] E. Even-Dar, A. Kesselman, and Y. Mansour, *Convergence time to Nash equilibrium in load balancing*, ACM Trans. Algorithms **3**, no. 3 (2007), p. 32.

[16] A. Fabrikant, C. Papadimitriou, and K. Talwar, *The complexity of pure Nash equilibria*, in Proc. 36th Annual ACM Sympos. Theory Comput., 2004, pp. 604–612.

[17] D. Feitelson and L. Rudolph, *Toward convergence in job schedulers for parallel supercomputers*, in Job Scheduling Strategies for Parallel Processing, D. Feitelson and L. Rudolph, eds., LNCS 1162, Springer, Berlin, Germany, 1996, pp. 1–26.

[18] A. Fiat, H. Kaplan, M. Levy, and S. Olonetsky, *Strong price of anarchy for machine load balancing*, in Proc. 34th Internat. Colloquium on Automata, Languages and Programming, L. Arge, C. Cachin, T. Jurdzinski, and A. Tarlecki, eds., LNCS 4596, 2007, pp. 583–594.

[19] M. Gairing, B. Monien, and K. Tiemann, *Routing (un-) splittable flow in games with player-specific linear latency functions*, in Proc. 33rd Internat. Colloquium on Automata, Languages and Programming, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., LNCS 4051, 2006, pp. 501–512.

[20] T. Harks, M. Hoefer, M. Klimm, and A. Skopalik, *Computing pure Nash and strong equilibria in bottleneck congestion games*, in Algorithms – ESA 2010, Part II, M. de Berg and U. Meyer, eds., LNCS 6347, 2010, pp. 29 – 38.

[21] T. Harks, M. Klimm, and R. Möhring, *Strong Nash equilibria in games with the lexicographical improvement property*, in Proc. 5th Internat. Workshop on Internet and Network Econom., S. Leonardi, ed., LNCS 5929, 2009, pp. 463–470.

[22] R. Holzman and N. Law-Yone, *Strong equilibrium in congestion games*, Games Econom. Behav. **21**, no. 1-2 (1997), pp. 85–101.

[23] R. Johari and J. Tsitsiklis, *Efficiency loss in a network resource allocation game*, Math. Oper. Res. **29**, no. 3 (2004), pp. 407–435.

[24] S. Keshav, *An engineering approach to computer networking: ATM networks, the Internet, and the telephone network*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997.

[25] H. Konishi, M. Le Breton, and S. Weber, *Equivalence of strong and coalition-proof Nash equilibria in games without spillovers*, Econom. Theory **9**, no. 1 (1996), pp. 97–113.

[26] H. KONISHI, M. LE BRETON, AND S. WEBER, *Equilibria in a model with partial rivalry*, J. Econom. Theory **72**, no. 1 (1997), pp. 225–237.

[27] H. KONISHI, M. LE BRETON, AND S. WEBER, *Pure strategy Nash equilibrium in a group formation game with positive externalities*, Games Econom. Behav. **21**, no. 1-2 (1997), pp. 161–182.

[28] E. KOUTSOUPIAS AND C. PAPADIMITRIOU, *Worst-case equilibria*, in Proc. 16th Internat. Sympos. Theoretical Aspects of Comput. Sci., 1999.

[29] M. LIU AND Y. WU, *Spectrum sharing as congestion games*, in Proc. Annual Allerton Conference on Comm., Control, and Comput., 2008.

[30] A. MAS-COLELL, M. D. WHINSTON, AND J. R. GREEN, *Microeconomic Theory*, Oxford University Press, 1995.

[31] M. MAVRONICOLAS, I. MILCHTAICH, B. MONIEN, AND K. TIEMANN, *Congestion games with player-specific constants*, in Proc. 32nd Internat. Sympos. Math. Foundations Comput. Sci., L. Kucera and A. Kucera, eds., LNCS 4708, 2007.

[32] V. MAZALOV, B. MONIEN, F. SCHOPPMANN, AND K. TIEMANN, *Wardrop equilibria and price of stability for bottleneck games with splittable traffic*, in Proc. 2nd Internat. Workshop on Internet and Network Econom., M. Mavronicolas and S. Kontogiannis, eds., LNCS 4286, 2006, pp. 331–342.

[33] I. MILCHTAICH, *Congestion games with player-specific payoff functions*, Games Econom. Behav. **13**, no. 1 (1996), pp. 111 – 124.

[34] D. MONDERER AND L. SHAPLEY, *Potential games*, Games Econom. Behav. **14**, no. 1 (1996), pp. 124 – 143.

[35] H. MOULIN, *The price of anarchy of serial, average and incremental cost sharing*, Economic Theory **36**, no. 3 (2008), pp. 379–405.

[36] J. NASH, *Non-cooperative games*, PhD thesis, Princteon, 1950.

[37] N. NISAN, T. ROUGHGARDEN, É. TARDOS, AND V. VAZIRANI, *Algorithmic Game Theory*, Cambridge University Press, New York, NY, USA, 2007.

[38] M. OSBORNE AND A. RUBINSTEIN, *A Course in Game Theory*, The MIT Press, 1994.

[39] L. QIU, Y. R. YANG, Y. ZHANG, AND S. SHENKER, *On selfish routing in internet-like environments*, IEEE/ACM Trans. Netw. **14**, no. 4 (2006), pp. 725–738.

[40] P. RENY, *On the existence of pure and mixed strategy Nash equilibria in discontinuous games*, Econometrica **67**, no. 5 (1999), pp. 1029–1056.

[41] R. ROSENTHAL, *A class of games possessing pure-strategy Nash equilibria*, Internat. J. Game Theory **2**, no. 1 (1973), pp. 65–67.

[42] O. ROZENFELD AND M. TENNENHOLTZ, *Strong and correlated strong equilibria in monotone congestion games*, in Proc. 2nd Internat. Workshop on Internet and Network Econom., P. Spirakis, M. Mavronicolas, and S. Kontogiannis, eds., LNCS 4286, 2006, pp. 74 – 86.

[43] R. SRIKANT, *The Mathematics of Internet Congestion Control*, Birkhäuser Boston, 2003.

[44] B. VÖCKING, *Selfish load balancing*, in Algorithmic Game Theory, N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, eds., Cambridge University Press, 2007, ch. 20, pp. 517–542.

[45] M. VOORNEVELD, *Congestion games and potentials reconsidered*, Internat. Game Theory Rev. **1**, no. 3 – 4 (1999), pp. 283 – 299.

[46] T. WERTH, H. SPERBER, AND S. KRUMKE, *Efficient computation of equilibria in bottleneck games via game transformation*, tech. report, TU Kaiserslautern, 2011.

Paper 8

Computing Pure Nash and Strong Equilibria in Bottleneck Congestion Games

Tobias Harks, Martin Hoefer, Max Klimm and Alexander Skopalik
Computing Pure and Strong Nash Equilibria in Bottleneck Congestion Games
Mathematical Programming (Series A), accepted s.t. minor revision. Conference version in Proc. of the 18th Annual European Symposium on Algorithms (ESA), 2010 LNCS 6347, Springer, pp. 29-38, 2010
Full version to appear in Mathematical Programming

Abstract. Bottleneck congestion games properly model the properties of many real-world network routing applications. They are known to possess strong equilibria – a strengthening of Nash equilibrium to resilience against coalitional deviations. In this paper, we study the computational complexity of pure Nash and strong equilibria in these games. We provide a generic centralized algorithm to compute strong equilibria, which has polynomial running time for many interesting classes of games such as, e.g., matroid or single-commodity bottleneck congestion games. In addition, we examine the more demanding goal to reach equilibria in polynomial time using natural improvement dynamics. Using unilateral improvement dynamics in matroid games pure Nash equilibria can be reached efficiently. In contrast, computing even a single coalitional improvement move in matroid and single-commodity games is strongly NP-hard. In addition, we establish a variety of hardness results and lower bounds regarding the duration of unilateral and coalitional improvement dynamics. They continue to hold even for convergence to approximate equilibria.

1. Introduction

One of the central challenges in algorithmic game theory is to characterize the computational complexity of equilibria. Results in this direction yield important indicators if game-theoretic solution concepts are plausible outcomes of competitive environments in practice. Probably the most prominent stability concept in (non-cooperative) game theory is the Nash equilibrium – a state, from which no player wants to unilaterally deviate – and its' complexity has been under increased scrutiny for quite some time. A drawback of Nash equilibrium is that in general it exists only in mixed strategies. There are, however, practically important classes of games that allow pure Nash equilibria (PNE), most prominently congestion games. In a congestion game [40], there is a set of resources, and the pure strategies of players are subsets of this set. Each resource has a delay function depending on the *load*, i.e., the number of players that select strategies containing the respective resource. The individual cost for a player in a ordinary congestion game is given by the *sum* over the delays of the resources in his strategy.

Congestion games are an elegant model to study the effects of resource usage and congestion with strategic agents. They have been used frequently to model competitive network routing scenarios [41]. For these games the complexity of exact and approximate PNE is now well-understood. A detailed characterization in terms of, e.g., the structure of strategy spaces [18, 2] or the delay functions [11, 8, 44] has been derived. However, ordinary congestion games have shortcomings, especially as models for the prominent application of routing in computer networks. The throughput of a stream of packets is usually determined by the delay experienced due to available bandwidth or capacity of links. Here the throughput of a player is closely related to the performance of the most congested (bottleneck) link (see, e.g., [30, 6, 12, 39]). A model that captures this aspect more realistically are *bottleneck congestion games*, in which the individual cost of a player is the maximum (instead of sum) of the delays in his strategy. Despite being a more realistic model for network routing, they have not received similar attention in the literature. For classes of non-atomic (with infinitesimally small players) and atomic splittable games (finite number of players with arbitrarily splittable demand) existence of PNE and bounds on the price of anarchy were considered in [12, 35]. For atomic games with unsplittable demand PNE do always exist [6]. In fact, Harks et al. [25] establish the finite improvement property via a lexicographic potential function. Interestingly, they are able to extend these conditions to hold even if *coalitions* of players are allowed to change their strategy in a coordinated way. This implies that bottleneck congestion games do admit even (pure) *strong equilibria (SE)*, a solution concept introduced by Aumann [5]. In an SE, no coalition (of any size) can deviate and strictly decrease the individual cost of each member. Every SE is a PNE, but the converse holds only in special cases (e.g., for singleton games [27]).

SE represent a very robust and appealing stability concept. In general games, however, they are quite rare, which makes the existence guarantee in bottleneck congestion games even more remarkable. For instance, even in dominant strategy games such as the Prisoner's Dilemma there might be no SE. Not surprisingly, for ordinary congestion games with linear aggregation the existence of SE is not guaranteed [31, 27] and, in fact, NP-hard to decide [26]. The existence of PNE and SE in bottleneck congestion games raises a variety of important questions regarding their computational complexity. In which cases can PNE and SE be computed efficiently? As the games have the finite improvement property, another important issue is the duration of natural (coalitional) improvement dynamics. More fundamentally, it is not obvious that even a single such coalitional improving move can be found efficiently. These are the main questions that we address in this paper.

1.1. Our Results

We examine the computational complexity of PNE and SE in bottleneck congestion games. In Section 2 we focus on computing PNE and SE using (centralized) algorithms. Our first main result is a generic algorithm that computes an SE for any bottleneck congestion game. The algorithm iteratively decreases capacities on the resources and relies on a *strategy packing oracle*. The oracle decides if a given set of capacities allows to pack a collection of feasible strategies for all players and outputs a feasible packing if one exists. The running time of the algorithm is essentially determined by the running time of this oracle, i.e., the problem of computing SE can be reduced to solving the strategy packing problem. As a characterization we also prove the reverse direction: The class of set packing problems addressed by strategy packing oracles can be solved efficiently if we can efficiently compute SE in bottleneck congestion games. A slight drawback is that the games constructed in this reduction exhibit a slightly different combinatorial structure than the packing problem. For the case of two players we can circumvent this problem and show polynomial equivalence between packing and SE computation even when we fix the underlying combinatorial structure.

In terms of complexity, we prove a number of upper and lower bounds for specific classes of games. For upper bounds we focus on three classes of games: single-commodity networks, branchings, and matroids (see Section 2.2 for the definition of single-commodity networks and branchings and Section 1.3 for the definition of matroids). Single-commodity network games represent a natural and frequently studied class of network routing. Branchings model a natural scenario when players strive to implement a broadcast from a set of source nodes to all other nodes in the network. Finally, matroid games have been studied prominently in ordinary congestion games [2] and represent a straightforward extension of the popular singleton case. In all these cases, there are strategy packing oracles that can be implemented in polynomial time. Thus, our generic algorithm yields an *efficient algorithm* to compute SE for all these classes of games. For general games, however, we show that the problem of computing an SE is NP-hard, even in two-commodity networks.

In Section 3 we study the duration and complexity of sequential improvement dynamics that converge to PNE and SE. Note that quick convergence (i.e., in a polynomial number of rounds) implies efficient computation. Therefore, we focus particularly on the classes of games, for which we found positive results in terms of computation. In particular, we first observe that for every matroid bottleneck congestion game a variant of best response dynamics presented in [2] called "lazy best response" converge in polynomial time to a PNE. In contrast to this positive result for unilateral dynamics, we show that it is NP-hard to decide if a coalitional improving move exists, even for matroid and single-commodity network games, and even if the deviating coalition is fixed a priori. This highlights an interesting contrast for these two classes of games: While there are polynomial-time algorithms to compute an SE, it is impossible to decide efficiently if a given state is an SE – the decision problem is co-NP-hard.

For more general games, we observe in Section 3.2 that constructions of [44] regarding the hardness of computing PNE in ordinary games can be adjusted to yield similar results for bottleneck games. In particular, in (a) symmetric games with arbitrary delay functions and (b) asymmetric games with bounded-jump delay functions computing a PNE is PLS-complete. In addition, we show that in both cases there exist games and starting states, from which every sequence of improvement moves to a PNE is exponentially long. We extend this result to the case when moves of coalitions of size $\mathcal{O}(n^{1-\epsilon})$ are allowed, for any constant $\epsilon > 0$. In addition, we observe that all of these hardness results generalize to the computation of α-approximate PNE and SE (see Section 1.3 for the definition), for any polynomially bounded factor α.

We conclude the paper in Section 4 by outlining some interesting open problems regarding the convergence to approximate equilibria.

1.2. Related Work

Congestion games (in the ordinary sense) were introduced by Rosenthal [40] and further characterized by Monderer and Shapley [37]. Holzman and Law-Yone [27] studied the existence of SE in congestion games with monotone increasing cost functions. They showed that SE need not exist in such games and gave a structural characterization of the strategy space for symmetric (and quasi-symmetric) congestion games that admit SE. They also introduced the concept of a *strong potential function*: a function on the set of states that decreases for every profitable deviation of a coalition. More recently, Hoefer and Skopalik [26] showed that deciding existence of SE in congestion games is NP-hard, even for two players, and even in games which are both matroid and single-commodity network games. Rozenfeld and Tennenholtz [42] explored the existence of (correlated) SE in congestion games with non-increasing cost functions. Exact and approximate SE have also been considered in other games, e.g., in cost sharing congestion games [17, 3, 33].

A generalization of congestion games has been proposed by Milchtaich [36], where he allows for player-specific delay functions on the resources (see also [34, 24, 1, 26] for subsequent work on (weighted) congestion games with player-specific cost functions). For games with singleton strategies and monotonic delay functions, Milchtaich proves existence of PNE. As shown by Voorneveld et al. [46], the singleton games considered by Milchtaich are equivalent to the games considered by Konishi et al. [31]. This is worth noting as they established existence of SE in such games. Closely related, Andelman et al. [4] considered scheduling games on unrelated machines and proved that the load-lexicographically minimal schedule is an SE. Efficiency of strong equilibria in scheduling games has been studied by Fiat et al. [21], and notions of approximate strong equilibria were analyzed by Feldman and Tamir [19].

Bottleneck congestion games with network structure have been considered by Banner and Orda [6]. They proved existence of PNE in the unsplittable and splittable flow settings. Harks et al. [25] considered a generalization of bottleneck congestion games and proved that these games possess the strong finite improvement property. Epstein et al. [16] characterized network topologies for both ordinary and bottleneck network congestion games such that in the resulting games all PNE are socially optimal. The price of anarchy for PNE in bottleneck congestion games was studied in [9, 14, 29].

Bottleneck routing with *non-atomic* players and elastic demands has been studied by Cole et al. [12], who derived bounds on the price of anarchy. For subsequent work on the price of anarchy in bottleneck routing games with atomic and non-atomic players, we refer to the paper by Mazalov et al. [35].

1.3. Preliminaries

Bottleneck congestion games are strategic games $G = (N, \mathcal{S}, (c_i)_{i\in N})$, where $N = \{1, \dots, n\}$ is the non-empty and finite set of players, $\mathcal{S} = \bigtimes_{i\in N} \mathcal{S}_i$ is the non-empty set of *states* or *strategy profiles*, and $c_i : \mathcal{S} \to \mathbb{N}$ is the *individual cost* function that specifies the cost value of player i for each state $S \in \mathcal{S}$. A game is called finite if $\mathcal{S}$ is finite. For the sake of a clean mathematical definition, we define strategies and costs using the general notion of a congestion model. A tuple $\mathcal{M} = (N, R, \mathcal{S}, (d_r)_{r\in R})$ is called a *congestion model* if $N = \{1, \dots, n\}$ is a non-empty, finite set of players, $R = \{1, \dots, m\}$ is a non-empty, finite set of *resources*, and $\mathcal{S} = \bigtimes_{i\in N} \mathcal{S}_i$ is the set of states or profiles. For each player $i \in N$, the set $\mathcal{S}_i$ is a non-empty, finite set of *pure strategies* $S_i \subseteq R$. Given a state S, we define $\ell_r(S) = |\{i \in N : r \in S_i\}|$ as the number of players using r in S. Every resource $r \in R$ has a *delay function* $d_r : S \to \mathbb{N}$ defined as $d_r(S) = d_r(\ell_r(S))$. In this paper, all delay functions are non-negative and non-decreasing. Delay function d_r satisfies the *β-bounded-jump condition* if $d_r(x + 1) \le \beta \cdot d_r(x)$ for any $x \ge 1$. A congestion model $\mathcal{M}$ is called *matroid congestion model* if for every $i \in N$ there is a matroid $M_i = (R, \mathcal{I}_i)$ such that $\mathcal{S}_i$ equals the set of bases of M_i. We denote by $\mathrm{rk}(\mathcal{M}) = \max_{i\in N} \mathrm{rk}(M_i)$ the rank of the matroid congestion model. (Bottleneck) congestion games corresponding to matroid congestion models will be called *matroid (bottleneck) congestion games*. Matroids exhibit numerous nice properties, some of which are summarized in the Appendix 5. For a comprehensive overview see standard textbooks [32, Chapter 13] and [43, Chapters 39 – 42].

Let $\mathcal{M}$ be a congestion model. The corresponding *bottleneck congestion game* is the strategic game $G(\mathcal{M}) = (N, \mathcal{S}, (c_i)_{i\in N})$ in which c_i is given by $c_i(S) = \max_{r\in S_i} d_r\big(\ell_r(S)\big)$. We drop $\mathcal{M}$ whenever it is clear from context. We define the corresponding *ordinary congestion game* in the same way, the only difference is that $c_i(S) = \sum_{r\in S_i} d_r\big(\ell_r(S)\big)$. For a coalition $C \subseteq N$ we denote by $-C$ its complement and by $\mathcal{S}_C = \bigtimes_{i \in C} \mathcal{S}_i$ the set of states of players in C. A pair $\big(S, (S'_C, S_{-C})\big) \in \mathcal{S} \times \mathcal{S}$ is called an *α-improving move* of coalition C if $c_i(S) > \alpha c_i(S'_C, S_{-C})$ for all $i \in C$ and $\alpha \geq 1$. For $\alpha = 1$ we call $\big(S, (S'_C, S_{-C})\big)$ an *improving move* (or *profitable deviation*). A state S is a k-strong equilibrium (k-SE), if there is no improving move $(S, \cdot)$ for a coalition of size at most k. We say S is a strong equilibrium (SE) if and only if it is an n-SE. Similarly, S is a pure Nash equilibrium (PNE) if and only if it is a 1-SE. We call a state S an α-approximate SE (PNE) if no coalition (single player) has an α-improving move $(S, \cdot)$. We denote by $I(S)$ the set of all possible α-improving moves (S, S') to other states $S' \in \mathcal{S}$. We call a sequence of states $(S^0, S^1, \dots)$ an *improvement path* if every tuple $(S^k, S^{k+1}) \in I(S^k)$ for all $k = 0, 1, 2, \dots$. Intuitively, an improvement path is a path in a so-called *state graph* $\mathcal{G}(G)$ derived from G, where every state $S \in \mathcal{S}$ corresponds to a node in $\mathcal{G}(G)$ and there is a directed edge (S, S') if and only if $(S, S') \in I(S)$.

2. Computing Strong Equilibria

In this section, we investigate the complexity of computing a SE in bottleneck congestion games. We first present a generic algorithm that computes a SE for an arbitrary bottleneck congestion game. It uses an oracle that solves a strategy packing problem (see Definition 2.1), which we term *strategy packing oracle*. For games in which the strategy packing oracle can be implemented in polynomial time, we obtain an efficient algorithm computing a SE. We then examine games for which this is the case. In general, however, we prove that computing a SE is NP-hard, even for two-commodity bottleneck congestion games.

2.1. The Dual Greedy

The general approach of our algorithm is to introduce upper bounds u_r (capacities) on each resource r. The idea is to iteratively reduce upper bounds of costly resources as long as the residual capacities admit a feasible strategy packing, see Definition 2.1 below. Our algorithm can be interpreted as a dual greedy, or *worst out algorithm* as studied, e.g., in the field of network optimization, see Schrijver [43].

Algorithm 1 Dual Greedy, the strategy packing oracle is denoted by $\mathfrak{O}$.

Input: Bottleneck congestion game $G(\mathcal{M})$ to the model $\mathcal{M} = (N, R, \mathcal{S}, d)$
Output: SE of G

```
set N' = N, u_r = n, l_r = 0 for all r ∈ R, and S' = 𝔒(R, S_{N'}, u_r)
while {r ∈ R : u_r > 0} ≠ ∅ do
  choose r' ∈ arg max_{r∈R:u_r>0} {d_r(u_r + l_r)}
  u_{r'} := u_{r'} − 1
  if 𝔒(R, S_{N'}, u_r) = ∅ then
    u_{r'} := u_{r'} + 1
    for all j ∈ N' with r' ∈ S'_j do
      S_j := S'_j
      set l_r := l_r + 1, u_r := u_r − 1 for all r ∈ S'_j
      N' := N' \ {j}
    end for
  end if
  S' = 𝔒(R, S_{N'}, u_r)
end while
return S
```

Definition 2.1 (Strategy packing oracle).
INPUT: Finite set of resources R with upper bounds $(u_r)_{r\in R}$, and n collections $\mathcal{S}_1,\dots,\mathcal{S}_n \subseteq 2^R$ given implicitly by a certain combinatorial property.
OUTPUT: Sets $S_1 \in \mathcal{S}_1,\dots,S_n \in \mathcal{S}_n$ such that $|i \in \{1,\dots,n\} : r \in S_i| \leq u_r$ for all $r \in R$, or the information, that no such sets exist.

More specifically, when the algorithm starts, no strategy has been assigned to any player and each resource can be used by n players, thus, $u_r = n$. If r is used by n players, its cost equals $d_r(n)$. The algorithm now iteratively reduces the maximum resource cost by picking a resource r' with maximum delay $d_r(u_r)$ and $u_r > 0$. The number of players allowed on r' is reduced by one and the strategy packing oracle checks, if there is a feasible strategy profile obeying the capacity constraints. If the strategy packing oracle outputs such a feasible state S, the algorithm reiterates by choosing a (possibly different) resource that has currently maximum delay. If the strategy packing oracle returns $\emptyset$ after the capacity of some $r' \in R$ was reduced to $u_{r'} - 1$, we fix the strategies of those $u_{r'}$ many players that used r' in the state the strategy packing oracle computed in the previous iteration and decrease the bounds u_r of all resources used in the strategies accordingly. This ensures that r' is *frozen*, i.e., there is no residual capacity on r' for allocating this resource in future iterations of the algorithm. The algorithm terminates after at most $n \cdot m$ calls of the oracle. For a formal description of the algorithm see Algorithm 1.

Theorem 2.2. Dual Greedy computes a SE.

Proof. Let S denote the output of the algorithm. In addition, we denote by N_k, $k = 1\dots,K$, the sets of players whose strategies are determined after the strategy packing oracle (denoted by $\mathfrak{O}$) returned $\emptyset$ for the k-th time. Clearly, $c_i(S) \leq c_j(S)$ for all $i \in N_k$, $j \in N_l$, with $k \geq l$. We will show by complete induction over k that the players in $N_1 \cup \dots \cup N_k$ will not participate in any improving move of any coalition.

We start with the case $k = 1$. Let $(u_r)_{r\in R}$ be the vector of capacities in the algorithm after the strategy packing oracle returned $\emptyset$ in line 5 for the first time and $u_{r'}$ is updated in line 6.

Suppose there is a coalition $C \subseteq N$ with $C \cap N_1 \neq \emptyset$ that deviates profitably from S to $T = (S'_C, S_{-C})$. We distinguish two cases.

Case 1: $\ell_r(T) \leq u_r$ for all $r \in R$. Let $\tilde{u}_r = u_r - 1,$ if $r = r'$ and $\tilde{u}_r = u_r,$ else. Since $\mathfrak{O}(R, \mathcal{S}, \tilde{u}) = \emptyset$, at least $|N_1|$ players use r' in T. Using $d_{r'}(T) \geq d_r(S)$ for all $r \in R$, we obtain a contradiction to the fact that every member of C must strictly improve.

Case 2: There is $\tilde{r} \in R$ such that $\ell_{\tilde{r}}(T) > u_r$. Using that Dual Greedy iteratively reduces the capacity of those resources with maximum delay (line 3), we derive that $d_{\tilde{r}}(T) \geq d_r(S)$ for all $r \in R$. Using $\ell_{\tilde{r}}(T) > u_r$, there is at least one player $i \in C$ with $\tilde{r} \in S'_i$, hence, this player does not strictly improve.

For the induction step $k \to k+1$, suppose the players in $N_1 \cup \dots \cup N_k$ stick to their strategies and consider the players in N_{k+1}. As the strategies of the players in $N_1 \cup \dots \cup N_k$ are fixed, the same arguments as above imply that no subset of N_{k+1} will participate in a profitable deviation from S. ☐

It is worth noting that the dual greedy algorithm applies to arbitrary strategy spaces. If the strategy packing problem can be solved in polynomial time, this algorithm computes a SE in polynomial time. Hence, the problem of computing a SE is polynomial-time reducible to the strategy packing problem. For general bottleneck congestion games the converse is also true.

Theorem 2.3. The strategy packing problem is polynomial-time reducible to the problem of computing a SE in a bottleneck congestion game.

Proof. Given an instance of the strategy packing problem Π we construct a bottleneck congestion game G_Π. Let Π be given as set of resources R with upper bounds $(u_r)_{r\in R}$, and n collections

$\mathcal{S}_1, \dots, \mathcal{S}_n \subseteq 2^R$. The game G_Π consists of the resources $R \cup \{r_1, \dots, r_n\}$ and the players $1, \dots, n+1$. The set of strategies of player $i \in \{1, \dots, n\}$ is $\{S_i \cup \{r_i\} \mid S_i \in \mathcal{S}_i\}$. Player $n+1$ has the strategies R and $\{r_1, \dots, r_n\}$. For each resource $r \in R$ the delay is 0 if used by at most $u_r + 1$ players and 2 otherwise. For each resource $r \in \{r_1, \dots, r_n\}$ the delay is 0 if used by at most one player and 1 otherwise.

If a strategy profile of players $1, \dots, n$ violates an upper bound u_r on a resource $r \in R$, player $n+1$ has delay of 2 if he plays strategy R. If he plays $\{r_1, \dots, r_n\}$, he and all other players have delay of 1. Hence, if there is a feasible strategy packing, every SE of the game yields delay 0 for every player. Otherwise, every SE yields delay 1 for every player. Therefore, the state of the players $1, \dots, n$ in a SE of G_Π corresponds to a solution for the strategy packing problem Π, if such a solution exists. On the other hand, if there is no solution for Π, every player in every SE in G_Π has delay of 1. □

Note that while the previous theorem establishes a reduction in general, the game G_Π constructed from the instance Π of the packing problem has a different combinatorial structure than Π. More concretely, G_Π is based on a larger set of resources and different strategy sets than the ones used in Π. The next theorem shows that for games with two players, we can obtain a stronger equivalence result without changing the underlying combinatorial structure. It remains an open problem to extend this stronger result to games with an arbitrary number of players and more general packing problems.

Theorem 2.4. Let R be a finite set and $\mathcal{S}_1, \mathcal{S}_2 \subseteq 2^R$ two collections of subsets of R. Then the following two problems are polynomially equivalent:

(1) Compute a SE of $G(\mathcal{M})$ for the congestion model $\mathcal{M} = (\{1,2\}, R, \mathcal{S}, (d_r)_{r\in R})$ where $(d_r)_{r\in R}$ is an arbitrary set of non-decreasing delay functions.
(2) Compute $S_1 \in \mathcal{S}_1, S_2 \in \mathcal{S}_2$ such that $|i \in \{1,2\} : r \in S_i\}| \leq u_r$ or decide that no such strategies exist where $u_r \in \{1,2\}$ for all $r \in R$ is arbitrary.

Proof. "2. $\to$ 1.": As the dual greedy algorithm computes a SE using polynomial many calls of the strategy packing oracle the first problem is polynomially reducible to the second one.

"1. $\to$ 2.": Suppose we are given an instance $\big(R, \mathcal{S}, (u_r)_{r\in R}\big)$ of the second problem. We regard the congestion model $\mathcal{M} = \big(\{1,2\}, R, \mathcal{S}, (d_r)_{r\in R}\big)$ where d_r is defined as

$$d_r(\ell) = \begin{cases} 0, & \text{if } \ell \leq u_r \\ 1, & \text{otherwise.} \end{cases}$$

Now, let G be a corresponding bottleneck congestion game and let S^* be a SE of G. We claim that $c_1(S^*) = c_2(S^*) = 0$ and S^* is a solution of the strategy packing problem if such a solution exists, and $c_1(S^*) = c_2(S^*) = 1$, otherwise. At first, we remark that either $c_1(S^*) = c_2(S^*) = 1$ or $c_1(S^*) = c_2(S^*) = 0$. So suppose that $c_1(S^*) = c_2(S^*) = 1$ and assume for a contradiction that there is solution $S' = (S'_1, S'_2)$ to the strategy packing problem. Then, by the definition of d_r we get that $c_1(S') = c_2(S') = 0$ and, thus, the deviation from S^* to S' is profitable both for player 1 and 2. This is a contradiction to the fact that S^* is a SE. Hence, no such state S' exists.

For the other direction, it is easy to check that $c_1(S^*) = c_2(S^*) = 0$ only if the strategies S_1^* and S_2^* obey the upper bounds on each resource. □

2.2. Complexity of Strategy Packing

In the previous section we have characterized the computation of SE in terms or a set packing problem. In this section, we examine the computational complexity of strategy packing and SE computation. In particular, we consider three classes of games, in which strategy packing can be done efficiently. For the general case, we show that computation becomes NP-hard. A more detailed characterization as to which structural properties are crucial for efficient strategy packing or hardness is an interesting avenue for future work.

Our first result is for matroid games, which represent a natural extension of singleton games. In a singleton game we have $|S_i| = 1$ for every strategy $S_i \in \mathcal{S}_i$ of every player i. In such games, SE are exactly the PNE and computation of SE can trivially be done in polynomial time. Also, strategy packing reduces to perfect matching in a bipartite graph.[1] For matroid games, we have to resort to more advanced algorithmic techniques.

Theorem 2.5. The strategy packing problem can be solved in polynomial time for matroid bottleneck congestion games where the strategy set of player i equals the set of bases of a matroid $M_i = (R, \mathcal{I}_i)$ given by a polynomial independence oracle.

Proof. For each matroid $M_i = (R, \mathcal{I}_i)$, we construct a matroid $M_i' = (R', \mathcal{I}_i')$ as follows. For each resource $r \in R$, we introduce u_r resources $r^1, \dots, r^{u_r}$ to R'. We say that r is the *representative* of $r^1, \dots, r^{u_r}$. Then, a set $I' \subset R'$ is independent in M_i' if the set I that arises from I' by replacing resources by their representatives is independent in M_i. This construction gives rise to a polynomial independence oracle for M_i'.

Now, we regard the matroid union $M' = M_1' \vee \cdots \vee M_n'$, see Definition 5.1 in the Appendix, which again is a matroid. Using the algorithm proposed by Cunningham [13] we can compute a maximum-size set $\mathcal{B}$ in $I_1' \vee \cdots \vee I_n'$ in time polynomial in n, m, $rk(\mathcal{M})$, and the maximum complexity of the n independence oracles.

Clearly, if $|\mathcal{B}| < \sum_{i \in N} \mathrm{rk}(M_i)$, there is no feasible packing of the bases of $M_1, \dots, M_n$. If, in contrast, $|\mathcal{B}| = \sum_{i \in N} \mathrm{rk}(M_i)$, we obtain the corresponding strategies $(S_1, \dots, S_n)$ using the algorithm. $\square$

Let us now consider strategy spaces defined as *a-arborescences*, which are in general not matroids. Let $\mathcal{D} = (V, R)$ be a directed graph with $|R| = m$. For a distinguished node in $a \in V$, we define an *a-arborescence* as a directed spanning tree, where a has in-degree zero and every other vertex has in-degree one. In this case, we can regard $a \in V$ as a common source, and each player strives to make a broadcast with source a by allocating a tree.

Theorem 2.6. The strategy packing problem can be solved in time $\mathcal{O}(m^2\, n^2)$ for a-arborescence games in which the set of strategies of each player equals the set of a-arborescences in a directed graph $\mathcal{D} = (V, R)$.

Proof. The problem of finding k disjoint a-arborescences in G can be solved in polynomial time $\mathcal{O}(m^2\, k^2)$, see Gabow [23, Theorem 3.1]. Introducing u_r copies for each edge $r \in R$, the problem of finding admissible strategies in the original problem is equivalent to finding n disjoint a-arborescences. $\square$

Recently, the polynomial packing algorithm for a-arborescences has been extended to branchings. Formally, we are given for each player i a root set $R_i \subseteq V$ and a convex[2] set $U_i \subseteq V$ with $R_i \subseteq U_i$. For any vector of capacities $(u_r)_{r \in R}$, the polynomial algorithm of Bérczi and Frank [7] computes for every player a branching which is rooted in R_i and spans U_i, that is, the in-degree of every vertex $v \in R_i$ is zero and the in-degree of every vertex $v \in U_i \setminus R_i$ is one, such that the capacity restriction of every edge is satisfied. This more general framework allows to model situations in which the players wish to broadcast from multiple broadcasting stations, where the broadcasts need not cover all vertices. It is worth mentioning that the convexity of U_i is

[1] We construct the graph as follows. The first partition contains a node for each player, the second partition contains u_r nodes for each $r \in R$. The node of player i is connected to all nodes of each $r \in \mathcal{S}_i$.

[2] In this context, a subset of vertices $U \subseteq V$ is called *convex* if there is no vertex $v \in V \setminus U$ such that there is both a directed path from v to some vertex $u \in U$ and a directed path from some node $u' \in U$ to v.

necessary for efficient computation, because otherwise, the corresponding decision problem turns out to be NP-complete.

When we turn to single-commodity networks, then efficient computation of a SE is possible using well-known flow algorithms to implement the oracle. For more general cases with two commodities, however, a variety of problems concerning SE become NP-hard by a simple construction.

Theorem 2.7. The strategy packing problem can be solved in time $\mathcal{O}(m^3)$ for single-commodity bottleneck congestion games.

Proof. Assigning a capacity of u_r to each edge and using the algorithm of Edmonds and Karp we obtain a maximum flow within $\mathcal{O}(m^3)$. Clearly, if the value of the flow is smaller than n, no admissible strategies exist and we can return $\emptyset$. If the flow is n or larger we can decompose it in at least n unit flows and return n of them. □

Theorem 2.8. In two-commodity network bottleneck games it is strongly NP-hard to (1) compute a SE, (2) decide for a given state whether any coalition has an improving move, and (3) decide for a given state and a given coalition if it has an improving move.

Proof. We reduce from the 2 Directed Arc-Disjoint Paths (2DADP) problem, which is strongly NP-hard, see Fortune et al. [22]. The problem is to decide if for a given directed graph $\mathcal{D} = (V, A)$ and two node pairs (s_1, t_1), (s_2, t_2) there exist two arc-disjoint (s_1, t_1)- and (s_2, t_2)-paths. For the reduction, we define a corresponding two-commodity bottleneck game by introducing non-decreasing delay functions on every arc r by $d_r(x) = 0$, if $x \leq 1$ and 1, else. We associate every commodity with a player. For proving (1), we observe that 2DADP is a Yes-instance if and only if every SE provides a payoff of zero to every player. For proving (2) and (3), we simply construct a solution in which the strategies for both players are not arc-disjoint. □

3. Convergence of Improvement Dynamics

In the previous section, we have outlined some prominent classes of games, for which SE can be computed in polynomial time. Furthermore, it is known [25] that sequential improvement dynamics converge to PNE and SE. In this section, we consider the duration of improvement dynamics in these games. As polynomial-time convergence implies polynomial-time computation, we first focus on classes of games, in which we have shown efficient computation, i.e., matroid and single-commodity network games. For matroid games we show polynomial-time convergence to a PNE using unilateral improving moves. For the convergence to SE we have to consider coalitional improving moves, but we show that deciding if such a move exists is NP-hard even in matroid games or single-commodity network games. This implies that even in these specialized classes of games with efficient computation of a SE, recognition of a state as a SE is co-NP-hard.

In more general games, hardness of recognition is not the only source of difficulty. In particular, we prove that in general games even computing an α-approximate PNE is PLS-hard. There are games and starting states, for which every sequence of unilateral improving moves is exponentially long. Perhaps surprisingly, this also holds when we consider coalitional improving moves of coalitions of size $O(n^{1-\epsilon})$, for any constant $\epsilon > 0$.

3.1. Matroid and Single-Commodity Network Games

We first observe that bottleneck congestion games can be transformed into ordinary congestion games while preserving useful properties regarding the convergence to PNE. This allows to show fast convergence to PNE in matroid bottleneck games and mirrors a prominent result for ordinary matroid games [2].

3.1.1. Convergence to Pure Nash Equilibria

The following lemma establishes a connection between bottleneck and ordinary congestion games. For a bottleneck congestion game G we denote by G^{sum} the ordinary congestion game with the same congestion model as G except that we choose $d'_r(S) = m^{d_r(\cdot)}, r \in R$.

Lemma 3.1. Every PNE for G^{sum} is a PNE for G.

Proof. Suppose S is a PNE for G^{sum} but not for G. Thus, there is player $i \in N$ and strategy $S'_i \in \mathcal{S}_i$, such that $\max_{r\in S_i} d_r(\ell_r(S)) > \max_{r\in S'_i} d_r(\ell_r(S'_i, S_{-i}))$. We define the value $\bar{d} := \max_{r\in S'_i} d_r(\ell_r(S'_i, S_{-i}))$. This implies $\max_{r\in S_i} d_r(\ell_r(S)) \geq \bar{d} + 1$. We obtain a contradiction by observing

$$\sum_{r\in S_i} d'_r(\ell_r(S)) \geq \max_{r\in S_i} d'_r(\ell_r(S)) \geq m^{\bar{d}+1} > (m-1)\, m^{\bar{d}} \geq \sum_{r\in S'_i} d'_r(\ell_r(S'_i, S_{-i})).$$

□

We analyze the lazy best response dynamics considered for ordinary matroid congestion games presented in [2]. Note that in matroid games, a player always picks as strategy a basis of a matroid. A *lazy* best response means that a player only exchanges a minimum number of resources that is needed to arrive at a basis representing a best response strategy (for details see [2]). Our analysis here is quite simple and does not explicitly rely on these details. In particular, we transform the game to an ordinary game as outlined in Lemma 3.1. Then we use the lazy best response dynamics in the ordinary game and the convergence result of [2] as a "black box" with the slight adjustment that we only execute moves yielding a strict improvement in the bottleneck resource of the moving player. This allows to establish the following result.

Theorem 3.2. Let G be a matroid bottleneck congestion game. Then the lazy best response dynamics converges to a PNE in at most $n^2 \cdot m \cdot \text{rk}(\mathcal{M})$ steps.

Proof. We consider the lazy best response dynamics in the corresponding game G^{sum}. In addition, we suppose that a player accepts a deviation only if his bottleneck value is strictly reduced. This might lead to even earlier termination of the dynamics. Thus, the duration is still bounded from above by $n^2 \cdot m \cdot \text{rk}(\mathcal{M})$ moves as shown in [2]. □

3.1.2. Convergence to Strong Equilibria

For matroid bottleneck congestion games we have shown above that there are polynomially long sequences of unilateral improving moves to a PNE from every starting state. While previous work [25] also establishes convergence to SE for every sequence of coalitional improving moves, it may already be hard to find one such move. In fact, we show that even an α-improving move can be strongly NP-hard to find, for any polynomial-time computable α, even if strategy spaces have simple matroid structures. This implies that deciding whether a given state is an α-approximate SE is strongly co-NP-hard – even if all delay functions satisfy the β-bounded-jump condition, for any $\beta > \alpha$.

Theorem 3.3. In matroid bottleneck congestion games it is strongly NP-hard to decide for a given state S if there is some coalition $C \subseteq N$ that has an α-improving move, for every polynomial-time computable α.

Proof. We reduce from SET PACKING. An instance of SET PACKING is given by a set of elements E and a set $\mathcal{U}$ of sets $U \subseteq E$, and a number k. The goal is to decide if there are k mutually disjoint sets in $\mathcal{U}$. Given an instance of SET PACKING we show how to construct a matroid game

G and a state S such that there is an improving move for some coalition of players C if and only if the instance of Set Packing has a solution.

The game will include $|N| = 1 + |\mathcal{U}| + |E| + \sum_{U\in\mathcal{U}} |U|$ many players. First, we introduce a master player p_1, which has two possible strategies. He can either pick a *coordination resource* r_c or the *trigger resource* r_t. For each set $U \in \mathcal{U}$, there is a *set player* p_U. Player p_U can choose either r_t or a *set resource* r_U. For each set U and each element $e \in U$, there is an *inclusion player* $p_{U,e}$. Player $p_{U,e}$ can use either the set resource r_U or an *element resource* r_e. Finally, for each element e, there is an *element player* p_e that has strategies $\{r_c, r_e\}$ and $\{r_c, r_a\}$ for some absorbing resource r_a.

The state S is given as follows. Player p_1 is on r_c, all set players use r_t, all inclusion players the corresponding set resources r_U, and all element players the strategies $\{r_c, r_e\}$. The coordination resource r_c is a bottleneck for the master player and all element players. The delays are $d_{r_c}(x) = \alpha + 1$, if $x > |E|$ and 1, otherwise. The trigger resource has delay $d_{r_t}(x) = 1$, if $x \leq |\mathcal{U}| - k + 1$, and $\alpha + 1$, otherwise. For the set resources r_U the delay is $d_{r_U}(x) = 1$, if $x \leq 1$ and $\alpha+1$, otherwise. Finally, for the element resources the delay is $d_{r_e}(x) = 1$ if $x \leq 1$ and $\alpha+1$ otherwise.

Suppose that the underlying Set Packing instance is a Yes-instance, then an α-improving move is as follows. The master player moves to r_t, the k set players corresponding to a solution choose their set resources, the respective inclusion players move to the element resources, and all element players move to r_a. The delay of r_c reduces from $\alpha + 1$ to 1, and the delay of r_t reduces from $\alpha+1$ to 1. Thus, the master player, all set players, and all element players improve their bottleneck by a factor of $\alpha + 1$. The migrating inclusion players do not interfere with each other on the element resources. Thus, they also improve the delay of their bottleneck resource by factor $\alpha + 1$, and we have constructed an α-improving move for the coalition of all migrating players, all set players, and all element players.

Suppose that the underlying Set Packing instance is a No-instance. For contradiction, assume that there is a coalition C that has an α-improving move. Consider any player $p \in C$. We will show that for any player $p \neq p_1$, i.e., any set, inclusion, or element player, $p_1 \in C$ is a prerequisite for achieving any strict improvement. We first note that the master player can never strictly improve without changing his strategy, because all element players will always use r_c in their strategy. A move from r_c to r_t is an improvement if and only if at least k set players drop r_t. These players must switch to the corresponding resources. However, for a set player p_M such a move is an improvement if and only if all inclusion players on r_U drop this resource from their strategy. These inclusion players must switch to the element resources. An inclusion player $p_{U,e}$ improves by such a move if and only if the element player drops the resource and $p_{U,e}$ is the only inclusion player moving to r_e. This implies that the moving set players must correspond to sets that are mutually disjoint. Finally, the element players move from r_e to r_a with delay $d_{r_a} = 0$, and this is an improvement if and only if the master player moves away from r_c. This last argument establishes that $p \in C$ implies $p_1 \in C$.

However, if the master player $p_1 \in C$, then we again follow the chain of reasoning above and see that the players corresponding to at least k mutually disjoint sets must move and therefore be in C. This is a contradiction to having a No-instance.

Finally, we can add the resource r_a to every strategy of the master, set, and inclusion players. In this way, the combinatorial structure of all strategy spaces is the same – a partition matroid M with $rk(M) = 2$ and partitions of size 1 and 2 – only the mapping to resources is different for each player. □

The previous theorem shows hardness of the problem of finding a suitable coalition and a corresponding improving move. Even if we specify the coalition in advance and search only for strategies corresponding to an improving move, the problem remains strongly NP-hard.

Corollary 3.4. In matroid bottleneck congestion games it is strongly NP-hard to decide for a given state S and a given coalition $C \subseteq N$ if there is an α-improving move for C, for every polynomial-time computable α.

Proof. We will show this corollary using the games constructed in the previous proof by fixing the coalition $C = N$. Consider the construction in the previous proof. The coalition described above that has an improving move for a Yes-instance consists of the master player, all set players, all element players and the inclusion players that correspond to the sets of the solution to SET PACKING. However, the inclusion players are only needed to transfer the chain of dependencies to the element players. We can set the strategy space of player $p_{U,e}$ to $\{r_h, r_l\} \times \{r_U, r_e\}$. Here r_h and r_l are two resources with delays $d_{r_h} = \alpha + 1$ and $d_{r_l} = 0$. In S we assign the inclusion players to strategies $\{r_h, r_U\}$. Then an improving move for the inclusion players that remain on r_U is to exchange r_h by r_l. Thus, the problem of finding an arbitrary coalition with an improving move becomes trivial. However, we strive to obtain an improving move for $C = N$, and this must generate improvements for the master player and the set players. Thus, we still must reassign some inclusion players from the resources r_U to the element resources r_e. Here we need to resolve conflicts as before, because otherwise inclusion players end up with a delay of $\alpha + 1$ on r_e and do not improve. Following the previous reasoning we have an α-improving move if and only if the underlying SET PACKING instance is solvable. Finally, by appropriately adding dummy resources, we can again ensure that the combinatorial structure of all strategy spaces is the same. $\square$

We can adjust the previous two hardness results on matroid games to hold also for single-commodity network games.

Theorem 3.5. In single-commodity network bottleneck congestion games it is strongly NP-hard to decide for a given state S (1) if there is some coalition $C \subseteq N$ that has an α-improving move, and (2) if a given coalition $C \subseteq N$ has an α-improving move, for every polynomial-time computable α.

Proof. We transform the construction of Theorem 3.3 into a symmetric network bottleneck congestion game, see Fig. 1 for an example. First, we introduce for each resource r_c, r_t, r_U for all $U \in \mathcal{U}$ and r_e for all $e \in E$ an edge with the corresponding delay function as before. Additionally, we identify players and their strategies by routing them through a set of gadgets composed of edges, which have capacities implemented by cost functions that are 1 up to a capacity bound and $\alpha + 10$ above.

The first gadget is to separate the players into groups. An edge with capacity 1 identifies the master player, an edge with capacity $|\mathcal{U}|$ the set players, an edge with capacity $\sum_{U \in \mathcal{U}} |U|$ the inclusion players, and an edge with capacity $|E|$ the element players. The set and inclusion players are then further divided into their particular identities by edges of capacity 1. The element players route all over r_c. In addition, the master player has the alternative to route over r_c or r_t. After the players have passed r_c they again split into specific element players using edges of capacity 1. One player is allowed to route directly to the source t. This is meant to be the master player, but it does not hurt our argument if this is not the case.

After the players have routed through the capacitated gadgets, they can be assumed to reach an identification point (indicated by gray nodes in Fig. 1) and obtain an identity. Then they decide on a strategy from the previous game by routing over one of two allowed paths. In particular, we can allow the set players to route either over r_t or their r_U, the inclusion players over r_U or r_e, and the element players over r_e or directly to the sink t.

We can create the corresponding state S as before by assigning the master player to route over r_c directly to the sink, the set players over r_t, the inclusion players over r_U and the element players over r_e. This assignment is such that every player receives one identity (i.e., routes over

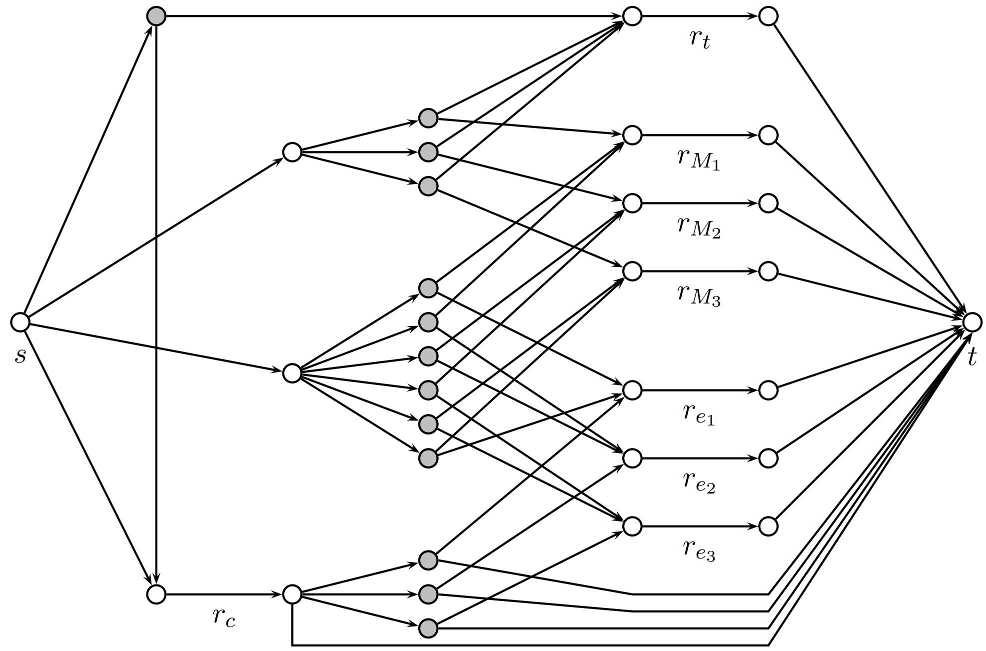


Figure 1. Network construction for a Set Packing instance with $\mathcal{U} = \{\{e_1, e_2\}, \{e_2, e_3\}, \{e_3, e_1\}\}$. Gray nodes serve as identification for players as discussed in the text.

exactly one gray node) and every identity is taken (i.e., every gray node is reached by exactly one player). This property also holds for every improving move – with the exception of one element player, who might route directly from r_c to the sink, but as noted before this does not hurt the argument.

Our network structure allows to reconstruct the reasoning as before. Any improving move must include the master player, which improves if and only if he moves together with players corresponding to a solution to the Set Packing instance. Note that even by switching player identities, we cannot create an improving move when the underlying Set Packing instance is unsolvable. This proves the first part of the theorem.

For the second part, we use the same adjustment as in Corollary 3.4 to ensure that inclusion players can always improve. Directly before the middle fan out (see Figure 1) that results in identification of inclusion players we simply insert a small gadget with 2 parallel edges r_l and r_h. In this way, all inclusion players must route over one of r_l or r_h and one of their corresponding r_U or r_e. This resembles the strategy choices in the matroid game and yields hardness of computing an improving move for the coalition $C = N$. This proves the theorem. □

3.2. General Games and Approximation

The results of the previous sections imply hardness of the computation of SE or coalitional deviations, even in network games. Therefore, when considering general games we here restrict ourselves mostly to unilateral improving moves and PNE. Unfortunately, even in this restricted case the hardness results for ordinary congestion games in Skopalik and Vöcking [44] immediately imply identical results for bottleneck congestion games. The main result of [44] shows that computing an approximate PNE is PLS-hard. The proof is a reduction from CircuitFlip, a prominent PLS-complete problem for feedback-free Boolean circuits. The problem is to find a local optimum, i.e., a bit string x such that the output resulting from applying the circuit to x cannot be improved lexicographically by switching a single bit in x [28].

We can regard the resulting congestion game in the reduction of [44] as a bottleneck congestion game. It is straightforward to adjust all arguments in the proof of [44] to remain valid for bottleneck congestion games. This simple fact has been observed before, e.g. in [45], and we include it here for completeness. We provide some details on the construction of the class $G(n)$ of games used in the reduction in the Appendix 6. A standard transformation [18] immediately yields the same result even for symmetric games, in which $\mathcal{S}_i = \mathcal{S}_j$ for all $i, j \in N$.

Corollary 3.6. Finding an α-approximate PNE in a symmetric bottleneck congestion game with positive and increasing delay functions is PLS-complete, for every polynomial-time computable $\alpha > 1$.

A second result in [44] reveals that sequences of α-improving moves do not reach an α-approximate PNE quickly – even if all delay functions satisfy the β-bounded-jump condition with a constant β. Again, the proof remains valid if one regards the game as an asymmetric bottleneck congestion game. This yields the following corollary.

Corollary 3.7. For every $\alpha > 2$, there is a $\beta > 1$ such that, for every $n \in \mathbb{N}$, there is a bottleneck congestion game $G(n)$ and a state S with the following properties. The description length of $G(n)$ is polynomial in n. The length of every sequence of α-improving moves leading from S to an α-approximate equilibrium is exponential in n. All delay functions of $G(n)$ satisfy the β-bounded-jump condition.

Using the same trick as before to convert an asymmetric game in a symmetric one yields a similar result for symmetric games. However, we must sacrifice the β-bounded-jump condition of the delay functions, for every β polynomial in n.

Despite the fact that (coalitional) improving moves are NP-hard to compute, one might hope that the state graph becomes sufficiently dense such that it allows short improvement paths. Unfortunately, we can show that this is not true, even if we consider all improving moves of coalitions of size up to $\mathcal{O}(n^{1-\epsilon})$, for any constant $\epsilon > 0$. Again, the same result holds for symmetric games when sacrificing the bounded-jump condition.

Theorem 3.8. For every $\alpha > 2$, there is a $\beta > 1$ such that, for every $n \in \mathbb{N}$ and for every $k \in \mathbb{N}$, there is a bottleneck congestion game $G(n, k)$ and a state S with the following properties. The description length of $G(n, k)$ is polynomial in n and k. The length of every sequence of α-improving moves of coalitions of size at most k leading from S' to an α-approximate k-SE is exponential in n. All delay functions of $G(n, k)$ satisfy the β-bounded-jump condition.

Proof. Our proof adjusts the construction of [44], which we recapitulate in the Appendix 6. The main idea of our adjustment is to construct a bottleneck congestion game $G(n, k)$ by generating k copies of the game $G(n)$. We then add resources to the strategies. These resources make sure that there is an improvement step for a player in $G(n)$ if and only if there is an improvement step of corresponding k players of the k copies in $G(n, k)$.

For each $j \in \{1, \ldots, 9\}$, $i \in \{1, \ldots, n\}$, and $m \in \{1, \ldots, k\}$, we add a resource $A^j_{i,k}$ to strategy j of player Main_i in copy m. Additionally, we add this resource to all strategies $j' \neq j$ of all players Main_i of every other copy $m' \neq m$. Each of these resources has delay of δ^{i-1} if it is allocated by at most one player and δ^{i+3} otherwise. Analogously, we add resources to the strategies of the auxiliary players. That is, for every player Block^j_i of every copy $m \in \{1, \ldots, k\}$, we add a resource B^j_i in his strategy 1. We also add this resource in every strategy 2 of every player Block^j_i of every copy $m' \neq m$. Similarly, for every player Block^j_i of every copy $m \in \{1, \ldots, k\}$, we add a resource C^j_i in his strategy 2, which we also add to every strategy 1 of every player Block^j_i of every copy $m' \neq m$. Each of these resources has a delay of δ^{i-1} if it is allocated by at most one player and δ^{i+3} otherwise. Finally, we have to increase δ slightly.

We obtain the initial strategy profile s' of $G(n,k)$ if every player of every copy m of $G(n)$ plays according to the initial strategy profile S of his copy. It it easy to see, that no coalition of less than k players of a copy m has an incentive to change their strategies. At least one of them would have to allocate a A-, B-, or C-resource that is already in use by another player. Thus, it is not an improvement step for these players. We, therefore, can conclude that all k copies of a player always choose the same strategy. On the other hand, if there is an improving move of one player in $G(n)$, there is a coalitional improving move of all k copies of that player in $G(n,k)$. If all players mimic this deviation in their copies, by construction, no two players allocate the same A-, B-, or C-resource. Furthermore, if the improvement step decreases the delay in $G(n)$, it does so for every copy of the player in $G(n,k)$.

Finally, note that as long as k is polynomial in n we obtain a reduction of polynomial size. In particular, for $k = n^{1/\epsilon - 1}$ we obtain a new game with nk players, for which the unilateral moves of $G(n)$ are exactly moves of coalitions of size $(nk)^{1-\epsilon}$ and no smaller coalitions have improving moves. This proves the theorem.

$\square$

4. Conclusion

We have provided a detailed study of the computational complexity of exact and approximate pure Nash and strong equilibria in bottleneck congestion games. However, some important and fascinating open problems remain. A major open problem is to find other interesting classes of games, for which efficient computation of and/or fast convergence to SE can be shown. As computation postulates less stringent requirements in terms of locality, there is generally more hope to derive positive results. In particular, what can be said about efficient computation of α-approximate SE?

For convergence to SE, we have provided a series of quite strong lower bounds. In this case, it natural to consider weaker concepts of stability that avoid our hardness results. For instance, we did not succeed in translating positive results known for ordinary congestion games and convergence to approximate PNE [11, 8, 10]. In addition, there are open problems regarding the duration of unilateral dynamics in symmetric network games and hardness of computing PNE in asymmetric networks. Finally, it is a major open problem how to augment the concept of PNE with resilience to coalitional deviations and avoid the hardness results we have observed. It would be interesting to consider computation and convergence characteristics of, e.g., k-SE, for $1 < k < n$, or partition equilibria [20].

Appendix

5. Basics in Matroid Theory

In the following, we will briefly introduce the notion of matroids. For a comprehensive introduction as well as for the proofs of the mentioned results we refer the reader to the textbooks of Korte and Vygen [32, Chapter 13] and Schrijver [43, Chapters 39 – 42].

Let F be a finite set. A tuple $M = (F, \mathcal{I})$ where $\mathcal{I} \subset 2^F$ is called a *matroid* if (i) $\emptyset \in \mathcal{I}$, ($ii$) if $I \in \mathcal{I}$ and $J \subseteq I$, then $J \in \mathcal{I}$, and (iii) if $I, J \in \mathcal{I}$ and $|J| < |I|$, then there exists an $i \in I \setminus J$ with $J \cup \{i\} \in \mathcal{I}$. A set $A \subseteq F$ is called *independent* if $A \in \mathcal{I}$ and *dependent*, otherwise. The set of (inclusion wise) maximal independent subsets of F is called the *basis* of M.

For given F, a matroid $(F, \mathcal{I})$ may be of exponential size, thus, one frequently assumes that a matroid comes with an *independence oracle* that returns for all sets $A \subseteq F$ whether $A \in \mathcal{I}$ or not. It shall be noted that for many subclasses of matroids an independence oracle can be implemented in polynomial time.

Another way of representing matroids is via a *rank function* $\mathrm{rk} : 2^F \to \mathbb{N}$. Every sub-cardinal, monotonic and sub-modular function rk gives rise to a matroid whose independent sets then are defined as $\{A \subseteq F : \mathrm{rk}(A) = |A|\}$. If the independent sets are known a priori via an independence oracle the rank function is defined as $\mathrm{rk}(A) = \max_{I \in \mathcal{I} : I \subseteq A} |I|$. With a slight abuse of notation, we define for a matroid $M = (F, \mathcal{I})$ the rank of the matroid itself as $\mathrm{rk}(M) = \mathrm{rk}(F)$.

To present our positive results for matroid bottleneck congestion games in a general framework we give the definition of matroid union. This concept has been introduced by Nash-Williams [38] and Edmonds [15].

Definition 5.1 (Matroid union)**.** Let $M_1 = (S_1, \mathcal{I}_1), \ldots, M_k = (S_k, \mathcal{I}_k)$ be matroids. Define the union of these matroids as $M_1 \vee \cdots \vee M_k = (S_1 \cup \cdots \cup S_k, \mathcal{I}_1 \vee \cdots \vee \mathcal{I}_k)$ where

$$\mathcal{I}_1 \vee \cdots \vee \mathcal{I}_k = \{I_1 \cup \cdots \cup I_k : I_1 \in \mathcal{I}_1, \ldots, I_k \in \mathcal{I}_k\}.$$

Nash-Williams proved that for k matroids $M_1 = (S_1, \mathcal{I}_1), \ldots, M_k = (S_k, \mathcal{I}_k)$ their union $M_1 \vee \cdots \vee M_k$ is a matroid again. The maximum cardinality of an independent set in $\mathcal{I}_1 \vee \cdots \vee \mathcal{I}_k$ equals the maximum cardinality of a common independent set of two suitably constructed matroids. This observation reduces the problem of finding a maximum-size set in $\mathcal{I}_1 \vee \cdots \vee \mathcal{I}_k$ to the intersection problem of two matroids, which can be solved in polynomial time, see Cunningham [13].

6. Description of $G(n)$

In this section, we recapitulate the construction of $G(n)$ from [44]. This shows that (bottleneck) congestion games do not converge quickly to a PNE even if the players only perform unilateral α-improving moves.

We construct a (bottleneck) congestion game $G(n)$ that resembles a recursive run of n *programs*, i.e., sequences of unilateral α-improving moves. After its activation, program i triggers a run of program $i-1$, waits until it finishes its run, and triggers it a second time. These sequences are deterministic apart from the order in which some auxiliary players make their improvement steps.

A program i is implemented by a gadget G_i consisting of a main player that we call Main_i and eight auxiliary players called $\mathrm{Block}_i^1, \ldots, \mathrm{Block}_i^8$. The main player has nine strategies numbered from 1 to 9. Each auxiliary player has two strategies, a first and a second one. A gadget G_i is *idle* if all of its players play their first strategy. Gadget G_{i+1} *activates* gadget G_i by increasing the delay of (the bottleneck resource in) the first strategy of player Main_i. In the following sequence

Strategies of Block_i^j	Resources	Delays
(1)	t_i^j b_i^j	$\delta^{i-1}/2\alpha^2\delta^{i-1}$ $\delta^{i-1}/\delta^{i+2}$
(2)	c_i^1	$2\alpha\delta^{i-1}/\delta^{i+2}$

Figure 2. Definition of the strategies of the players Block_i^j

Strategy	Resources	Delays
(1)	e_i^1	$\delta^i/9\alpha^9\delta^i$
(2)	e_i^2 $c_{i-1}^1, \ldots, c_{i-1}^9$ t_i^1	$8\alpha^8\delta^i$ $2\alpha\delta^{i-2}/\delta^{i+1}$ $\delta^{i-1}/2\alpha^2\delta^{i-1}$
(3)	e_i^3 e_{i-1}^1 t_i^2 b_i^1	$7\alpha^7\delta^i$ $\delta^{i-1}/9\alpha^9\delta^{i-1}$ $\delta^{i-1}/2\alpha^2\delta^{i-1}$ $\delta^{i-1}/\delta^{i+2}$
(4)	e_i^4 b_{i-1}^8 t_i^3 b_i^2	$6\alpha^6\delta^i$ $\delta^{i-2}/\delta^{i+1}$ $\delta^{i-1}/2\alpha^2\delta^{i-1}$ $\delta^{i-1}/\delta^{i+2}$
(5)	e_i^5 t_i^4 b_i^3	$5\alpha^5\delta^i$ $\delta^{i-1}/2\alpha^2\delta^{i-1}$ $\delta^{i-1}/\delta^{i+2}$
(6)	e_i^6 $c_{i-1}^1, \ldots, c_{i-1}^9$ t_i^5 b_i^4	$4\alpha^4\delta^i$ $2\alpha\delta^{i-2}/\delta^{i+1}$ $\delta^{i-1}/2\alpha^2\delta^{i-1}$ $\delta^{i-1}/\delta^{i+2}$
(7)	e_i^7 e_{i-1}^1 t_i^6 b_i^5	$3\alpha^3\delta^i$ $\delta^{i-1}/9\alpha^9\delta^{i-1}$ $\delta^{i-1}/2\alpha^2\delta^{i-1}$ $\delta^{i-1}/\delta^{i+2}$
(8)	e_i^8 b_{i-1}^8 t_i^7 b_i^6	$2\alpha^2\delta^i$ $\alpha\delta^{i-2}/\delta^{i+1}$ $\delta^{i-1}/2\alpha^2\delta^{i-1}$ $\delta^{i-1}/\delta^{i+2}$
(9)	e^9 t_i^8 b_i^7	$\alpha\delta^i$ $\delta^{i-1}/2\alpha^2\delta^{i-1}$ $\delta^{i-1}/\delta^{i+2}$

Figure 3. Definition of the strategies of the players Main_i. The delay of resource e_n^1 is constantly $9\alpha^9\delta^n$.

of improvement steps the player Main_i successively changes to the strategies $2, \ldots, 8$. We call this sequence a *run* of G_i. During each run, $Main_i$ activates gadget G_{i-1} twice by increasing the delay of the (bottleneck resource in the) first strategy of Main_{i-1}. Gadget G_{i+1} is blocked (by player Block_i^8) until player Main_i reaches its strategy 9. Then G_{i+1} continues its run, that is, it decreases the delay of the bottleneck resource in the first strategy of player Main_i, waits until gadget G_i becomes idle again, and afterwards triggers a second run of G_i. The role of the auxiliary players of G_i is to control the strategy changes of Main_i and Main_{i+1}.

In the initial state s, every gadget G_i with $1 \leq i \leq n-1$ is idle. Gadget G_n is activated. In every improvement path starting from s, gadget G_i is activated 2^{n-i} times, which yields the theorem.

Now we go into the details of our construction. The (bottleneck) congestion game $G(n)$ consists of the gadgets $G_1, \ldots, G_n$. Each gadget G_i consists of a player Main_i and the players $\text{Block}_i^1, \ldots, \text{Block}_i^8$. The nine strategies of a player Main_i are given in Figure 3. The two strategies of a player Block_i^j are given in Figure 2. $\delta = 10\alpha^9$ is a scaling factor for the delay functions.

The auxiliary players implement a locking mechanism. The first strategy of player Block_i^j is $\{t_i^j, b_i^j\}$ and its second strategy is $\{c_i^j\}$. The delays of the resources b_i^j and c_i^j are relatively small (δ^{i-1} and $2\alpha\delta^{i-1}$, respectively) if allocated by only one player. If they are allocated by two or more players, however, then each of them induce a significantly larger delay of δ^{i+2}. Theses resources are also part of the strategies of Main_i or Main_{i+1}. Note, that neither Main_i nor Main_{i+1} has an incentive to change to a strategy having a delay of δ^{i+2} or more. The delay of the resource t_i^j is chosen such that Block_i^j has an incentive to change to its second strategy if Main_i allocates this resource. If Main_i neither allocates this resource nor the resource b_i^j, it has an incentive to change to its first strategy. Due to scaling factor δ^{i-1} the delays of the resource t_i^j do not affect the preferences of Main_i.

These definitions yield the following properties. If auxiliary player Block_i^j of gadget G_i plays its first strategy then this prevents Main_i from choosing strategy $j+2$. Player Block_i^j has an incentive to change to its second strategy only if player Main_i chooses its strategy $j+1$. By this mechanism, we ensure that Main_i chooses the strategies 1 to 8 in the right order. In addition, the first strategy of Block_i^8 prevents Main_{i+1} from going to strategy 4 or 8. This ensures that Main_{i+1} waits until the run of player Main_i is completed. Furthermore, Main_{i+1} can enter into strategy 3 or 7 only if all auxiliary players of gadget G_i use their first strategy. This ensures that a run starts with all auxiliary players being in their first strategy.

This shows that in every sequence of improvement steps from s to a Nash equilibrium in the (bottleneck) congestion game $G(n)$ each gadget i is activated 2^{n-i} times. One can easily check that every improvement step of a player decreases its delay (of the bottleneck resource) by a factor of at least α and every delay function satisfies the β-bounded-jump condition with $\beta = \delta^3$ with $\delta = 10\alpha^9$.

References

[1] H. ACKERMANN, P. BERENBRINK, S. FISCHER, AND M. HOEFER, *Concurrent imitation dynamics in congestion games*, in Symp. Principles of Distributed Computing (PODC), 2009, pp. 63–72.

[2] H. ACKERMANN, H. RÖGLIN, AND B. VÖCKING, *On the impact of combinatorial structure on congestion games*, J. ACM **55**, no. 6 (2008).

[3] S. ALBERS, *On the value of coordination in network design*, SIAM J. Comput. **38**, no. 6 (2009), pp. 2273–2302.

[4] N. ANDELMAN, M. FELDMAN, AND Y. MANSOUR, *Strong price of anarchy*, Games Econom. Behav. **65**, no. 2 (2009), pp. 289–317.

[5] R. AUMANN, *Acceptable points in general cooperative n-person games*, in Contributions to the Theory of Games IV, Annals of Mathematics Study 40, Princeton University Press, 1959, pp. 287–324.

[6] R. BANNER AND A. ORDA, *Bottleneck routing games in communication networks*, IEEE J. Sel. Area Comm. **25**, no. 6 (2007), pp. 1173–1179.

[7] K. BÉRCZI AND A. FRANK, *Packing arborescences*, Tech. Report TR-2009-04, Egerváry Research Group on Combinatorial Optimization, 2009.

[8] A. Bhalgat, T. Chakraborty, and S. Khanna, *Nash dynamics in congestion games with similar resources*, in Proc. 5th Intl. Workshop Internet & Network Economics (WINE), 2009, pp. 362–373.

[9] C. Busch and M. Magdon-Ismail, *Atomic routing games on maximum congestion*, Theoret. Comput. Sci. **410**, no. 36 (2009), pp. 3337–3975.

[10] T. Chakraborty and S. Khanna, *Nash dynamics in constant player and bounded jump congestion games*, in Proc. 2nd Intl. Symp. Algorithmic Game Theory (SAGT), 2009, pp. 196–207.

[11] S. Chien and A. Sinclair, *Convergence to approximate Nash equilibria in congestion games*, Games Econom. Behav. **71**, no. 2 (2011), pp. 315–327.

[12] R. Cole, Y. Dodis, and T. Roughgarden, *Bottleneck links, variable demand, and the tragedy of the commons*, in Proc. 17th Symp. Discrete Algorithms (SODA), 2006, pp. 668–677.

[13] W. Cunningham, *Improved bounds for matroid partition and intersection algorithms*, SIAM J. Comput. **15**, no. 4 (1986), pp. 948–957.

[14] B. de Keijzer, G. Schäfer, and O. Telelis, *On the inefficiency of equilibria in linear bottleneck congestion games*, in Proc. 3rd Internat. Sympos. Algorithmic Game Theory, S. Konogiannis, E. Koutsoupias, and P. Spirakis, eds., 2010, pp. 335–346.

[15] J. Edmonds, *Matroid partition*, in Mathematics of the Decision Sciences, G. Dantzig and A.F.Veinott, eds., AMS, 1968, pp. 335–345.

[16] A. Epstein, M. Feldman, and Y. Mansour, *Efficient graph topologies in network routing games*, Games Econom. Behav. **66** (2009), pp. 115–125.

[17] A. Epstein, M. Feldman, and Y. Mansour, *Strong equilibrium in cost sharing connection games*, Games Econom. Behav. **67**, no. 1 (2009), pp. 51–68.

[18] A. Fabrikant, C. Papadimitriou, and K. Talwar, *The complexity of pure Nash equilibria*, in Proc. 36th Symp. Theory of Computing (STOC), 2004, pp. 604–612.

[19] M. Feldman and T. Tamir, *Approximate strong equilibrium in job scheduling games*, J. Artificial Intelligence Res. **36** (2009), pp. 387–414.

[20] M. Feldman and M. Tennenholtz, *Structured coalitions in resource selection games*, ACM Trans. Intell. Syst. Tech. **1**, no. 1 (2010), p. 4.

[21] A. Fiat, H. Kaplan, M. Levy, and S. Olonetsky, *Strong price of anarchy for machine load balancing*, in Proc. 34th Intl. Coll. Automata, Languages and Programming (ICALP), 2007, pp. 583–594.

[22] S. Fortune, J. Hopcroft, and J. Wyllie, *The directed subgraph homeomorphism problem*, Theoret. Comput. Sci. **10** (1980), pp. 111–121.

[23] H. Gabow, *A matroid approach to finding edge connectivity and packing arborescences*, J. Comput. Syst. Sci. **50**, no. 2 (1995), pp. 259–273.

[24] M. Gairing, B. Monien, and K. Tiemann, *Routing (un-)splittable flow in games with player-specific linear latency functions*, ACM Trans. Algorithms **7**, no. 3 (2011), pp. 1–31.

[25] T. Harks, M. Klimm, and R. Möhring, *Strong Nash equilibria in games with the lexicographical improvement property*, in Proc. 5th Intl. Workshop Internet & Network Economics (WINE), 2009, pp. 463–470.

[26] M. Hoefer and A. Skopalik, *On the complexity of pareto-optimal nash and strong equilibria*, in Intl. Symp. Algorithmic Game Theory (SAGT), 2010, pp. 312–322.

[27] R. Holzman and N. Law-Yone, *Strong equilibrium in congestion games*, Games Econom. Behav. **21**, no. 1-2 (1997), pp. 85–101.

[28] D. Johnson, C. Papadimitriou, and M. Yannakakis, *How easy is local search?*, J. Comput. Syst. Sci. **37** (1988), pp. 79–100.

[29] R. Kannan and C. Busch, *Bottleneck congestion games with logarithmic price of anarchy*, in Proc. 3rd Internat. Sympos. Algorithmic Game Theory, S. Konogiannis, E. Koutsoupias, and P. Spirakis, eds., 2010, pp. 222–233.

[30] S. Keshav, *An engineering approach to computer networking: ATM networks, the Internet, and the telephone network*, Addison-Wesley, 1997.

[31] H. Konishi, M. L. Breton, and S. Weber, *Equilibria in a model with partial rivalry*, J. Econ. Theory **72**, no. 1 (1997), pp. 225–237.

[32] B. KORTE AND J. VYGEN, *Combinatorial Optimization: Theory and Algorithms*, Springer Verlag, 2002.

[33] S. LEONARDI AND P. SANKOWSI, *Network formation games with local coalitions*, in Proc. 26th Symp. Principles of Distributed Computing (PODC), 2007, pp. 299–305.

[34] M. MAVRONICOLAS, I. MILCHTAICH, B. MONIEN, AND K. TIEMANN, *Congestion games with player-specific constants*, in Proc. 32nd Intl. Symp. Mathematical Foundations of Computer Science (MFCS), 2007, pp. 633–644.

[35] V. MAZALOV, B. MONIEN, F. SCHOPPMANN, AND K. TIEMANN, *Wardrop equilibria and price of stability for bottleneck games with splittable traffic*, in Proc. 2nd Intl. Workshop Internet & Network Economics (WINE), 2006, pp. 331–342.

[36] I. MILCHTAICH, *Congestion games with player-specific payoff functions*, Games Econom. Behav. **13**, no. 1 (1996), pp. 111–124.

[37] D. MONDERER AND L. SHAPLEY, *Potential games*, Games Econom. Behav. **14** (1996), pp. 1124–1143.

[38] C. NASH-WILLIAMS, *An application of matroids to graph theory*, in Theory of Graphs; Proc. of an International Symposium in Rome 1966, P. Rosenstiehl, ed., 1967, pp. 263–265.

[39] L. QIU, Y. R. YANG, Y. ZHANG, AND S. SHENKER, *On selfish routing in internet-like environments*, IEEE/ACM Trans. Netw. **14**, no. 4 (2006), pp. 725–738.

[40] R. ROSENTHAL, *A class of games possessing pure-strategy Nash equilibria*, Intl. J. Game Theory **2** (1973), pp. 65–67.

[41] T. ROUGHGARDEN, *Routing games*, in Algorithmic Game Theory, N. Nisan, É. Tardos, T. Roughgarden, and V. Vazirani, eds., Cambridge University Press, 2007, ch. 18.

[42] O. ROZENFELD AND M. TENNENHOLTZ, *Strong and correlated strong equilibria in monotone congestion games*, in Proc. 2nd Intl. Workshop Internet & Network Economics (WINE), 2006, pp. 74–86.

[43] A. SCHRIJVER, *Combinatorial Optimization: Polyhedra and Efficiency*, Springer Verlag, 2003.

[44] A. SKOPALIK AND B. VÖCKING, *Inapproximability of pure Nash equilibria*, in Proc. 40th Symp. Theory of Computing (STOC), 2008, pp. 355–364.

[45] V. SYRGKANIS, *Equilibria in congestion game models: Existence, complexity and efficiency*, master's thesis, National Technical University of Athens, 2009.

[46] M. VOORNEVELD, P. BORM, F. VAN MEGEN, S. TIJS, AND G. FACCHINI, *Congestion games and potentials reconsidered*, Intl. Game Theory Rev. **1**, no. 3 (1999), pp. 283–299.

Paper 9

Approximation Algorithms for Capacitated Location Routing

Tobias Harks, Jannik Matuschke and Felix G. König

Approximation Algorithms for Capacitated Location Routing
Transportation Science, to appear

Abstract. An *approximation algorithm* for an optimization problem runs in polynomial time for all instances and is guaranteed to deliver solutions with bounded optimality gap. We derive such algorithms for different variants of *capacitated location routing*, an important generalization of vehicle routing where the cost of opening the depots from which vehicles operate is taken into account. Our results originate from combining algorithms and lower bounds for different relaxations of the original problem, and besides location routing we also obtain approximation algorithms for *multi-depot capacitated vehicle routing* by this framework. Moreover, we extend our results to further generalizations of both problems, including a *prize-collecting* variant, a *group* version, and a variant where *cross-docking* is allowed. We finally present a computational study of our approximation algorithm for capacitated location routing on benchmark instances and large-scale randomly generated instances. Our study reveals that the quality of the computed solutions is much closer to optimality than the provable approximation factor.

1. Introduction

The broad realm of vehicle routing addresses the omnipresent logistic challenge of minimizing the cost of operating vehicles performing pickups and/or deliveries of goods for clients from a given set of depots. In many logistics applications, however, the cost of opening these depots constitutes a second major cost driver. Integrating this aspect of location decisions into the model leads to an additional and distinct optimization challenge. The two families corresponding to the routing and location subproblems, namely *vehicle routing* and *facility location*, have been studied extensively from practical as well as theoretical points of view. The integrated problem of jointly making location and routing decisions is known as *location routing* and has received significant attention in the operations research community as well.

A basic variant of location routing is the *capacitated location routing problem* (CLR) defined as follows. We are given an undirected connected graph $G = (V, E)$, where the node set $V =$

$\mathcal{C} \dot{\cup} \mathcal{F}$ of the graph is partitioned into a set of *clients* $\mathcal{C}$ and a set of *facilities* $\mathcal{F}$. We will use the term *facilities* interchangeably with the term *depots*. There are cost functions $c : E \to \mathbb{R}^+$ on the edge set, and $\phi : \mathcal{F} \to \mathbb{R}^+$ associated with the depots modeling *opening costs*. Every potential depot maintains an unbounded fleet of vehicles, each with a *uniform capacity* $u > 0$. Each client $v \in \mathcal{C}$ has a demand $d_v > 0$. A feasible solution to CLR is given by a tuple $(F, \mathcal{T})$, where $F \subseteq \mathcal{F}$ is a set of open depots and $\mathcal{T}$ is a set of tours $\{T_1, \ldots, T_k\}$ such that (1) every tour starts at an open depot and returns to the same depot at the end, (2) the demand of every client is served by the tours by which it is visited, and (3) the total demand served by a tour does not exceed u. The total cost of a solution is defined by $\sum_{T \in \mathcal{T}} c(T) + \sum_{w \in F} \phi(w)$, where $c(T) = \sum_{e \in T} c_e$ denotes the routing cost of tour T. Note that we may assume w.l.o.g. that G is complete and the edge costs c satisfy the triangle inequality: If this is not the case, we replace G by its metric closure. Furthermore, note that this model also implicitly covers depot-dependent fixed costs per tour, i.e., each vehicle sent out from depot v incurs a cost of $a_v \in \mathbb{R}^+$. This can be easily modeled by adding $\frac{1}{2}a_v$ to the cost of all edges incident to v, as each tour originating at v contains exactly two of these edges.

In the version of the problem described above, a client's demand may be split up and served by multiple facilities, which is not always desired or even possible in practice. This motivates the following terminology. A solution to CLR fulfills the *single-assignment property* (cf. [37, 31]), if the demand of each client is served by exactly one facility. A solution fulfills the *single-tour property*, if each client's demand is served by exactly one tour. Clearly, this latter property can only be fulfilled if $d_v \leq u$ for all $v \in \mathcal{C}$.

The special case of CLR where location decisions have already been made (i.e., $\phi \equiv 0$) is the *multi-depot capacitated vehicle routing problem* (MDCVR). Note that in the uncapacitated case ($u = \infty$), CLR and MDCVR are equivalent: By triangle inequality, every optimal solution to either problem can be transformed such that each depot is visited by at most one tour (without increasing cost). Hence, facility opening cost can be modeled by adding $\frac{1}{2}\phi(v)$ to $c(e)$ for all edges e incident to a facility $v \in \mathcal{F}$.

Not surprisingly, CLR contains *NP*-hard combinatorial optimization problems as a special case. When there is only one facility and infinite vehicle capacity, for instance, the problem becomes the travelling salesman problem. Or, when demands are uniform and match the vehicle capacity, CLR becomes the (metric) uncapacitated facility location (UFL) problem, as an optimal routing corresponds to finding shortest paths from each client to an open facility.

Because of this intrinsic hardness, an *exact* solution method for most location routing problems including CLR is very likely to perform poorly on some problem instances. Speaking more formally, its *worst-case running time* is likely to grow *exponentially* with problem size [25]. In fact, even for simple variants of vehicle routing problems, only relatively small instances are solved to optimality, see the book by [41] and references therein. On the other hand, problem sizes encountered in real-life problems have grown tremendously over the past years (and are expected to grow further), thus fast heuristics are becoming increasingly important for solving location and vehicle routing problems [15, 18]. While *(meta-)heuristics* used today deliver feasible solutions to larger instances in reasonable time, there is usually *no guaranteed bound* regarding solution quality. Merely for some restricted special cases, there are heuristics for which such bounds are known, see [25].

To address this apparent dilemma regarding worst-case running time and guaranteed solution quality, we use *approximation algorithms* in this paper, a solution methodology in the intersection of mathematics, computer science, and operations research. An approximation algorithm for an *NP*-hard combinatorial optimization problem is a heuristic enjoying two desirable properties: Its worst-case running time is bounded by a polynomial in problem size, and there are provable a priori bounds (constant numbers in the best case) on the worst-case quality of the solution generated:

Definition 1.1. An algorithm *ALG* for a minimization problem P is a *ρ-approximation algorithm* if it runs in time polynomial in the input size, and for every instance I of P, we have

$$\mathrm{ALG}(I) \leq \rho \cdot \mathrm{OPT}(I),$$

where $\mathrm{ALG}(I)$ and $\mathrm{OPT}(I)$ denote the objective values of the solution returned by *ALG* and of an optimal solution for I, respectively.

While this worst-case guarantee gives theoretical evidence for the reasonability of the algorithm, the quality of solutions may be much closer to optimality in practice than the approximation factor indicates. A standard reference containing approximation algorithms for a multitude of hard optimization problems is the book of [27]. Another recent and very good reference, containing a detailed introduction to the various techniques used in the design of approximation algorithms, is the book by [46].

Within this framework, we devise a constant factor approximation algorithm for CLR with arbitrary demands. For MDCVR with arbitrary demands, we obtain an improved approximation factor, which is, to the best of our knowledge, the best constant factor approximation algorithm for this problem to date. Moreover, we consider three practically relevant extensions of the above model. Suppose a company has the option not to serve all clients' demands itself, but to outsource any number of transports to clients at given customer-dependent prices. This extended model is known as the *prize-collecting* capacitated location routing problem (PC-CLR). In the second extension, we consider *group* capacitated location routing (G-CLR) where the set of clients is partitioned into groups $\mathcal{C}_1, \dots, \mathcal{C}_k$, with $\mathcal{C} = \bigcup_{i=1}^{k} \mathcal{C}_i$. In a feasible solution, only one client from each group needs to be served. Applications include intermodal transport networks, where goods can be transferred from one logistics network to the next at one of several hub locations. In the third extension, *cross-docking* is allowed: We allow consolidation tours which do not visit a facility, but contain one node where they meet with other tours. From there, spare capacity on the latter tours is utilized jointly to forward all demand of the consolidation tour to facilities. Being of profound practical interest (see e.g. [43, 45]), cross-docking operations may significantly improve capacity utilization and hence reduce total cost. We extend our constant factor approximations to all three of these variants, where for the group version, our approximation guarantee depends on the cardinality of the largest group.

1.1. Previous Results

Location routing (as the integration of vehicle routing and facility location) has occupied a central place in the operations research literature over the past decades. Since hundreds of papers have been published in this broad area, we will give pointers to text books and survey articles when referring to the main streams in location routing. However, we give a concise overview of works regarding approximation algorithms on the subject.

Location Routing. Perhaps one of the earliest models of location routing appears in the paper by [44]. [29] gives a comprehensive overview of the literature prior to the late 80s. More recent survey articles summarizing heuristic algorithms and mathematical programming formulations for many variants of location routing can be found in [36] and [37]. Very recently, there have been several works on integer programming formulations for CLR with capacitated facilities using strengthened cut inequalities, see [5] and [14].

There are only a few works that are concerned with approximation theory for location routing problems. For unbounded vehicle capacity, a $(2 - \frac{1}{|V|-1})$-approximation algorithm is given by [21]. [20] generalize this result to the case of (uncapacitated) group location routing, where one is given a system of groups of clients, and only one client from each group needs to be served. Among other results, they derive a $(2 - \frac{1}{|V|-1})L$-approximation algorithm, where L denotes the cardinality of the largest group. Finally, [11] provide a 24-approximation for location routing with soft facility capacities (i.e., facilities can be installed multiple times, each copy capable to serve a limited amount of demand, while vehicle loads are still unbounded).

Vehicle Routing. When facilities can be opened at no cost, location routing becomes the multi-depot vehicle routing problem, for which countless exact and heuristic solution methods have been proposed. For an overview of the rich literature in this field, we refer the reader to the books edited by [41], [23], and the surveys of [3], [15, 16], [25], [30], and [32]. For vehicle routing problems with additional side constraints (such as time-windows, heterogeneous fleets, fleets of limited size) see also [6], [3], and [18].

There is a large body of work regarding the classical capacitated vehicle routing problem (with a single depot) including the seminal PTAS of [24] for geometric distances. [33] consider the multi-depot capacitated vehicle routing problem (MDCVR) and present, among other results, a $(2 + 2\rho_{\text{TSP}})$-approximation algorithm for arbitrary, unsplittable demands, where ρ_{TSP} denotes the factor of an approximation algorithm for the travelling salesman problem. This is the best previously known approximation algorithm for this version of the problem, with $\rho_{\text{TSP}} = 3/2$ using the algorithm by [12]. There is also a PTAS for the case of Euclidean distances and uniform demands, albeit with running time exponential in vehicle capacity as well as the number of depots [8].

[10] studied the related k-delivery TSP in which a single vehicle with capacity k needs to transport n (unit-sized) items located at arbitrary locations to given demand points. For this problem they derive a 5-approximation.

Facility Location. Approximation algorithms for (metric) uncapacitated facility location (UFL) constitute a central topic in combinatorial optimization. As a reference, we point the reader to the 1.52-approximation of [35]. Using ideas of [13], a recent publication by [7] improves this factor to 1.5, also introducing a bifactor approximation that provides separate approximation ratios for connection and opening costs with respect to an initially solved LP relaxation.

[40] study the related *capacitated cable facility location problem*. As in CLR, one is given a complete undirected graph with costs on the edges. A set of clients needs to be served from facilities with associated opening costs. Facilities need to be opened and clients need to be connected to open facilities by Steiner trees, where an edge e of a tree is associated with a number of cables bought for the corresponding connection, each at price $c(e)$. Each cable has uniform capacity u, and each connection needs to comprise enough cables to provide capacity no less than the number of clients depending on it. The authors propose a $(\rho_{\text{UFL}} + \rho_{\text{ST}})$-approximation algorithm, where ρ_{UFL} and ρ_{ST} denote the approximation factors of algorithms for UFL and Steiner tree, respectively, which are used as subroutines. Their algorithm computes a feasible solution by merging a UFL and a Steiner tree solution. The merging procedure first routes the entire demand along the Steiner tree and then iteratively relieves overloaded subtrees of excessive demand by rerouting it to a closest open facility in the UFL solution.

The approximation algorithms in this paper use a similar technique of merging two solutions (UFL and a minimum spanning tree) by iteratively rerouting demand from overloaded subtrees of the spanning tree to a closest open facility of the UFL solution. Since our model is tour-based, however, we cannot argue on individual link capacities, or use corresponding flow arguments. The merging procedure in [40] crucially relies on the flexibility to install sufficient cable capacity on individual edges, and to fractionally route flow under these capacity constraints. In contrast, we have to decide about buying complete tours from open facilities, requiring a different rerouting procedure.

Extended Models. In the *prize-collecting* (PC) version of the above problems, a feasible solution does not have to serve all clients. Instead, an individual penalty may be paid for each unserved client. Thereby, PC can precisely model outsourcing decisions and is hence of profound practical interest. For the PC version of UFL, [28] claim to obtain a 2-approximation, improving the 3-approximation by [9], but omitting a complete proof. We are not aware of any previous approximation results for PC vehicle routing or PC location routing.

In the *group* variant, the set of clients is partitioned into disjoint subsets, or groups of clients, and only one client from every group has to be served. Group facility location is closely related to unweighted set cover, as we shall see in Section 4. For the group case of uncapacitated vehicle and location routing, the only previous result we know of is the algorithm by [20] mentioned above.

Finally, in capacitated location routing and multi-depot capacitated vehicle routing, *cross-docking* may be allowed in certain application scenarios. Here, some clients are served by consolidation tours which do not connect directly to a facility, but meet with other tours having spare capacity. These latter tours jointly forward all demand required by the consolidation tour to their respective facilities. Cross-docking plays a significant role in numerous logistics applications, and some heuristic approaches have recently been proposed for vehicle routing with cross-docking [43, 45]. This model also exhibits strong similarity to a practically relevant problem called mixed truck delivery which is studied in [34]. Here, delivery tours are sought as well, and clients may be served by tours either from facilities or from hubs, which are in turn served by facilities. The authors develop a heuristic solution approach and present computational results suggesting that routing cost can be reduced on average by around 10% for random instances when allowing cross-docking. Our model corresponds to the case where each client node may also function as a hub.

1.2. Our contribution and structure of the paper

In Section 2 we develop a framework for combining approximation algorithms for facility location with spanning or Steiner tree algorithms in order to obtain approximation algorithms for capacitated location routing and multi-depot capacitated vehicle routing problems. We apply our technique to devise a constant factor approximation algorithm for CLR with arbitrary demands. We are not aware of any previous results regarding constant factor approximations for CLR. For MDCVR, we obtain an improved approximation guarantee which is, to the best of our knowledge, the best approximation factor to date. In Sections 3, 4 and 5 we study the prize-collecting, group, and cross-docking variants. We extend our approximation algorithm to all three variants. While we derive constant factor approximations for the prize-collecting and cross-docking versions, the approximation guarantee for the group version depends on the cardinality of the largest group. In fact, we show that this version of the problem does not allow for a constant factor approximation by providing a lower bound on the achievable approximation factor depending on the number of groups. In Section 6 we present a computational study of our algorithm for CLR, where we compare solution quality and running time with those of other algorithms for CLR from the literature on benchmark instances. It turns out that in practice, the algorithm's performance greatly exceeds its theoretically proven approximation guarantee. On the benchmark test set, the quality of our solutions is on average within a factor of 1.1–1.2 of best known solutions. While the increase in cost over other algorithm is mild, our algorithm's running time is several magnitudes faster, taking only negligible time on benchmark instances. To further demonstrate this computational efficiency we test our algorithm on a set of large-scale randomly generated instances (1000–10000 customers, 100–1000 facilities per instance). We are not aware of any previous work solving CLR instances of comparable size. We conclude the paper in Section 7 with a brief summary and a discussion of open problems.

2. Approximation Algorithm for Capacitated Location Routing

In this section, we present our main approximation result. After deriving two lower bounds, we present our algorithm for CLR followed by its analysis. Finally, we describe a specialization for multi-depot capacitated vehicle routing yielding an improved approximation guarantee.

Before we start, we introduce some additional notation. As described in the introduction, a feasible solution to CLR consists of a set of open facilities F and a set of tours (or, in mathematical terms, closed walks) $\mathcal{T}$ such that (1) each tour visits an open facility, (2) the demand of each client is served by the tours by which it is visited, and (3) the demand transported by a tour does not exceed the vehicle capacity u. The second and third condition can be expressed by the existence of *demand assignments*, i.e., non-negative values x_{vi} for each $v \in \mathcal{C}$ and $i \in \mathcal{T}$ with $x_{vi} > 0$ only if client v is visited by tour T_i, and fulfilling (2) $\sum_{i \in \mathcal{T}} x_{vi} = d_v$ for all $v \in \mathcal{C}$ and (3) $\sum_{v \in \mathcal{C}} x_{vi} \leq u$ for all $i \in \mathcal{T}$. Although these demand assignments are not part of the actual solution output, they can be computed efficiently from the tuple $(F, \mathcal{T})$ and we will use them in proofs throughout the paper.

2.1. Two Lower Bounds

We provide two lower bounds on the optimal solution, which will be used to derive a constant approximation factor for our algorithm.

Lemma 2.1. Given an instance of CLR, consider an uncapacitated facility location (UFL) instance defined as follows. The sets of clients and facilities remain the same as in CLR, but we set the costs of edges to $\tilde{c} := \frac{2}{u}c$. Then, the cost of an optimal solution to UFL (w.r.t. $\tilde{c}$) is at most the cost of an optimal solution to CLR (w.r.t. c).

Proof. Consider a feasible solution $(F, \mathcal{T})$ of CLR with demand assignments x_{vi}. Construct a solution U of the UFL instance by opening all facilities that were opened by CLR and connecting each client $v \in \mathcal{C}$ to a closest open facility $w(v) \in F$. The connection cost of U is $\tilde{c}(U) = \sum_{v \in \mathcal{C}} \tilde{c}_{vw(v)} d_v$. We will show that $\tilde{c}(U) \leq \sum_{T \in \mathcal{T}} c(T)$, which proves the lemma.

Consider a flow f constructed from the CLR solution as follows. For every client $v \in \mathcal{C}$ and each tour T_i serving v, partition T_i into two paths from the facility to the client and send x_{vi} units of flow along each path. Note that the amount of flow carried by edge $e \in E$ is at most u times the number of tours containing e and, thus, e is contained in at least $\left\lceil \frac{f_e}{u} \right\rceil$ tours. Denoting the cost of f w.r.t. $\tilde{c}$ by $\tilde{c}(f)$ we deduce

$$\frac{1}{2}\tilde{c}(f) = \frac{1}{u}\sum_{e \in E} c_e f_e \leq \sum_{e \in E} c_e \left\lceil \frac{f_e}{u} \right\rceil \leq \sum_{e \in E} \sum_{T \in \mathcal{T} : e \in T} c_e = \sum_{T \in \mathcal{T}} c(T).$$

Note that the construction of the flow f induces a path decomposition. Let $\mathcal{P}_v$ be the set of all paths from a facility to client $v \in \mathcal{C}$ used in the construction of f and let f_P be the flow value assigned to that path. Note that $\sum_{P \in \mathcal{P}_v} f_P = \sum_{i=1}^{k} 2x_{vi} = 2d_v$, because every tour contributes two paths for every client it serves. Furthermore, $\tilde{c}(P) \geq \tilde{c}_{vw(v)}$, i.e., the length of any of the facility-client-paths is at least the distance to a closest facility. Thus, we obtain

$$\tilde{c}(f) = \sum_{v \in \mathcal{C}} \sum_{P \in \mathcal{P}_v} \tilde{c}(P) f_P \geq \sum_{v \in \mathcal{C}} \tilde{c}_{vw(v)} \sum_{P \in \mathcal{P}_v} f_P = \sum_{v \in \mathcal{C}} \tilde{c}_{vw(v)} 2d_v = 2\tilde{c}(U)$$

showing that $\tilde{c}(U) \leq \sum_{T \in \mathcal{T}} c(T)$. □

Lemma 2.2. Given an instance of CLR, consider the graph $G' = (V \cup \{r\}, E \cup E')$, where $E' = \{\{r, w\} : w \in \mathcal{F}\}$ and define costs $c'_{rw} = 0$, $c'_{vw} = c_{vw} + \frac{1}{2}\phi(w)$ for all $v \in \mathcal{C}$, $w \in \mathcal{F}$, and $c'_{vw} = c_{vw}$ for all other $\{v, w\} \in E$. Then the cost of a minimum spanning tree in G' with respect to the costs c' is a lower bound on the cost of an optimal solution of CLR (w.r.t. c).

Proof. Consider a feasible solution $(F, \mathcal{T})$ to CLR. We will construct a spanning tree in G' that has at most the same cost. For every open facility $w \in F$, let $T_1, \ldots, T_j$ be the tours based at w with $T_i = (w, v_1^i, \ldots, v_{l_i}^i, w)$ where l_i is the number of clients in T_i. For $i = 1, \ldots, j-1$, replace the last edge $\{v_{l_i}^i, w\}$ of T_i and the first edge $\{w, v_1^{i+1}\}$ of T_{i+1} by the direct edge $\{v_{l_i}^i, v_1^{i+1}\}$. Also remove the final edge $\{v_{l_j}^j, w\}$ of T_j. As a result, we get a walk P_w from w to $v_{l_j}^j$ along all clients that are served by w. Note that $c'(P_w) = \sum_{e \in P_w} c'(e) \leq \sum_{i=1}^{j} c(T_i) + \frac{1}{2}\phi(w)$ by triangle inequality and the fact that P_w contains only one edge incident to w.

Now let S be the union of all P_w for $w \in F$ and all edges in E'. As S spans all facilities and contains a walk from any client to a facility, it contains a spanning tree of G' with cost at most $c'(S) \leq \sum_{T \in \mathcal{T}} c(T) + \sum_{w \in F} \phi(w)$. □

2.2. Algorithm

We construct an approximate solution to CLR from an approximate solution to the UFL instance described in Lemma 2.1 and a minimum spanning tree on the graph G' as described in Lemma 2.2. Essentially, the idea is to decompose the spanning tree into subtrees with demands between $u/2$ and u, which can then be turned into tours by doubling edges. These tours are serviced by facilities opened by either the spanning tree or the UFL solution. The cost of the resulting solution is bounded by the sum of twice the cost of the spanning tree, twice the connection cost of the UFL solution, and once the opening cost of the UFL solution. Using the bifactor approximation algorithm of [7] for UFL, we obtain a total approximation factor of 4.38 for CLR.

We now describe the algorithm in more detail. After solving the UFL instance approximately and computing a minimum spanning tree, we open all facilities that are opened in the UFL solution and also all facilities w that are incident to an edge other than $\{w, r\}$ in the spanning tree S. Any client with demand $d_v \geq u$ is assigned to a closest open facility and served by $\lceil \frac{d_v}{u} \rceil$ tours comprising only the assigned facility and the client. We proceed to describe how to construct tours for the remaining demands by merging the given spanning tree on G' with a UFL solution to obtain a feasible solution to CLR (this will later be referred to as the "merge phase"). For a better understanding, direct the spanning tree towards the root r and denote the subtree rooted at node v by S_v with D_v being the sum of the demands of all clients in S_v.

If z is a facility and the total demand in S_z is at most u, we turn this subtree into a tour based at z by doubling edges and short-cutting by triangle inequality. If the total demand in S_z exceeds u, we will relieve this subtree by rerouting excessive demand to other open facilities, charging the costs to the UFL solution, until the remaining demand is at most u. This last step resembles a technique introduced by [40].

We now describe our rerouting procedure in detail. Let v be a node in S_z such that $D_v > u$ but $D_w \leq u$ for all children w of v. Let I be the set containing all subtrees S_w with w being a child of v and the set $\{v\}$ itself. We want to make sure that less than u units of demand have to be routed to the parent of v in the tree and the rest of the demand is connected with additional edges paid for by the UFL solution. To this end, we greedily partition I into groups $I_0, \ldots, I_k$ such that the sum of demands of all subtrees in each group I_j is at most u but at least $u/2$ (unless $j = 0$). We keep the connection of all trees in I_0 to the node v, but we extract the trees of all other groups from the spanning tree (including the edges connecting them with v). For each $j = 1, \ldots, k$, the subtrees in group I_j together with the edges to v form one single tree which can be turned into a tour by doubling edges and short-cutting. Among all clients on this tour we choose one with the cheapest connection cost to an open facility and insert this facility into the tour, paying at most twice the cost of the corresponding edge by triangle inequality. Observe that this edge carries at least $u/2$ units of demand. We repeat this procedure until the total demand in the subtree S_z is at most u. Then we turn the remainder of S_z into a tour, again by doubling edges.

Algorithm 1 Algorithm for CLR.

Input: An instance of CLR.
Output: A feasible solution to CLR.

UFL phase:
Create an UFL instance with edge costs $\tilde{c} = \frac{2}{u}c$ as described in Lemma 2.1.
Apply the bifactor approximation algorithm of Byrka and Aardal with $\gamma = 2.38$ on this instance and let F_1 be the set of facilities opened in the resulting UFL solution.
Tree phase:
Construct the graph G' with edge costs c' as described in Lemma 2.2 and compute a minimum spanning tree S.
Let F_2 be the set of facilities that are incident to an edge in $S \cap E$.
Large demand phase:
Open all facilities in $F_1 \cup F_2$.
for all $v \in \mathcal{C}$ with $d_v \geq u$ **do**
  Construct $\lceil \frac{d_v}{u} \rceil$ copies of a tour from v to a closest open facility.
  Add the tours to $\mathcal{T}$ and remove the corresponding demand d_v.
end for
Merge phase:
for all $z \in F_2$ **do**
  while $D_z > u$ **do**
    Let $v \in V(S_z)$ such that $D_v > u$ but $D_w \leq u$ for all children w of v.
    Let $I = \{V(S_w) \,:\, w \text{ is a child of } v\} \cup \{\{v\}\}$.
    Find a partition $I = I_0 \dot\cup \ldots \dot\cup I_k$, such that $\sum_{v \in I_j} d_v \leq u$ for all $j \in \{0, \ldots, k\}$ and $\sum_{v \in I_j} d_v > \frac{u}{2}$ for all $j \in \{1, \ldots, k\}$.
    for all $j \in \{1, \ldots, k\}$ **do**
      Find a pair (w, z') such that w is a vertex of a tree in I_j, $z' \in F_1 \cup F_2$ and c_{wz} is minimal.
      Construct a tour visiting all vertices of trees in I_j and z' by doubling wz and the edges of all trees in I_j and short-cutting.
      Add the tour to $\mathcal{T}$ and remove the corresponding subtrees in I_j from S.
    end for
  end while
  Construct a tour from S_z by doubling all edges and short-cutting.
  Add the tour to $\mathcal{T}$.
end for
Clean-up phase:
Remove all facilities from $F_1 \cup F_2$ that are not on any of the tours in $\mathcal{T}$.
return $(F_1 \cup F_2, \mathcal{T})$

2.3. Analysis

We analyze the algorithm presented in the previous section to show that it is a 4.38-approximation for CLR. We start by estimating the cost of the solution produced in the merge phase against the cost of the spanning tree and the facility location solution.

Lemma 2.3. The solution to CLR constructed by Algorithm 1 in the large demand and merge phases from the spanning tree S and the UFL solution U has cost at most $2c'(S) + 2\tilde{c}(U) + \phi(U)$.

Proof. Every tour constructed in the large demand phase for a client $v \in \mathcal{C}$ has cost at most $2\lceil\frac{d_v}{u}\rceil c_{vw(v)}$, where $w(v)$ is a closest open facility in U. This is bounded by twice the connection cost for v in the UFL solution as $2\lceil\frac{d_v}{u}\rceil c_{vw(v)} \leq 2 \cdot 2\frac{d_v}{u} c_{vw(v)} \leq 2\,\tilde{c}_{vw(v)}d_v$, because $\frac{d_v}{u} \geq 1$.

Consider a tour T constructed during an iteration of the inner "for" loop of the merge phase in Algorithm 1. The cost of the tour is at most twice the cost of the edges of the corresponding subtree plus $2c_{wz'}$. Observe that, by the choice of w and z', the edge $\{w, z'\}$ is at most as expensive as any other edge used in U to connect any of the clients x on the tour to its facility $y(x)$. As the sum of the demands on the tour is at least $\frac{u}{2}$, we obtain

$$\sum_{x \in V(T)} \tilde{c}_{xy(x)} d_x \geq \tilde{c}_{wz'} \sum_{x \in V(T)} d_x \geq \tilde{c}_{wz'} \frac{u}{2} = c_{wz'}.$$

Thus, the total cost of all tours constructed in the inner loop amounts to at most twice the connection cost of U plus twice the costs of the corresponding subtrees. The tours constructed in the outer loop and the opening costs of all facilities in F_2 are bounded by twice the costs of the remaining subtrees S_z (w.r.t. c'), and the opening costs of all facilities in F_1 are $\phi(U)$. As all subtrees are pairwise disjoint, summing everything up yields the desired result. $\square$

Consequently, if S is a minimum spanning tree and U is a ρ-approximation to a minimum cost solution to the UFL instance, the merge phase of Algorithm 1 returns an $(2+2\rho)$-approximation to CLR. Note, however, that in this analysis $\phi(U)$ is counted twice while the actual solution only pays it once. We can improve the approximation factor by using a bifactor approximation algorithm for UFL of [7]. Given a parameter $\gamma > 1.678$, this algorithm returns a solution whose opening cost exceeds the opening cost of an initially computed optimal fractional LP solution U_{LP} by at most a factor of γ, and whose connection cost exceeds the connection cost of the fractional solution by at most $1+2e^{-\gamma}$. In this way, we obtain a solution U with $2\,\tilde{c}(U)+\phi(U) \leq 2\,(1+2e^{-\gamma})\,\tilde{c}(U_{\mathrm{LP}})+\gamma\,\phi(U_{\mathrm{LP}})$, which is bounded by $\gamma(\tilde{c}(U_{\mathrm{LP}})+\phi(U_{\mathrm{LP}}))$ for all $\gamma \geq 2.38$. Choosing $\gamma = 2.38$, Lemma 2.3 yields our main result.

Theorem 2.4. Algorithm 1 is a 4.38-approximation algorithm for CLR. The solution it produces fulfills the single-assignment property. If $d_v \leq u$ for all $v \in \mathcal{C}$, it furthermore fulfills the single-tour property.

On the other hand, the approximation ratio of our algorithm improves naturally in cases where an approximation algorithm for UFL with an approximation ratio better than 1.19 can be used. One example are graphs with Euclidean edge cost. Here, a PTAS for UFL [1] can be applied to obtain a $(4+\epsilon)$-approximation for CLR.

2.4. Special Case: Multi-Depot Capacitated Vehicle Routing

The special case of CLR, where opening facilities does not incur cost ($\phi \equiv 0$) is the *multi-depot capacitated vehicle routing problem* (MDCVR) as considered in [33, 8]. By a slight modification of Algorithm 1, we obtain an improved approximation ratio for this problem: Instead of solving the UFL instance approximately in the UFL phase, we solve it exactly by opening all facilities and assigning clients to facilities along shortest client-facility paths. We thus can replace the factors incurred by the bifactor UFL-algorithm by 1 and obtain the following result.

Theorem 2.5. When solving the UFL instance by shortest path computation, Algorithm 1 is a 4-approximation algorithm for MDCVR. The solution it produces fulfills the single-assignment property. If $d_v \leq u$ for all $v \in \mathcal{C}$, it furthermore fulfills the single-tour property.

Note that this improves the previously best known approximation guarantee of 5 for MDCVR in [33] yielding the single-assignment property.

3. Prize-Collecting Location Routing

We now apply our algorithmic framework for CLR and MDCVR to the *prize-collecting* (PC) variant of these problems. In a prize-collecting setting, we can decide for each client whether to serve it by our solution, or to pay a penalty for not serving it. Note that prize-collecting can naturally be viewed as a way of incorporating outsourcing decisions into an optimization model: In this case, a customer's penalty corresponds to the cost of having it served by an outside service provider. As outsourcing is an important option in many logistics applications, the prize-collecting variants of CLR and MDCVR are highly relevant in practice. Moreover, it is not hard to see that PC-CLR and PC-MDCVR are generalization of CLR and MDCVR, respectively: By setting penalties high enough, we can force any optimal solution to serve all clients.

Formally, an instance of PC-CLR comprises an instance of CLR together with a penalty function $p : \mathcal{C} \to \mathbb{R}^+$, and a solution is now a three-tuple $(F, \mathcal{T}, C)$, where $F \subseteq \mathcal{F}$ is a set of open facilities as before, $C \subseteq \mathcal{C}$ is the set of clients served, and $\mathcal{T}$ is a set of tours as before, except that we require only the demands of clients in C to be served by $\mathcal{T}$. The cost of a solution to PC-CLR is $\sum_{T \in \mathcal{T}} c(T) + \sum_{w \in F} \phi(w) + \sum_{v \in \mathcal{C} \setminus C} p(v)$. As before, PC-MDCVR is the special case of PC-CLR where $\phi \equiv 0$.

3.1. Algorithm

The key challenge in solving the prize-collecting variant by our algorithm lies in the choice of C: On the one hand, both our solution to UFL (Lemma 2.1) and our spanning tree (Lemma 2.2) need to serve the same set of clients in order for our rerouting procedure to work. On the other hand, we need to ensure that the sum of the costs of these partial solutions remains a lower bound for the original problem. We accomplish this by utilizing an approximation algorithm for PC-UFL, and an LP-based approximation algorithm for the prize-collecting Steiner tree to determine two respective sets of customers served. We then compute a solution to PC-CLR serving exactly those customers served by both the tree and the facility location solution.

A formal description of the algorithm is given in Algorithm 2. We will prove that it is a $(\rho_{\text{PC-ST}} + 2\rho_{\text{PC-UFL}})$-approximation algorithm for PC-CLR, where $\rho_{\text{PC-ST}}$ and $\rho_{\text{PC-UFL}}$ denote the approximation factors of the approximation algorithms used for prize-collecting Steiner tree (w.r.t. the undirected cut relaxation) and PC-UFL, respectively. Currently, the best known approximation algorithm for PC-UFL achieves an approximation ratio of $\rho_{\text{PC-UFL}} = 2$ [28], while for prize-collecting Steiner tree the algorithm of [21] achieves an approximation factor of $2 - \frac{1}{|V|}$, meeting the integrality gap of the LP relaxation. Using these algorithms results in an approximation factor of 6 for our algorithm.

First note that an equivalent of Lemma 2.1 still holds in a prize-collecting setting: In its proof, we constructed a feasible solutions to a scaled instance of UFL from any feasible solution to CLR without increasing cost. It is easy to see that this construction adapts naturally when transferring the set of clients served from an optimal PC-CLR solution to a feasible solutions to PC-UFL: The penalties for customers not served are exactly the same in both solutions.

To obtain the second, tree based lower bound, we consider a prize-collecting Steiner tree instance defined as follows. We add a root node r to the network and connect it to all facilities, i.e., we consider the graph $G' = (V \cup \{r\}, E \cup E')$ with $E' = \{\{r, w\} : w \in \mathcal{F}\}$ as constructed in Lemma 2.2. We then extend the cost function c to E' by defining cost $c_{rw} = \frac{1}{2}\phi_w$ for each $w \in \mathcal{F}$ and define new penalties by setting $p' := \frac{1}{2}p$. We let $R = \mathcal{C} \cup \{r\}$ be the set of terminals. We will use an approximation algorithm on this prize-collecting Steiner tree instance that is based on the following undirected cut relaxation.

Algorithm 2 Algorithm for PC-CLR.

Input: An instance of PC-CLR.
Output: A feasible solution to PC-CLR.

1: **UFL phase:**
2: Create a UFL instance as in the UFL phase of Algorithm 1. Add p to obtain an instance of PC-UFL.
3: Run an approximation algorithm for PC-UFL. Let U be the returned UFL solution and C_1 denote the set of served clients and F_1 be the set of opened facilities in U.
4: **Steiner tree phase:**
5: Construct the graph G' as in Lemma 2.2.
6: Run an approximation algorithm for prize-collecting Steiner tree on the instance given by G', the terminal set $\mathcal{C} \cup \{r\}$ and penalties p. Let S be the resulting tree and C_2 denote set of connected customers and F_2 be the set of connected facilities in the Steiner tree.
7: **Merge phase:**
8: Set $C := C_1 \cap C_2$.
9: Run the large demand and merge phases of Algorithm 1 using U and S, serving only clients in C. Let $\mathcal{T}$ be the resulting set of tours.
10: **return** $(F_1 \cup F_2, \mathcal{T}, C)$.

$$\text{(PC-ST}_{\text{LP}}) \quad \begin{aligned} \min \quad & \sum_{e \in E \cup E'} c(e)y(e) + \sum_{N \subseteq \mathcal{C}} \Big(\sum_{v \in N} p'(v)\Big) z(N) \\ \text{s.t.} \quad & \sum_{e \in \delta_{G'}(S)} y(e) \; + \sum_{N \subseteq \mathcal{C}:\; S \cap \mathcal{C} \subseteq N} z(N) \; \geq 1 \qquad \forall S \subseteq V,\; S \cap \mathcal{C} \neq \emptyset \\ & y \geq 0 \end{aligned}$$

Here, $\delta_{G'}(S)$ denotes the cut in G' induced by the vertex set S, i.e., the set of all edges of G' that have one endpoint in S and one endpoint outside of S. The intuition for the LP formulation is the following: Given a feasible solution to prize-collecting Steiner tree, define $z(N) = 1$ for the set N of clients that are not connected to the Steiner tree, and $z(N) = 0$ for all other sets of clients. Moreover, set $y(e) = 1$ if edge e is in the Steiner tree, $y(e) = 0$ otherwise. The inequalities follow from the fact that any cut that separates a served terminal from the root has to be crossed by at least one edge of the tree.

Lemma 3.1. $\mathrm{OPT}(\text{PC-ST}_{\text{LP}}) \leq \frac{1}{2} OPT(\text{PC-CLR})$

Proof. Let $(F, \mathcal{T}, C)$ be an optimal solution to PC-CLR. Construct a solution $(\tilde{z}, \tilde{y})$ to PC-ST$_{\text{LP}}$ by setting $\tilde{z}(N^* := \mathcal{C} \setminus C) = 1$ and $\tilde{z}(N) = 0$ for all other $N \subseteq \mathcal{C}$, and $\tilde{y}(\{r, w\}) = 1$ for all $w \in F$, $\tilde{y}(\{r, w\}) = 0$ for all $w \in \mathcal{F} \setminus F$, and $\tilde{y}(e) = \frac{1}{2}|\{T \in \mathcal{T} : e \in E(T)\}|$. It is easy to observe that the constructed solution $(\tilde{y}, \tilde{z})$ has cost $\frac{1}{2}\sum_{w \in F} \phi(w) + \frac{1}{2}\sum_{v \in \mathcal{C} \setminus C} p(v) + \frac{1}{2}\sum_{T \in \mathcal{T}} c(T)$.

It remains to show that $(\tilde{z}, \tilde{y})$ is feasible for PC-ST$_{\text{LP}}$. So let S denote an arbitrary subset of V with $S \cap \mathcal{C} \neq \emptyset$. If S contains an open facility w, then $\{r, w\} \in \delta_{G'}(S)$, and by definition of $\tilde{y}$, the constraint for S is fulfilled. Else, if $S \cap C = \emptyset$, then S contains only unserved clients and the set $\{N \subseteq \mathcal{C} : \; S \cap \mathcal{C} \subseteq N\}$ contains N^*. Hence, by definition of $\tilde{z}$, the constraint for S is satisfied as well. Finally, if S does not contain an open facility and $S \cap C \neq \emptyset$, then there is a a client $v \in C \cap S$ connected to an open facility outside of S by a tour. At least two edges of this tour lie in the cut $\delta_{G'}(S)$, hence the constraint for S is again satisfied by definition of $\tilde{y}$. □

Theorem 3.2. Using the algorithm of [21] in its Steiner tree phase, Algorithm 2 is a $(2 + 2\rho_{\text{PC-UFL}})$-approximation algorithm for PC-CLR. The solution it produces fulfills the single-assignment property. If $d_v \leq u$ for all $v \in \mathcal{C}$, it furthermore fulfills the single-tour property.

Proof. Since the algorithm uses the large demand and merge phases of Algorithm 1, the claims of the theorem regarding single-assignment and single-tour properties follow directly from Theorem 2.4.

Moreover, by Lemma 2.3, the cost of the solution returned in the merge phase is bounded by

$$\begin{aligned} 2c(S) + 2\tilde{c}(U) + \phi(U) + \sum_{v \in \mathcal{C} \setminus C} p(v) &\leq 2c(S) + \sum_{v \in \mathcal{C} \setminus C_1} 2p'(v) + 2\tilde{c}(U) + \phi(U) + \sum_{v \in \mathcal{C} \setminus C_2} p(v) \\ &\leq 2 \cdot 2 \text{ OPT}(\text{PC-ST}_{\text{LP}}) + 2\,\rho_{\text{PC-UFL}} \cdot \text{OPT}(\text{PC-UFL}) \\ &\leq (2 + 2\rho_{\text{PC-UFL}}) \text{ OPT}, \end{aligned}$$

where the second to last inequality stems from the fact that the algorithm by [21] is a 2-approximation, and the last from Lemma 3.1. □

Similar to Section 2.4, we can replace the algorithm for PC-UFL by shortest path computations for the case of PC-MDCVR, which solve this subproblem to optimality: A client is connected to a facility if and only if the shortest path distance to its closest facility is no greater than its penalty. This yields an improved approximation ratio for PC-MDCVR.

Theorem 3.3. For the case $\phi \equiv 0$, PC-UFL can be solved exactly by shortest path computations. Thereby, Algorithm 2 becomes a 4-approximation algorithm for PC-MDCVR. The solution it produces fulfills the single-assignment property. If $d_v \leq u$ for all $v \in \mathcal{C}$, it furthermore fulfills the single-tour property.

4. Group Location Routing

We now consider a group version of location routing (G-CLR) where the set of clients is partitioned into groups $\mathcal{C}_1, \ldots, \mathcal{C}_k$, with $\mathcal{C} = \bigcup_{i=1}^k \mathcal{C}_i$ and only one client from each group needs to be served. The uncapacitated version of this problem was studied by [20], who give a $(2 - \frac{1}{|V|-1})L$-approximation algorithm with L being the cardinality of the largest group. Their idea is to solve an LP relaxation of the problem and use the resulting fractional solution to decide which client is to be served from each group. We extend this approach to the capacitated case which is significantly more complex: In the absence of vehicle capacities, facility opening costs can be transferred to edges of the graph, i.e. location routing is equivalent to multi-depot vehicle routing in this case. In contrast to [20], our LP relaxation has to explicitly incorporate the facility location aspect of the problem.

The dependence of our approximation factor on the parameter L gives rise to the question whether there is a constant factor approximation algorithm for G-CLR that is independent of any parameters in the input. At the end of this section, we answer this question in the negative by showing that there is no $o(\log(k))$-approxmation algorithm for G-CLR.

4.1. LP relaxation

In order to obtain an approximation for G-CLR, we describe how to transform a solution of G-CLR into a multi-commodity flow variable assignment on the arcs and vertices of a directed graph. We then prove a set of valid inequalities fulfilled by all assignments obtained from feasible G-CLR solutions. The LP relaxation resulting from these inequalities can be used to decide on a set of representatives, one for each client group. Replacing each group by its representative, we obtain an instance of (non-group) CLR which can be approximated by an adaption of Algorithm 1 with the spanning tree replaced by a Steiner tree. We will show that the resulting solution to G-CLR is a $4.38L$-approximation.

While the problem remains based on an undirected graph, it is more convenient to consider its directed equivalent in our LP relaxation: We replace each undirected edge e by two oppositely directed arcs a_e^+ and a_e^- with costs $c(a_e^+) = c(a_e^-) = c(e)$ and denote the set of all such arcs by A. We start constructing a multi-commodity flow on the edges in A from a given (undirected) solution of G-CLR by fixing an arbitrary orientation for every tour. Let $y(a)$ be the number of tours using arc $a \in A$. Let $T_{v\leftarrow w}(a)$ and $T_{v\rightarrow w}(a)$ be the index sets of all tours that serve client $v \in \mathcal{C}$ from facility $w \in \mathcal{F}$ with an occurrence of arc $a \in A$ on the path from w to v or, respectively, from v to w. Accordingly, define variables $x_{v\leftarrow w}(a) = \sum_{i \in T_{v\leftarrow w}(a)} x_{vi}$ and $x_{v\rightarrow w}(a) = \sum_{i \in T_{v\rightarrow w}(a)} x_{vi}$ for all arcs, where the x_{vi} are the demand assignments introduced at the beginning of Section 2. Finally, for each facility $w \in \mathcal{F}$, let $z(w) = 1$ if w is open and $z(w) = 0$ otherwise.

The values $x_{v\leftarrow w}(a)$ and $x_{v\rightarrow w}(a)$ can be interpreted as multi-commodity flow with two commodities $v \leftarrow w$ and $v \rightarrow w$ for each pair $v \in \mathcal{C}$ and $w \in \mathcal{F}$, respectively. The first commodity corresponds to goods transported from facility w to client v, the second commodity $v \rightarrow w$ emulates the empty truck capacity on the tour returning from v to w. We define the flow balance of node $v \in V$ with respect to commodity $h \in \{v \leftarrow w, v \rightarrow w \ : \ v \in \mathcal{C}, w \in \mathcal{F}\}$ as $b_h(v) := \sum_{a\in\delta^+(v)} x_h(a) - \sum_{a\in\delta^-(v)} x_h(a)$.

First observe that the total amount of flow on any arc can at most be the capacity u times the number of tours using the arc, i.e.,

$$\sum_{v\in\mathcal{C}} \sum_{w\in\mathcal{F}} (x_{v\leftarrow w}(a) + x_{v\rightarrow w}(a)) \leq u y(a) \quad \forall a \in A. \tag{1}$$

Furthermore, we obtain

$$\sum_{v\in\mathcal{C}_i} \sum_{w\in\mathcal{F}} \tfrac{1}{d_v}(x_{v\leftarrow w}(a) + x_{v\rightarrow w}(a)) \leq y(a) \quad \forall a \in A, i \in \{1,\dots,k\} \tag{2}$$

by observing that the left hand side of the equation is at most 1 per tour that is using the arc: Only one client v in a group is served, only d_v units are transported to this client in total, and in any tour, each arc occurs either before or after v but never both.

By construction of x, flow conservation holds for each commodity at all nodes that neither correspond to its facility nor to its client. Furthermore, at clients $v \in \mathcal{C}$, the value of any commodity $v \rightarrow w$ for some $w \in \mathcal{F}$ leaving the client equals the value of $v \leftarrow w$ entering it:

$$b_{v\rightarrow w}(p) = 0 = b_{v\leftarrow w}(p) \qquad \forall v \in \mathcal{C}, w \in \mathcal{F}, p \in V \setminus \{v, w\} \tag{3}$$

$$b_{v\leftarrow w}(v) = -b_{v\rightarrow w}(v) = b_{v\rightarrow w}(w) = -b_{v\leftarrow w}(w) \qquad \forall v \in \mathcal{C}, w \in \mathcal{F} \tag{4}$$

Moreover, as one client from every group needs to be served, the variables fulfill

$$\sum_{v\in\mathcal{C}_i} \sum_{w\in\mathcal{F}} \tfrac{1}{d_v} b_{w\rightarrow v}(v) = 1 \qquad \forall i \in \{1,\dots,k\}. \tag{5}$$

Finally, at most d_v units of flow are sent from an open facility to client v and thus

$$\sum_{v\in\mathcal{C}_i} \tfrac{1}{d_v} b_{v\leftarrow w}(v) \leq z_w \quad \forall w \in \mathcal{F}, i \in \{1,\dots,k\}. \tag{6}$$

We conclude that the value of an optimal solution to the group location routing problem is at least the value of an optimal solution of the following LP.

$$(G\text{-CLR}_{\mathrm{LP}}) \qquad \begin{array}{rl} \min & \sum_{a\in A} c(a)y(a) + \sum_{w\in\mathcal{F}} \phi_w z_w \\ \text{s.t.} & x, y, z \ \text{ fulfill } \ (1) - (6) \\ & x, y, z \geq 0 \end{array}$$

Let (x^*, y^*, z^*) be an optimal solution to G-CLR$_{\text{LP}}$. For $i \in \{1, \ldots, k\}$, let $r_i \in \mathcal{C}_i$ be a client with $\sum_{w \in \mathcal{F}} \frac{b^*_{v \leftarrow w}}{d_v}$ maximum over all $v \in \mathcal{C}_i$. We now define the set of group representatives as $R := \{r_1, \ldots, r_k\}$. The following inequality will be useful for deriving lower bounds on OPT(G-CLR$_{\text{LP}}$).

Lemma 4.1. Let $L := \max\{|\mathcal{C}_i| : i \in \{1, \ldots, k\}\}$. Then $L \cdot \sum_{w \in \mathcal{F}} b^*_{r_i \leftarrow w} \geq d_{r_i}$ for all $i \in \{1, \ldots, k\}$.

Proof. By (5), in each group C_i, there has to be at least one client $v \in C_i$ with $\sum_{w \in \mathcal{F}} \frac{b^*_{v \leftarrow w}}{d_v} \geq \frac{1}{L}$, and thus this inequality holds for r_i in particular. □

Now denote the instance of (non-group) CLR defined by the set of representatives R by CLR(R). Consider the following LP relaxation for the uncapacitated facility location problem arising from CLR(R) as described in Lemma 2.1. We will use it to derive a lower bound on the value of an optimal solution to G-CLR$_{\text{LP}}$.

$$
(\text{UFL}_{\text{LP}}(R)) \quad
\begin{aligned}
\min \quad & \sum_{v \in R} \sum_{w \in \mathcal{F}} \tilde{c}_{vw} x_{vw} + \sum_{w \in \mathcal{F}} \phi_w z_w & \\
\text{s.t.} \quad & \sum_{w \in \mathcal{F}} x_{vw} \geq d_v & \forall v \in R \\
& \tfrac{1}{d_v} x_{vw} \leq z_w & \forall v \in R, w \in \mathcal{F} \\
& x, z \geq 0 &
\end{aligned}
$$

Lemma 4.2. OPT(UFL$_{\text{LP}}(R)$) $\leq L \cdot$ OPT(G-CLR$_{\text{LP}}$).

Proof. Consider the solution $(\tilde{x}, \tilde{z})$ to UFL$_{\text{LP}}(R)$ obtained by setting $\tilde{z}_w = L \cdot z^*_w$ and $\tilde{x}_{vw} = L \cdot b^*_{v \leftarrow w}(w)$ for all $v \in R, w \in \mathcal{F}$. Observe that by Lemma 4.1, we have for each representative r_i

$$\sum_{w \in \mathcal{F}} \tilde{x}_{r_i w} = L \cdot \sum_{w \in \mathcal{F}} b^*_{r_i \leftarrow w}(r_i) \geq d_{r_i}.$$

Together with (6), this immediately implies that $(\tilde{x}, \tilde{z})$ is a feasible solution to UFL$_{\text{LP}}$. The flow of each commodity $v \leftarrow w$ $(v \rightarrow w)$ can be decomposed into flow on v-w-paths (w-v-paths), each of which has at length at least c_{vw} by triangle inequality. Combining this with (1), we obtain

$$
\begin{aligned}
\sum_{a \in A} c(a) y^*(a) &\geq \sum_{a \in A} \tfrac{c_a}{u} \sum_{v \in \mathcal{C}} \sum_{w \in \mathcal{F}} (x^*_{v \leftarrow w}(a) + x^*_{v \rightarrow w}(a)) \geq \sum_{v \in \mathcal{C}} \sum_{w \in \mathcal{F}} \tfrac{2}{u} c_{vw} b^*_{v \leftarrow w}(v) \\
&\geq \tfrac{1}{L} \cdot \sum_{v \in R} \sum_{w \in \mathcal{F}} \tfrac{2}{u} c_{vw} \tilde{x}_{vw} = \tfrac{1}{L} \cdot \sum_{v \in R} \sum_{w \in \mathcal{F}} \tilde{c}_{vw} \tilde{x}_{vw}.
\end{aligned}
$$

Furthermore, $L \cdot \sum_{w \in \mathcal{F}} \phi_w z^*_w = \sum_{w \in \mathcal{F}} \phi_w \tilde{z}_w$ by construction, which implies OPT(UFL$_{\text{LP}}$) $\leq L \cdot$ OPT(G-CLR$_{\text{LP}}$). □

A second lower bound can be obtained from the LP relaxation of a Steiner tree instance defined similar to that in Section 3. Again, we consider the graph $G' = (V \cup \{r\}, E \cup E')$ with $E' = \{\{r, w\} : w \in \mathcal{F}\}$ as constructed in Lemma 2.2. We then extend the cost function c to E' by defining cost $c_{rw} = \frac{1}{2} \phi_w$ for each $w \in \mathcal{F}$. We now consider the undirected cut relaxation of the Steiner tree instance on G' with terminals $R \cup \{r\}$.

$$(\mathrm{ST_{LP}}(R)) \qquad \begin{aligned} \min \quad & \sum_{e \in E \cup E'} c(e)y(e) \\ \text{s.t.} \quad & \sum_{e \in \delta_{G'}(S)} y(e) \geq 1 \quad \forall S \subseteq V,\ S \cap R \neq \emptyset \\ & y \geq 0 \end{aligned}$$

Lemma 4.3. $\mathrm{OPT}(\mathrm{ST_{LP}}(R)) \leq \frac{1}{2}L \cdot \mathrm{OPT}(G\text{-}\mathrm{CLR_{LP}})$

Proof. Consider the solution $\tilde{y}$ to $\mathrm{ST_{LP}}(R)$ obtained by setting $\tilde{y}(\{v,w\}) = \frac{1}{2}L \cdot (y^*(vw) + y^*(wv))$ for all $v, w \in V$ and $\tilde{y}(\{r,w\}) = L \cdot z^*_w$ for all $w \in \mathcal{F}$. Let $S \subseteq V$ with $r_i \in S$ for some $i \in \{1, \dots k\}$. By flow conservation and (6) we obtain

$$\sum_{a \in \delta^+(S)} x^*_{r_i \to w}(a) + \sum_{w \in S} d_{r_i} z_w \geq b^*_{r_i \to w}(r_i) \quad \text{and} \quad \sum_{a \in \delta^-(S)} x^*_{r_i \leftarrow w}(a) + \sum_{w \in S} d_{r_i} z^*_w \geq b^*_{r_i \to w}(r_i),$$

where $\delta^+(S) = \{vw \in A : v \in S, w \in V \setminus S\}$ and $\delta^-(S) = \{vw \in A : v \in V \setminus S, w \in S\}$. By construction of $\tilde{y}$ and inequality (2) we obtain

$$\begin{aligned} \sum_{e \in \delta_{G'}(S)} \tilde{y}(e) &= \frac{1}{2}L \left(\sum_{a \in \delta^+(S)} y^*(a) + \sum_{a \in \delta^-(S)} y^*(a) \right) + \sum_{w \in S} z^*_w \\ &\geq \frac{L}{2d_{r_i}} \left(\sum_{a \in \delta^+(S)} x^*_{r_i \to w}(a) + \sum_{a \in \delta^-(S)} x^*_{r_i \leftarrow w}(a) + 2 \cdot \sum_{w \in S} d_{r_i} z^*_w \right) \\ &\geq \tfrac{L}{d_{r_i}} \cdot b^*_{r_i \to w}(r_i). \end{aligned}$$

The last expression is at least 1 by Lemma 4.1. Thus, $\tilde{y}$ is a feasible solution to $\mathrm{ST_{LP}}(R)$ with $\sum_{e \in E \cup E'} c(e)\tilde{y}(e) = \frac{1}{2}L \left(\sum_{a \in A} c(a)y^*(a) + \sum_{w \in \mathcal{F}} \phi_w \right) = \frac{1}{2}L \cdot \mathrm{OPT}(G\text{-}\mathrm{CLR_{LP}})$. ☐

Remark 4.4. The LP relaxation presented in this section also yields an alternative proof of the minimum spanning tree lower bound in Lemma 2.2 for the non-group case, using the bidirected cut formulation of the spanning tree polytope. However, the direct and combinatorial proof of Lemma 2.2 given in Section 2.1 appears to be more intuitive and elegant.

4.2. Algorithm

Lemma 4.2 and Lemma 4.3 immediately lead to a $4.38L$-approximation algorithm for G-CLR: Compute an optimal solution to G-$\mathrm{CLR_{LP}}$, obtain a set of representatives R from this solution and compute an approximation to the resulting instance $\mathrm{CLR}(R)$ with Algorithm 1, using an LP-based Steiner tree 2-approximation algorithm instead of a minimum spanning tree computation.

Theorem 4.5. Algorithm 3 is a $4.38L$-approximation for G-CLR. There is a $4L$-approximation for G-MDCVR.

Proof. The cost of the Steiner tree computed by the algorithm of [21] is a at most $2 \cdot \mathrm{OPT}(\mathrm{ST_{LP}}(R))$. The UFL solution U computed in Algorithm 1 approximates the opening cost of an optimal solution to $\mathrm{UFL_{LP}}(R)$ by γ, and its connection cost by $(1 + 2e^{-\gamma})$, because the LP relaxation is equivalent to the one used in the algorithm of [7]. Thus, Lemmas 4.2 and 4.3 yield $2c(S) + 2\tilde{c}(U) + \phi(U) \leq 2 \cdot \mathrm{OPT}(\mathrm{ST_{LP}}(R)) + \gamma \cdot \mathrm{OPT}(\mathrm{UFL_{LP}}(R)) \leq 4.38L \cdot \mathrm{OPT}(\mathrm{GCLR})$. ☐

Algorithm 3 Algorithm for GCLR.

Input: An instance of G-CLR.
Output: A feasible solution to G-CLR.

1: Compute an optimal solution (x^*, y^*, z^*) to G-CLR$_{\text{LP}}$.
2: **for all** $i \in \{1, \ldots, k\}$ **do**
3: Let $r_i \in \mathcal{C}_i$ be a client with $\sum_{w \in \mathcal{F}} \frac{b^*_{v \leftarrow w}}{d_v}$ maximum over all $v \in \mathcal{C}_i$.
4: $R = R \cup \{r_i\}$
5: **end for**
6: Construct the graph G' with extended edge costs $c_{rw} = \frac{1}{2}\phi_w$ for $w \in \mathcal{F}$.
7: Apply the algorithm of Goemans and Williamson to obtain a Steiner tree S with terminal set $R \cup \{r\}$ in G'.
8: Apply Algorithm 1 on the instance CLR(R) with the minimum spanning tree computed in the tree phase replaced by the Steiner tree S.
9: Return the computed solution.

4.3. Lower bound on the approximability

Observing that the approximation guarantee of Algorithm 3 depends on the cardinality of the largest group, it is natural to ask whether the group version of CLR is indeed considerably harder than the standard version or whether there is a constant factor approximation whose performance is independent of any instance parameters. We answer this question by showing that there is no approximation algorithm for G-CLR with a factor better than $\mathcal{O}(\log(k))$.

In fact, the inapproximability result already holds for the special case of G-CLR with unit demands and unit capacity, which corresponds to the group version of (metric) uncapacitated facility location (G-UFL), as well as for the uncapacitated case considered in [20]. It is derived by a straightforward reduction from unweighted set cover.

Proposition 4.6. There exists a constant $\alpha > 0$ such that there is no $\alpha \log(k)$-approximation for G-UFL, unless $P = NP$.

Proof. We reduce the unweighted set cover problem, for which the same $\log(n)$-approximability-threshold has been proven by [19], to G-UFL. An instance of unweighted set cover consists of a ground set H and a set system $\mathcal{S} \subseteq 2^H$ together with costs c_S for every $S \in \mathcal{S}$. The task is to choose a subset $\mathcal{S}'$ of $\mathcal{S}$ such that every element of the ground set is covered, i.e., $\bigcup_{S \in \mathcal{S}'} S = H$, while minimizing the total cost $\sum_{S \in \mathcal{S}'} c_S$.

We create a G-UFL instance by introducing a facility w_S for each $S \in \mathcal{S}$ and setting $\phi(w_S) := c_S$. For every $h \in H$ and every $S \in \mathcal{S}$ with $h \in S$ we introduce a client v_{hS}. We also introcude a client group $\mathcal{C}_h$ for each element $h \in H$ of the ground set and let it contain all clients v_{hS}. Finally, we set $c_{w_S, v_{hS'}} = 0$, whenever $S = S'$, and to ∞ otherwise.

Note that any feasible solution to this G-UFL instance with finite costs corresponds to a feasible solution to set cover with the same costs, by selecting the sets corresponding to open facilities. As for every client group there is an open facility with connection cost 0 to one of its members, every set is covered. Likewise, every feasible solution to set cover induces a solution to G-UFL by opening the facilities corresponding to the chosen sets. Since every element of the ground set is covered, for every client group there is a member that has connection cost 0 to an open facility. Thus, any γ-approximation for G-UFL (or the group version of UFL) immediately implies a γ-approximation for set cover. Choosing the same α as used in [19] for set cover, we conclude that there is no $\alpha \log(k)$-approximation for G-UFL (unless $P = NP$), since this would imply a $\alpha \log(|H|)$-approximation for set cover (note that $|H|$ is the number of groups in the constructed G-UFL instance). $\square$

Corollary 4.7. There exists a constant $\alpha > 0$ such that there is no $\alpha \log(k)$-approximation for G-CLR (even if $u = \infty$), unless $P = NP$.

Proof. As G-UFL is a special case of G-CLR, the inapproximability also holds for the latter. The reduction also works if $u = \infty$, as connection costs are either 0 or ∞ and so serving all clients at a facility on one tour instead of serving them separately does not change the costs. □

5. Location Routing with Cross-Docking

A major trade-off in classic vehicle routing applications is good capacity utilization versus low cost of each tour conducted. Especially in applications where clients with small demands are located far away from facilities, significant cost savings can be realized by allowing *consolidation tours*. In such a tour, a vehicle is positioned at a client node to collect goods from other vehicles passing through. Then, it starts on its own tour to distribute the goods collected. Essentially, the demand of a tour of clients is consolidated at one node and forwarded to facilities via other tours from there. The necessitated process of loading goods from one vehicle to another at a client node is commonly referred to as *cross-docking*. The example in Figure 1 shows that cross-docking may indeed lead to cost savings.

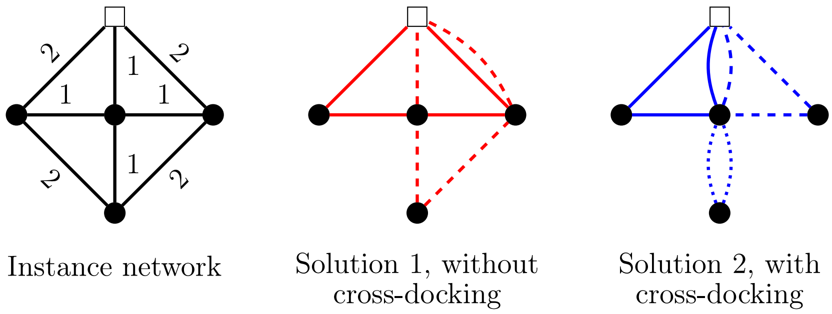


Figure 1. A CLR instance with $u = 5$. The numbers on the edges indicate the edge costs. The demand at the central client is 1, the demand at the other clients is 3. The optimal routing scheme in Solution 1 without cross-docking has total cost 12. The routing scheme in Solution 2 uses cross-docking to consolidate the tours at the central vertex. Its total cost is 10.

Formally, a *solution with cross-docking* to CLR is again a tuple $(F, \mathcal{T})$, where $F \subseteq \mathcal{F}$ is a set of open facilities and $\mathcal{T} = \{T_1, \ldots, T_k\}$ is a set of tours, but $\mathcal{T} = \mathcal{T}_{\mathrm{F}} \,\dot\cup\, \mathcal{T}_{\mathrm{H}}$ is now partitioned into a set of *facility tours* $\mathcal{T}_{\mathrm{F}}$ and a set of *consolidation tours* $\mathcal{T}_{\mathrm{H}}$. We now require that (1a) every facility tour visits an open facility, (1b) in every consolidation tour $T_i \in \mathcal{T}_{\mathrm{H}}$, exactly one client $h_i \in V(T_i)$ is designated as the *hub* consolidating the hub-demand $d_i^h := \sum_{v \in \mathcal{C}} x_{vi}$ of all clients served by the tour, (2) the demand of every client (including the additional demand occurring if the client is the hub of one or more consolidation tours) is served by the tours by which it is visited, and (3) the demand served by a tour does not exceed u. More precisely, there are non-negative values x_{vi} such that $x_{vi} > 0$ only if client v is visited by tour i and $\sum_{i=1}^{k} x_{vi} = d_v + \sum_{T_i \in \mathcal{T}_H : v = h_i} d_i^h$ for all $v \in \mathcal{C}$, and $\sum_{v \in \mathcal{C}} x_{vi} \leq u$ for all $i \in \{1, \ldots, k\}$. We say that a solution fulfills the *single-vehicle-to-client* property if each client's demand arrives on a single vehicle, which can be important in practice. Note that, in contrast to the single-tour property, single-vehicle-to-client deliveries still can be split up on earlier segments of the transportation route or even originate from distinct facilities.

5.1. Algorithm

We now describe how to adapt Algorithm 1 in order to allow for cross-docking. As before, we start by computing a solution to a UFL instance as defined in Lemma 2.1 and a minimum spanning tree for the modified graph G' as defined in Lemma 2.2. Then, we modify our rerouting procedure as stated formally in Algorithm 4.

Algorithm 4 Algorithm for CLR with cross-docking.

Input: An instance of CLR.
Output: A feasible solution to CLR with cross-docking.

UFL phase:
Create an UFL instance with edge costs $\tilde{c} = \frac{2}{u}c$ as described in Lemma 2.1.
Apply the 1.5-approximation algorithm of Byrka and Aardal on this instance and let F_1 be the set of facilities opened in the resulting UFL solution.
Run tree and large demand phase of Algorithm 1.
Merge phase:
for all $z \in F_2$ **do**
    while $D_z > u$ **do**
        Let $v \in V(S_z)$ such that $D_v > u$ but $D_w \leq u$ for all children w of v.
        Let $I = \{V(S_w) \, : \, w \text{ is a child of } v\} \cup \{\{v\}\}$.
        For every $R \in I$ find a pair (v_R, z_R) such that $v_R \in V(R)$ and $z_R \in F_1 \cup F_2$ and $c_{v_R z_R}$ is minimal.
        Order the sets in I non-decreasingly by $c_{v_R z_R}$ and include the first $\lfloor \frac{D_v}{u} \rfloor$ sets in I_{t}. Let $I_{\mathrm{s}} := I \setminus I_{\mathrm{t}}$.
        for all $R \in I_{\mathrm{t}}$ **do**
            Construct a tour visiting z_R, v and all vertices in R by adding $v_R z_R$ to the tree and then doubling edges and short-cutting.
            Add the tour to $\mathcal{T}_{\mathrm{F}}$ and remove the subtrees corresponding to the elements of R from S.
        end for
        for all $R \in I_{\mathrm{s}}$ **do**
            Construct a tour visiting v and all vertices in R by doubling edges and short-cutting.
            Add the tour to $\mathcal{T}_{\mathrm{H}}$ with hub v and remove the subtrees corresponding to the elements of R from S.
        end for
    end while
    Construct a tour from S_z by doubling all edges and short-cutting.
    Add the tour to $\mathcal{T}_{\mathrm{F}}$.
end for
Run clean-up phase of Algorithm 1.

As in Section 2.2, we consider a node v with $D_v > u$ but $D_w \leq u$ for all children w of v and let I be the set containing all subtrees S_w, with w being a child of v, and $\{v\}$ itself. For each of these sets $R \in I$, we determine a node v_R with cheapest connection cost $c_{v_R z_R}$ to an open facility z_R. We order the sets $R \in I$ non-decreasingly by $c_{v_R z_R}$ and define the set of *sink trees* I_{t} as the first $\lfloor \frac{D_v}{u} \rfloor$ elements of I. The remaining elements $I \setminus I_{\mathrm{t}}$ comprise the set of *source trees* I_{s}. Each sink tree $R \in I_t$ is turned into a facility tour by doubling edges and inserting the open facility closest to v_R, paying at most twice the tree edges plus the connection cost of v_R to its facility. Each source tree is turned into a consolidation tour with hub v by doubling the edges and short-cutting.

Note that by this construction, each facility tour visits v. Hence, any spare capacity on a facility tour can be filled by hub demands ensuing at v from consolidation tours. Furthermore, the sum of all demands that cannot be served by the facility tours constructed is strictly less than u.

5.2. Analysis

We first point out that our lower bounds from Section 2.1 remain valid when allowing cross-docking. These results can easily be obtained by slight modification of the corresponding proofs of Lemma 2.1 and Lemma 2.2, respectively.

Lemma 5.1. Consider a UFL instance as defined in Lemma 2.1. The cost of its optimal solution (w.r.t. $\tilde{c}$) is at most the cost of an optimal solution to CLR with cross-docking (w.r.t. c).

Proof. Consider an optimal solution $(F, \mathcal{T})$ of CLR with cross-docking and demand assignments x_{vi}. As in the proof of Lemma 2.1, we can construct a flow f from the CLR solution as follows. For every client $v \in \mathcal{C}$ and each tour $S_i \in \mathcal{T}$ serving v, partition S_i into two paths from the facility or hub of the tour to the client and send x_{vi} units of flow along either path. Since flow is sent along two paths for every client/tour pair and all hub demand is forwarded along further tours to facilities, the net flow transported to any client $v \in \mathcal{C}$ equals $2d_v$. We can thus apply flow decomposition on f to obtain a set of client-facility paths $\mathcal{P}_{\mathrm{P}}$ and cycles $\mathcal{P}_{\mathrm{C}}$, respectively, with corresponding flow values f_P for every $P \in \mathcal{P}_{\mathrm{P}} \cup \mathcal{P}_{\mathrm{C}}$. On this flow decomposition, we can apply the same arguments as in the proof of Lemma 2.1. $\square$

Lemma 5.2. The cost of a minimum spanning tree in the graph G' w.r.t. costs c' as defined in Lemma 2.2 is a lower bound on the cost of an optimal solution to CLR with cross-docking (w.r.t. c).

Proof. Let $(F, \mathcal{T})$ be a feasible solution to CLR with cross-docking. Note that $S = \bigcup_{T \in \mathcal{T}} T \cup E'$ spans all vertices of the graph G' since for every client there is a path from a facility to this client along edges used in the tours. We can modify S such that it spans the graph G' but every facility opened in the CLR solution is incident to at most one edge in E by applying the technique from the proof of Lemma 2.2 on the set of facility tours. This set contains a spanning tree of cost at most the cost of the CLR solution. $\square$

It turns out that guaranteeing demand u on each of the tours constructed in the rerouting procedure yields an improved approximation guarantee for the merge phase of our algorithm.

Lemma 5.3. The merge phase of Algorithm 4 constructs a solution to CLR with cross-docking with cost at most $2c'(S) + \tilde{c}(U) + \phi(U)$ from the spanning tree S and the UFL solution U.

Proof. Observe that each facility tour constructed in the inner loop serves a total demand of u. Thus, the central inequality in the proof Lemma 2.3 changes to $\sum_{x \in V(T)} \tilde{c}_{xy(x)} d_x \geq 2c_{wz'}$. Accordingly, the connection cost of the UFL solution is paid only once. $\square$

Intuitively, the improved bound in Lemma 5.3 arises from the tight capacity utilization of vehicles that are paid for by the UFL solution. We immediately obtain a better approximation guarantee for Algorithm 4 when using the 1.5-approximation of [7] for constructing the UFL solution U.

Theorem 5.4. Algorithm 4 is a 3.5-approximation algorithm for CLR with cross-docking. If $d_v \leq u$ for all $v \in \mathcal{C}$, the obtained solution satisfies the single-vehicle-to-client property.

Again, in the case of MDCVR with $\phi \equiv 0$, we can apply shortest path computations to solve the UFL instance exactly.

Theorem 5.5. When solving the UFL instance by shortest path computation, Algorithm 4 is a 3-approximation algorithm for MDCVR with cross-docking. If $d_v \leq u$ for all $v \in \mathcal{C}$, the obtained solution satisfies the single-vehicle-to-client property.

We remark that, while Algorithm 1 produces a solution without cross-docking, its approximation factor still holds for the case where cross-docking is allowed as all lower bounds used in Theorem 2.4 remain valid. Thus, we obtain the following bounds on the improvements realizable by cross-docking in CLR and MDCVR.

Corollary 5.6.

(1) Algorithm 1 is a 4.38-approximation for CLR with cross-docking and a 4-approximation for MDCVR with cross-docking. The produced solution fulfills the single-assignment property. If $d_v \leq u$ for all $v \in \mathcal{C}$, it fulfills the single-tour property.

(2) The value of an optimal solution for CLR without cross-docking is at most 4.38 times the value of a solution with cross-docking. The value of an optimal solution for MDCVR without cross-docking is at most 4 times the value of a solution with cross-docking.

We close this section by observing that the validity of the lower bounds extend to the cases of prize-collecting as well as group location routing with cross-docking. We can thus combine the merge phase of Algorithm 4 with the modifications introduced in Algorithm 2 for PC-CLR and Algorithm 3 for G-CLR, respectively.

Theorem 5.7. There is a $(2 + \rho_{\text{PC-UFL}})$-approximation algorithm for PC-CLR with cross-docking. There is a 3-approximation algorithm for PC-MDCVR with cross-docking.

Theorem 5.8. There is a $3.5\,L$-approximation algorithm for G-CLR with cross-docking. There is a $3\,L$-approximation algorithm for G-MDCVR with cross-docking.

6. Computational Study

In Section 2, we have proven that our polynomial time algorithm for CLR is guaranteed to compute solutions which are at most 4.38 times as expensive as the optimum. In this section, we shall see that the algorithm's performance in practice exceeds this theoretical worst-case estimate by far. We would like to emphasize that we do not expect our algorithm to compete with (meta-)heuristic approaches without an approximation guarantee. Rather, the question addressed in this computational study is how much solution quality on typical instances needs to be sacrificed in exchange for polynomial running time and a worst case performance guarantee across all instances.

For our experiments, we implemented Algorithm 1 with the following minor modifications: First, instead of using the bifactor approximation algorithm of Byrka and Aardal in the UFL phase, we implemented the greedy approximation algorithm of [28]. While the latter has a slightly worse approximation guarantee of 1.861, it is purely combinatorial, avoiding randomization and linear programming, and far easier to implement. Moreover, before applying Prim's algorithm (see e.g. [17]) in the tree phase, we set the opening costs of all facilities opened in the UFL phase to zero; doing so turns out to yield slightly improved results, while it does not interfere with our theoretical analysis of the algorithm. Finally, once the algorithm has computed all tours, we added an option to improve each single tour by solving the corresponding travelling salesman problem (TSP) using LKH, an implementation of the Lin-Kernighan heuristic described in [26].

Fact 1. Our implementation of Algorithm 1 has an approximation guarantee of 5.722.

Fact 2. The running time of our implementation of Algorithm 1 is $\mathcal{O}(n^2m)$, where n and m denote the number of clients and facilities, respectively.

Fact 1 results directly from Lemma 2.3 and the approximation factor of the greedy algorithm used in the UFL phase. The running time of the implementation is dominated by that of the UFL phase, cf. [28]. Moreover, experiments in [26] indicate that the practical running time of LKH is quite low (close to quadratic). Our study supports this observation, as the additional running time when employing the option for a-posteriori tour optimization by LKH turns out to be small, immeasurable on moderately sized instances.

We report results for two different sets of instances: The first, referred to as the benchmark set, comprises 45 instances appearing frequently in the location routing literature, see the references appearing below. Here, we compare our results with those obtained by recent (meta-)heuristic algorithms as well as best known solutions (bks) from the literature. While the benchmark instances are moderate in size (20–200 clients, 5–20 facilities), our second test set consists of 27 randomly generated instances which are considerably larger (up to 10000 clients and 1000 facilities). Our implementation was done in C++ using GCC 4.5 under SUSE Linux 11.3, and all computations were conducted on an Intel Core2 Duo E8400 processor at 3GHz with 4GB RAM.

6.1. Benchmark instances

Key properties of the benchmark instances used are listed in Table 1. The first 36 instances were introduced in [42], the last nine in [4]; we will refer to them as sets *TB* and *B*, respectively. While set *TB* is adopted as-is, our set *B* contains only those instances introduced in [4] which do not have a capacity limit on facilities, as only those mirror the location routing problem addressed here.

The best known solution values reported for *TB* were obtained in [39]. For *B*, some proven optima were already reported in [4], while the remaining instances were solved to proven optimality in [2], as reported in [14].

Table 2 contains gaps to bks and cpu times for our implementation of Algorithm 1, with and without a-posteriori optimization of tours using LKH, compared to those of four other algorithms for CLR: a greedy randomized adaptive search procedure (GRASP) proposed in [38]; a Lagrangean relaxation granular tabu search (LRGTS) developed in [39]; a two-phase tabu search (TS) studied in [42]; and finally an exact branch-and-cut-and-price approach (BCP) proposed in [2]. Results for algorithms GRASP and LRGTS are stated in [39] for all 45 benchmark instances, while results for TS and BPS are only available in the corresponding works for the instances in *TB* and *B*, respectively.

Please note that our algorithms, GRASP and LRGTS, TS, and BCP were tested on different machines, so the cpu times stated should not be compared directly. Since all tests were performed on modern desktop computers, however, we do believe that a comparison of the magnitudes of running times remains feasible.

On average, our approximation algorithm delivers solutions with cost about 19% above the bks value. This figure improves to 10% when LKH is used to optimize tours a-posteriori. Moreover, the running time of our algorithm is negligible on these instances, regardless of whether LKH is used or not. In comparison, the (meta-)heuristic algorithms GRASP, LRGTS, and TS compute solutions with objective 1–4% above that of bks on average, while their running times vary strongly from 1–7 seconds (TS) to up to 7 minutes (GRASP). The exact approach (BCP) is able to find optimal solutions for all instances in *B*, while its running time is naturally very high (up to several hours).

name	#facilities	#clients	∅ demand	vehicle capacity	bks value
111112	10	100	15.17	150	1468.40
111122	20	100	15.00	150	1449.20
111212	10	100	14.39	150	1396.46
111222	20	100	15.19	150	1432.29
112112	10	100	15.28	150	1167.53
112122	20	100	14.32	150	1102.70
112212	10	100	15.06	150	793.97
112222	20	100	14.73	150	728.30
113112	10	100	14.81	150	1238.49
113122	20	100	15.10	150	1246.34
113212	10	100	14.73	150	902.38
113222	20	100	14.78	150	1021.31
121112	10	200	14.95	150	2281.78
121122	20	200	15.15	150	2185.55
121212	10	200	14.81	150	2234.78
121222	20	200	14.94	150	2259.52
122112	10	200	15.24	150	2101.90
122122	20	200	14.47	150	1709.56
122212	10	200	14.69	150	1467.54
122222	20	200	15.21	150	1084.78
123112	10	200	15.13	150	1973.28
123122	20	200	14.66	150	1957.23
123212	10	200	15.09	150	1771.06
123222	20	200	15.29	150	1393.62
131112	10	150	14.79	150	1866.75
131122	20	150	14.93	150	1841.86
131212	10	150	15.02	150	1981.37
131222	20	150	14.71	150	1809.25
132112	10	150	14.95	150	1448.27
132122	20	150	14.75	150	1444.25
132212	10	150	14.91	150	1206.73
132222	20	150	15.15	150	931.94
133112	10	150	14.95	150	1699.92
133122	20	150	14.93	150	1401.82
133212	10	150	15.18	150	1199.51
133222	20	150	14.91	150	1152.86
Chr69-100x10	10	100	14.58	200	842.90*
Chr69-50x5	5	50	15.54	160	565.60*
Chr69-75x10	10	75	18.19	160	861.60*
Gas67-22x5	5	22	463.14	4500	585.11*
Gas67-29x5	5	29	439.66	4500	512.10*
Gas67-32x5	5	32	917.81	8000	562.20*
Gas67-32x5-2	5	32	917.81	11000	504.30*
Gas67-36x5	5	36	25.00	250	460.40*
Min92-27x5	5	27	311.48	2500	3062.00*

Table 1. Properties of benchmark instances and cost of a best known solution (bks, * denotes proven optimality). The bks values for the first 36 instances are from [39], those for the last nine from a series of papers by [2], [4], and [42].

Since gaps to bks for GRASP and LRGTS are no greater for the instances in *TB* than for those in *B*, where optimality has been proven, it seems reasonable to assume that the gap between bks and an optimum solution is generally small. In this case, our algorithm vastly outperforms its theoretical approximation guarantee of 5.722. When employing a simple post-optimization

instance	approx gap	approx cpu	approx+tsp gap	approx+tsp cpu	GRASP gap	GRASP cpu	LRGTS gap	LRGTS cpu	TS/BCP gap	TS/BCP cpu
111112	0.207	0.00	0.079	0.00	0.039	32.40	0.015	3.30	0.060	6.01
111122	0.235	0.00	0.117	0.00	0.054	40.70	0.016	6.50	0.057	5.71
111212	0.133	0.00	0.043	0.00	0.019	27.60	0.011	4.20	0.034	3.36
111222	0.342	0.00	0.246	0.00	0.035	36.20	0.008	7.40	0.055	5.52
112112	0.164	0.00	0.076	0.00	0.028	27.70	0.017	6.90	0.054	5.45
112122	0.133	0.00	0.095	0.01	0.019	34.30	0.012	6.80	0.027	2.66
112212	0.086	0.00	0.041	0.00	0.025	22.50	0.024	5.20	0.039	3.92
112222	0.119	0.00	0.070	0.00	0.027	37.30	0.020	5.90	0.017	1.68
113112	0.183	0.00	0.090	0.00	0.028	21.50	0.024	4.30	0.063	6.34
113122	0.201	0.00	0.131	0.00	0.021	36.00	0.008	6.30	0.023	2.26
113212	0.140	0.00	0.082	0.00	0.011	20.30	0.012	4.00	0.020	2.04
113222	0.166	0.00	0.126	0.00	0.004	38.40	0.007	4.90	0.023	2.05
131112	0.253	0.01	0.142	0.01	0.075	113.00	0.042	12.50	0.072	7.19
131122	0.230	0.01	0.110	0.01	0.026	161.40	0.018	18.50	0.028	2.77
131212	0.153	0.00	0.067	0.01	0.027	100.00	0.015	11.10	0.021	2.06
131222	0.206	0.01	0.102	0.01	0.026	132.40	0.006	15.80	0.025	2.53
132112	0.163	0.01	0.081	0.01	0.041	117.70	0.000	22.00	0.074	7.43
132122	0.301	0.01	0.230	0.02	0.009	166.10	0.034	28.00	0.024	2.39
132212	0.101	0.01	0.050	0.00	0.028	106.70	0.004	14.60	0.020	2.04
132222	0.170	0.00	0.123	0.01	0.010	142.40	0.005	13.70	0.018	1.75
133112	0.155	0.01	0.098	0.00	0.022	92.80	0.017	17.90	0.037	3.68
133122	0.127	0.01	0.075	0.01	0.017	128.40	0.016	18.50	0.062	6.17
133212	0.128	0.00	0.068	0.01	0.020	88.50	0.014	14.50	0.054	5.43
133222	0.081	0.00	0.029	0.01	0.068	134.90	0.008	14.30	0.026	2.55
121112	0.217	0.01	0.145	0.01	0.055	308.00	0.016	32.60	0.053	4.28
121122	0.139	0.01	0.050	0.02	0.047	410.00	0.010	39.60	0.012	1.20
121212	0.191	0.01	0.105	0.02	0.017	311.40	0.012	32.80	0.024	2.39
121222	0.225	0.02	0.122	0.02	0.042	418.90	0.004	40.20	0.047	4.26
122112	0.145	0.02	0.088	0.02	0.017	338.00	0.009	47.20	0.027	2.70
122122	0.179	0.02	0.125	0.02	0.057	370.00	0.017	59.30	0.045	4.53
122212	0.107	0.01	0.050	0.01	0.020	242.70	0.014	36.70	0.056	5.60
122222	0.119	0.01	0.049	0.00	0.010	308.50	0.005	38.70	0.026	2.60
123112	0.170	0.01	0.081	0.01	0.036	282.80	0.005	41.60	0.042	4.20
123122	0.126	0.01	0.050	0.02	0.068	399.20	0.015	51.80	0.023	2.31
123212	0.183	0.02	0.146	0.02	0.010	199.00	0.009	34.00	0.060	6.00
123222	0.182	0.01	0.134	0.01	0.011	296.30	0.005	43.20	0.015	1.52
Chr69-100x10*	0.283	0.00	0.108	0.00	0.022	25.50	0.000	28.20	0.000	13074.7
Chr69-50x5*	0.220	0.00	0.079	0.00	0.059	2.30	0.037	2.40	0.000	112.9
Chr69-75x10*	0.177	0.00	0.104	0.00	0.000	9.80	0.002	10.10	0.000	3413.5
Gas67-22x5*	0.244	0.00	0.021	0.00	0.000	0.20	0.004	0.20	0.000	6.0
Gas67-29x5*	0.279	0.00	0.165	0.00	0.006	0.40	0.000	0.40	0.000	178.2
Gas67-32x5*	0.245	0.00	0.179	0.00	0.017	0.60	0.040	0.60	0.000	63.4
Gas67-32x5-2*	0.205	0.00	0.123	0.00	0.000	0.50	0.001	0.50	0.000	117.9
Gas67-36x5*	0.448	0.00	0.094	0.00	0.000	0.80	0.035	0.70	0.000	2.9
Min92-27x5*	0.181	0.00	0.115	0.00	0.000	0.40	0.001	0.30	0.000	47.0
∅	**0.188**	0.01	**0.100**	0.01	**0.026**	128.54	**0.013**	17.96	**0.038** **0.000**	3.74 1890.69

Table 2. Gaps to best known solution (bks) and cpu times for various algorithms on benchmark instances (* signifies proven optimality of bks). Results for algorithm TS are only available for the first 36 instances, those for BCP only for the last nine, hence they share a column. The last row contains average values, with those for TS (first 36 instances) and BCP (last nine) one above the other.

step using LKH, it yields solutions within a factor of 1.25 of bks on all instances, within 1.1 on average. Moreover, its polynomial running time is reflected in very small cpu times on these benchmark instances. When compared to (meta-)heuristic algorithms, solution quality suffers only by a single-digit percentage on average, while cpu times are improved by several magnitudes. Moreover, recall that this improvement in running time comes in addition to the advantage of having a guarantee on solution quality across all possible instances, including malicious examples where (meta-)heuristics might perform very poorly. In light of its extremely fast running time, our algorithm can also be used to compute feasible start solutions for other search heuristics.

6.2. Larger, randomly generated instances

The extremely fast running time of our algorithm on benchmark instances, which are all of moderate size, suggests that our algorithm is suitable for larger instances as well. To the best of our knowledge, no instances of CLR which are significantly larger than those in the benchmark set have been solved in the literature; hence, we generated a random test set from three input parameters: size, facility opening cost, and vehicle capacity.

Instances were generated for three sizes: M (1000 clients, 100 facilities), L (5000, 500), and XL (10000, 1000). Facility opening costs were drawn uniformly at random from three different ranges: $[0; 100]$, $[100; 200]$, and $[200; 500]$. Vehicle capacities were set to either 9, 100, or 1000, while client demands were drawn uniformly at random from $[0; 10]$ in all cases. Finally, x- and y-coordinates for clients and facilities were drawn uniformly at random from $[0; 100]$, and Euclidean distances $d(i,j) := \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$ are used in all instances.

All possible combinations of the three input parameters yield 27 different instances, which we name by their size, indexed with their choice of facility opening cost and vehicle capacity. E.g., $\mathrm{M}_{2,2}$ is an instance with 1000 clients, 100 facilities, facility opening costs in $[100; 200]$, and vehicle capacity 100.

Key properties of the solutions computed by our algorithm, again with and without LKH, together with cpu times are depicted in Table 3. Cpu time for the largest instances is at most about twenty minutes. On average, using LKH to optimize tours a-posteriori reduces total cost by about 5%, while increasing cpu time by roughly 10%. Naturally, the effect of using LKH on both solution quality and cpu time is more significant when vehicle capacity is large (i.e., tours are long).

Since we did not compute lower bounds, we have no way to assess solution quality. However, we encourage the authors of other algorithms for CLR to perform experiments on our random test set, which we will gladly provide upon request, and compare their results to ours.

7. Summary & Outlook

Approximation algorithms combine efficient running times with provable a priori guarantees on solution quality. We applied this concept to several versions of capacitated location routing problems, which extend classical vehicle routing problems by depot location decisions. Variants of our algorithms also yield improved approximation guarantees for multi-depot capacitated vehicle routing.

We constructed a 4.38-approximation algorithm for capacitated location routing with arbitrary client demands, the first constant-factor approximation known for this problem. For the case of multi-depot capacitated vehicle routing, our algorithm improves the best known approximation ratio from 5 to 4. We then extended our algorithms to practically relevant generalizations of these problems, namely a prize-collecting version with penalties for non-served clients, and a group version, where one client from each group needs to be chosen. In all three cases, a variant where cross-docking is allowed leads to better approximation factors. All algorithms in our

name	#open fac.	fac. cost	#tours	approx		approx+tsp	
				cost	cpu	cost	cpu
$M_{1,1}$	8	31.8	10	3343.6	1.61	2478.9	1.65
$M_{1,2}$	13	79.4	61	4111.6	0.95	3499.1	1.00
$M_{1,3}$	33	779.4	760	13563.4	0.47	13478.9	0.55
$M_{2,1}$	1	102.6	6	3520.2	1.62	2620.8	1.67
$M_{2,2}$	5	528.2	57	5098.9	0.92	4468.7	0.98
$M_{2,3}$	18	2157.9	760	18086.2	0.52	17997.0	0.62
$M_{3,1}$	1	211.0	6	3656.8	0.83	2779.3	0.88
$M_{3,2}$	3	665.8	56	6012.0	1.25	5345.9	1.32
$M_{3,3}$	10	2370.3	757	23008.8	0.67	22926.8	0.76
$L_{1,1}$	16	42.1	36	7344.7	120.44	5463.7	121.95
$L_{1,2}$	47	337.3	289	9433.5	50.83	8106.1	59.29
$L_{1,3}$	128	2426.4	3709	32473.0	23.33	32325.9	35.90
$L_{2,1}$	2	206.0	30	8477.0	163.30	6624.8	165.08
$L_{2,2}$	10	1063.7	273	13435.5	65.84	12059.5	73.81
$L_{2,3}$	50	5900.5	3700	50380.6	28.81	50229.7	41.35
$L_{3,1}$	1	209.2	29	8835.3	210.84	6966.5	213.44
$L_{3,2}$	6	1273.0	271	15694.9	89.85	14372.4	97.81
$L_{3,3}$	31	7093.3	3696	64058.9	38.12	63905.1	50.94
$XL_{1,1}$	33	52.7	76	10400.0	742.56	7754.7	749.91
$XL_{1,2}$	78	405.1	583	13752.3	314.78	11872.0	369.47
$XL_{1,3}$	229	3394.8	7510	48879.5	136.27	48677.1	214.23
$XL_{2,1}$	4	407.6	57	12018.2	881.25	9296.1	886.83
$XL_{2,2}$	17	1752.7	555	20133.7	382.92	18159.1	434.58
$XL_{2,3}$	82	9264.9	7490	77796.3	164.96	77580.6	243.40
$XL_{3,1}$	2	409.7	57	13091.1	1317.87	10389.9	1322.69
$XL_{3,2}$	11	2255.2	552	23304.5	518.57	21341.5	570.46
$XL_{3,3}$	48	10593.1	7483	101676.0	227.77	101454.0	307.12
∅	32.85	2000.51	1439.59	22651.35	203.23	21562.00	221.03

Table 3. Solution properties, costs, and cpu times for random instances.

framework are based on computing an uncapacitated solution via a minimum spanning tree or Steiner tree, and rerouting excess demand according to the solution of a scaled facility location problem (or along shortest paths for multi-depot capacitated vehicle routing).

Finally, we demonstrated in a computational study that our algorithm for CLR is also of practical relevance. Our computational experiments revealed that the actual solution quality achieved by our algorithm is much closer to optimality than suggested by the theoretical bounds. On a benchmark set of instances from the literature, our algorithm for CLR computes solutions with cost within a factor of 1.1–1.2 of best known solutions on average. Moreover, we demonstrated that our algorithm is extremely fast, running in negligible time on benchmark instances. Thus, it might be a valuable tool for solving large-scale problems.

A further investigation of the algorithms in this paper would be of practical as well as theoretical interest: Given its fast running time, using its solution as a starting point in local search frameworks might lead to improved results of those heuristics without a significant increase in running time. Further experiments could be conducted on extended location routing models, e.g., with capacities on facilities, or heterogeneous vehicle fleets. Although the theoretical approximation guarantee might be lost in these cases, the algorithm could be adapted to still be an efficient heuristic for those problems.

On the theoretical side, it might be possible to sharpen our analysis and prove stronger theoretical approximation guarantees. Moreover, our paper does not address the issue of lower bounds on the possible approximation factor for basic capacitated location routing. It is easy to see that it cannot be approximated better than by a factor of 1.5 (unless $P = NP$), which is the best known lower bound for approximating a single-depot vehicle routing problem with uniform vehicle capacities [22]. However, an analysis that takes into account both the location and routing aspects of the problem might lead to stronger inapproximability results.

Moreover, our algorithms strongly rely on the technique of tree-to-tour-conversion, thereby incurring an additional factor of 2 in their approximation ratios. It would be interesting to find out if a more tour-specific approach, e.g., the tour partitioning techniques widely used for vehicle routing problems [33], could lead to better approximation factors.

The cross-docking model considered in this work assumes that the cost of actual cross-docking operations is negligible, and the operations can be performed at arbitrary client nodes. Deriving approximation algorithms for the case where cross-docking operations incur cost and are restricted to certain cross-docking facilities is an open problem. In Section 5, we point out that the possible improvement due to cross-docking is bounded by a factor of at most 4.38. We suspect the actual bound to be much smaller and leave its determination as a further open question for future research.

Acknowledgements: We would like to thank two anonymous referees, the guest editor, and the editor in chief for motivating us to conduct the computational study in Section 6.

References

[1] S. Arora, P. Raghavan, and S. Rao, *Approximation schemes for Euclidean k-medians and related problems*, in STOC '98, New York, NY, 1998, Association for Computing Machinery, pp. 106–113.

[2] R. Baldacci, A. Mingozzi, and R. Wolfer-Calvo, *The capacitated location routing problem.* Presented at ROUTE 2009, Roligted, Denmark, 2009.

[3] R. Baldacci, P. Toth, and D. Vigo, *Exact algorithms for routing problems under vehicle capacity constraints*, Annals of Operations Research **175**, no. 1 (2010), pp. 213–245.

[4] S. Barreto, C. Ferreira, J. Paixão, and B. S. Santos, *Using clustering analysis in a capacitated location-routing problem*, European Journal of Operational Research **179**, no. 3 (2007), pp. 968 – 977.

[5] J.-M. Belenguer, E. Benavent, C. Prins, C. Prodhon, and R. W. Calvo, *A branch-and-cut method for the capacitated location-routing problem*, Computers & OR **38**, no. 6 (2011), pp. 931–941.

[6] O. Bräysy and M. Gendreau, *Vehicle routing problem with time windows, part I: Route construction and local search algorithms*, Transportation Science **39**, no. 1 (2005), pp. 104–118.

[7] J. Byrka and K. Aardal, *An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem*, SIAM Journal on Computing **39**, no. 6 (2010), pp. 2212–2231.

[8] S. Cardon, E. Dommers, C. Eksin, R. Sitters, and L. Stougie, *A PTAS for the multiple depot vehicle routing problem*, SPOR Report 2008-03, Technische Universiteit Eindhoven, Eindhoven, 2008.

[9] M. Charikar, S. Khuller, D. M. Mount, and G. Narasimhan, *Algorithms for facility location problems with outliers*, in SODA '01, Philadelphia, PA, 2001, Society for Industrial and Applied Mathematics, pp. 642–651.

[10] M. Charikar, S. Khuller, and B. Raghavachari, *Algorithms for capacitated vehicle routing*, SIAM Journal on Computing **31**, no. 3 (2001), pp. 665–682.

[11] X. Chen and B. Chen, *Cost-effective designs of fault-tolerant access networks in communication systems*, Networks **53**, no. 4 (2009), pp. 382–391.

[12] N. Christofides, *Worst-case analysis of a new heuristic for the travelling salesman problem*, Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA, 1976.

[13] F. A. Chudak and D. B. Shmoys, *Improved approximation algorithms for the uncapacitated facility location problem*, SIAM Journal on Computing **33**, no. 1 (2003), pp. 1–25.

[14] C. Contardo, J.-F. Cordeau, and B. Gendron, *A branch-and-cut algorithm for the capactiated location problem*. unpublished manuscript, 2010.

[15] J. F. Cordeau, M. Gendreau, G. Laporte, J. Y. Potvin, and F. Semet, *A guide to vehicle routing heuristics*, Journal of the Operational Research Society **53** (2002), pp. 512–522.

[16] J. F. Cordeau, G. Laporte, M. W. P. Savelsbergh, and D. Vigo, *Vehicle routing*, in Transportation, C. Barnhart and G. Laporte, eds., Handbooks in Operations Research and Management Science 14, Elsevier, Amsterdam, 2007, pp. 367–428.

[17] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, The MIT Press, New York, 2001.

[18] M. Desrochers, J. K. Lenstra, M. W. P. Savelsbergh, and F. Soumis, *Vehicle routing with time windows: Optimization and approximation*, in Vehicle Routing: Methods and Studies, B. L. Golden and A. A. Assad, eds., North-Holland, Amsterdam, 1988, pp. 65–84.

[19] U. Feige, *A threshold of ln n for approximating set cover*, Journal of the ACM (JACM) **45**, no. 4 (1998), pp. 634–652.

[20] H. Glicksman and M. Penn, *Approximation algorithms for group prize-collecting and location-routing problems*, Discrete Applied Mathematics **156**, no. 17 (2008), pp. 3238–3247.

[21] M. X. Goemans and D. P. Williamson, *A general approximation technique for constrained forest problems*, SIAM Journal on Computing **24**, no. 2 (1995), pp. 296–317.

[22] B. Golden and R. Wong, *Capacitated arc routing problems*, Networks **11**, no. 3 (1981), pp. 305–315.

[23] *Vehicle Routing: Methods and Studies*, North-Holland, Amsterdam, 1988.

[24] M. Haimovich and A. H. G. Rinnoy Kan, *Bounds and heuristics for capacitated routing problems*, Mathematics of Operations Research **10**, no. 4 (1985), pp. 527–542.

[25] M. Haimovich, A. H. G. Rinnoy Kan, and L. Stougie, *Analysis of heuristics for vehicle routing problems*, in Vehicle Routing: Methods and Studies, B. L. Golden and A. A. Assad, eds., North-Holland, Amsterdam, 1988, pp. 47–61.

[26] K. Helsgaun, *An effective implementation of the Lin-Kernighan traveling salesman heuristic*, European Journal of Operations Research **126**, no. 1 (2000), pp. 106–130.

[27] D. S. Hochbaum, *Approximation algorithms for NP-hard problems*, PWS Publishing, Boston, MA, 1997.

[28] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani, *Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP*, Journal of the ACM **50**, no. 6 (2003), pp. 795–824.

[29] G. Laporte, *Location-routing problems*, in Vehicle Routing: Methods and Studies, B. L. Golden and A. A. Assad, eds., North-Holland, Amsterdam, 1988, pp. 163–198.

[30] G. Laporte, *Fifty years of vehicle routing*, Transportation Science **43**, no. 4 (2009), pp. 408–416.

[31] G. Laporte, Y. Nobert, and S. Taillefer, *Solving a family of multi-depot vehicle routing and location-routing problems*, Transportation Science **22**, no. 3 (1988), pp. 161–172.

[32] G. Laporte and F. Semet, *Classical heuristics for the capacitated VRP*, in The vehicle routing problem, P. Toth and D. Vigo, eds., Society for Industrial and Applied Mathematics, Philadelphia, PA, 2001, pp. 109–128.

[33] C. Li and D. Simchi-Levi, *Worst-case analysis of heuristics for multidepot capacitated vehicle routing problems*, ORSA Journal on Computing **2**, no. 1 (1990), pp. 64–73.

[34] J. Liu, C.-L. Li, and C.-Y. Chan, *Mixed truck delivery systems with both hub-and-spoke and direct shipment*, Transportation Research Part E: Logistics and Transportation Review **39**, no. 4 (2003), pp. 325–339.

[35] M. Mahdian, Y. Ye, and J. Zhang, *Approximation algorithms for metric facility location problems*, SIAM Journal on Computing **36**, no. 2 (2006), pp. 411–432.

[36] H. Mina, V. Jayaraman, and R. Srivastava, *Combined location-routing problems: A synthesis and future research directions*, European Journal of Operational Research **108**, no. 1 (1998), pp. 1–15.

[37] G. Nagy and S. Salhi, *Location-routing: Issues, models and methods*, European Journal of Operational Research **177**, no. 2 (2007), pp. 649–672.

[38] C. PRINS, C. PRODHON, AND R. W. CALVO, *Solving the capacitated location-routing problem by a grasp complemented by a learning process and a path relinking*, 4OR **4**, no. 3 (2006), pp. 221–238.

[39] C. PRINS, C. PRODHON, A. B. RUIZ, P. SORIANO, AND R. W. CALVO, *Solving the capacitated location-routing problem by a cooperative lagrangean relaxation-granular tabu search heuristic*, Transportation Science **41**, no. 4 (2007), pp. 470–483.

[40] R. RAVI AND A. SINHA, *Approximation algorithms for problems combining facility location and network design*, Operations Research **54**, no. 1 (2006), pp. 73–81.

[41] P. TOTH AND D. VIGO, *The Vehicle Routing Problem*, Monographs on Discrete Mathematics and Applications, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2002.

[42] D. TUZUN AND L. I. BURKE, *A two-phase tabu search approach to the location routing problem*, European Journal of Operational Research **116**, no. 1 (1999), pp. 87–99.

[43] B. VAHDANI AND M. ZANDIEH, *Scheduling trucks in cross-docking systems: Robust meta-heuristics*, Computers & Industrial Engineering **58**, no. 1 (2010), pp. 12–24.

[44] M. H. J. WEBB, *Cost functions in the location of depots for multiple-delivery journeys*, Operations Research Quarterly **19** (1968), pp. 311–320.

[45] M. WEN, J. LARSEN, J. CLAUSEN, J.-F. CORDEAU, AND G. LAPORTE, *Vehicle routing with cross-docking*, Journal of the Operational Research Society **60**, no. 11 (2009), pp. 1708–1718.

[46] D. WILLIAMSON AND D. SHMOYS, *The Design of Approximation Algorithms*, Cambridge University Press, 2011.