**Mathematical Programming manuscript No.**
(will be inserted by the editor)

George Karakostas · Anastasios Viglas

# Equilibria for networks with malicious users[*]

**Abstract.** We consider the problem of characterizing user equilibria and optimal solutions for selfish routing in a given network. We extend the known models by considering malicious behaviour. While selfish users follow a strategy that minimizes their individual cost, a *malicious* user will use his flow through the network in an effort to cause the maximum possible damage to this cost. We define a generalized model, present characterizations of flows at equilibrium and prove bounds for the ratio of the social cost of a flow at equilibrium over the cost when centralized coordination among users is allowed.

## 1. Introduction

The general framework of a system of non-cooperative users can be used to model many different optimization problems such as network routing, traffic or transportation problems, load balancing and distributed computing, auctions

George Karakostas: McMaster University, Dept. of Computing and Software, 1280 Main St. West, Hamilton, Ontario L8S 4K1, Canada, e-mail: gk@cas.mcmaster.ca. This research was supported in part by an NSERC Discovery research grant and MITACS.

Anastasios Viglas: University of Sydney, School of Information Technologies, Madsen Building F09, The University of Sydney, NSW 2006, Australia, e-mail: tasos@it.usyd.edu.au. This research was done when this author was a postdoctoral fellow at the University of Toronto, Canada.

[*] An extended abstract of this work appeared in the Proceedings of the 14th Annual International Symposium on Algorithms and Computation (ISAAC) 2003 [KV03]

and many more. Game Theoretic techniques can be used to model and analyze such systems in a natural way. The performance of a system of non-cooperative users is measured by an appropriate cost function which depends on the behaviour, or strategies of the users. For example in the case of network routing, the total, system-wide cost can be defined as the total routing cost, or the total latency experienced by all the users in the network. On the other hand, there is also a cost associated with each individual user (for example the latency experienced by the user). It is a well known fact that if each user optimizes her own cost, then they might choose a strategy that does not give the optimal total cost for the entire system, also known as *social cost* [P20]. In other words, the *selfish* behaviour of the users leads to a sub-optimal performance.

Koutsoupias and Papadimitriou [KP99] initiated the study of the *coordination ratio* (also referred to as the price of anarchy): How much worse is the performance of a network of selfish users where each user optimizes her own cost, compared to the best possible performance that can be achieved on the same system? This question has been studied in various different models (e.g. [RT02], [SM03]) and bounds for the coordination ratio have been shown for many interesting cases. The discrepancy in the network performance between selfish (uncoordinated) and optimized (coordinated) behavior is to be expected: typically, the selfish users use a local heuristic (e.g., the usage of the fastest path to send flow), that doesn't take into account the overall picture of the network, while the central coordinator is not restricted by the usage of a particular heuristic and has the whole network picture at her disposal. Therefore, the rather surprising fact is the existence of small coordination ratio bounds for non-trivial cases.

A basic assumption of the models considered so far is that the users are considered to be selfish and *non-malicious*: the user optimizes her own utility or payoff, and does not care about the performance of the system or the cost induced to other users by her strategy. In this work, we are motivated by the following setting: Suppose that a content provider supplies live feed to a number of clients through a network, using the fastest possible connection to each client at any time. The speed of the connection depends on the congestion of the network links (as congestion increases, the link delay also increases). The clients will remain connected to the provider's server as long as the feed delay is acceptable, otherwise they will terminate the connection. At the same time, a malicious user pretends to be a client stationed at any other point in the network, so it can request feed from the provider. The flow that this malicious user can request is limited (he cannot afford more than a certain (limited) number of simultaneous connections), but what he can do is to constantly look for switching to a different network node that will make the connection slower. If he knows that the provider tries to be fair to all its customers (i.e., they should all experience more or less the same delay), then his own delay is an indication of the delay the other customers are experiencing. Therefore, if at some point he manages to increase his delay to unacceptable levels, he knows that the other customers will deny the services of the provider. The natural question that arises then is, how much degradation of the network performance can be expected from such malicious behavior?

We try to formulate and (partially) answer such questions by extending the existing selfish routing models to include *malicious users*. As in the example above, a malicious user will choose a strategy that will cause the worst possible performance for the entire network. Such malicious behaviour can be found in practice in settings such as the internet (for example in 'denial of service' attacks). While in terms of Wardrop equilibria, the extension of the selfish model considered before is quite straight-forward, the existence of malicious users forces us to a different model for the 'social optimum'. We no longer have an objective function that can be minimized by the centralized coordination among the users, since in our setting some of the users still can be coordinated to *minimize* it, but at the same time there is a (malicious) user that tries to *maximize* it. This leads naturally to the formulation of the 'social optimum' objective as a *minimax* problem instead of just a minimization problem. As a result, we cannot refer to an 'social optimum' that is a global minimizer of the social cost objective. Instead, we have to compare the worst Wardrop equilibrium to the *saddle-points* of the minimax problem. We define the 'social optimum' as the minimum cost achieved by the set of saddle-points. The fact that this set is (usually) *non-convex* makes the exact characterization of the 'social optimum' (and therefore the coordination ratio) more difficult to characterize than the previous models. Nevertheless, in this paper we show that in the very general setting considered by Roughgarden and Tardos [RT02], their results can be extended to the case of systems with malicious users.

**Previous Work:** Many of the Game Theoretic tools used for analyzing systems of non-cooperative users derive from results in traffic models and transportation, including work of Dafermos and Sparrow [DS69], Beckmann, McGuire and Winsten [BMW56] and Aashtiani and Magnanti [AM81]. More recently, Nash equilibria and their applications were used for routing problems and the internet. Koutsoupias and Papadimitriou [KP99] considered the coordination ratio for load balancing problems (routing on a network of parallel links). The model they considered allowed multiple equilibria, and the coordination ratio compared the worst case equilibrium cost to the optimal routing cost. Their bounds were improved in subsequent work on the same model by Mavronicolas and Spirakis [MS01], and Czumaj and Vöcking [CV02]. Roughgarden and Tardos [RT02] considered a different model for selfish routing, where there is a unique Wardrop equilibrium and proved bounds for the coordination ratio, including results for the special case of linear utility functions. Other work in this model includes results on the topology of the underlying network [Rou01a, Rou02], algorithms and bounds for Stackelberg scheduling strategies [Rou01b], etc.

In Sections 2-4 we present the general setting of our study: first, we present the selfish routing model with malicious users and the notion of Wardrop equilibria for it (Section 2), and prove the existence of Wardrop equilibrium flows under certain constraints for the latency functions; then, in Section 3 we define the 'social optimum' as the solution (saddle-points) of a minimax program and give characterizations of these saddle-points; finally in Section 4 we connect the previous two sections with the use of the coordination ratio measure which compares the Wardrop equilibria of Section 2 to the saddle-points of Section 3. After this general setting, we study a set of particular disutility functions for the good and malicious users. These functions are defined in Section 5. For this case, the Wardrop equilibria are compared with the saddle-points of the minimax program in Section 4, first to get a bicriteria result (Section 6), and then to get upper bounds first for the instructive case of linear latency functions (Section 7.1) and then for more general latency functions that satisfy certain assumptions (Section 7.2). We conclude with a short discussion and open problems.

## 2. The selfish routing model

We are given a directed network $G = (V, E)$ and $k$ source-sink pairs of nodes $(s_i, t_i), i = 1 \dots k$. There are also two special nodes $s_M, t_M$ connected to $G$ with edges $(s_M, s_i), (t_i, t_M), i = 1 \dots k$. A commodity $i$ with demand $r_i$ is associated with each pair $(s_i, t_i), i = 1 \dots k$, and a commodity $M$ of demand $F$ is associated with pair $(s_M, t_M)$. Let $\mathcal{P}_i$ ($\mathcal{P}_M$) be the set of acyclic paths from $s_i$ to $t_i$ ($s_M$ to $t_M$). A latency function $l_P(\cdot)$ is associated with each path $P$. For a flow $f$ on $G$, $l_P(f)$ is the latency (cost) of path $P$ for this particular flow. Notice that in general this latency depends on the whole flow $f$, and not only on the flow $f_e$ through each edge $e \in P$. In this paper we adopt the *additive model* for the path latencies, i.e. $l_P(f) = \sum_{e \in P} l_e(f_e)$, where $l_e$ is the latency function for edge $e$ and $f_e$ is the amount of flow that goes through $e$. In fact, the flow $f$ is the combination of a *good* flow $f_e^G$ and a *malicious* flow $f_e^M$, so $l_e$ is in fact a function

$l_e(f_e^G, f_e^M) : \mathbb{R}_+^2 \to \mathbb{R}_+$. We also let $\mathcal{P}$ be the set of all available paths in the network and assume that for every source-sink pair there is at least one path joining the source to the sink. The *cost* for a flow $f = (f^G, f^M)$ through edge $e$ is given by a function $c_e : \mathbb{R}_+^2 \to \mathbb{R}_+$, defined as

$$c_e(F_e^G, f_e^M) := f_e^G \cdot l_e(f_e^G, f_e^M).$$

Note that this cost function gives the cost inflicted on the *good* flow. Let

$$C(f^M, f^G) := \sum_{e \in E} c_e(f_e^M, f_e^G))$$

be the *social cost* of $f$. We will use the shorthand $(G, r, F, l)$ to describe an instance of the model.

Commodities $i = 1 \ldots k$ model selfish, but otherwise 'good' users who want to just use the network in order to satisfy their demands with the smallest possible cost (i.e. latency for every unit of flow routed). Commodity $M$ models a selfish 'malicious' user who wants to use his own flow $F$ in such a way that will do the biggest possible damage to the total cost of the good players. Intuitively, the malicious user that we define later more precisely tries to route his/her flow so that the social cost $C(\cdot)$ is as big as possible. Of course this is just a specific malicious behavior we try to model here. This doesn't preclude the modelling of other kinds of malicious users in some future work.

For the definition of the classic *traffic* or *Wordrop* equilibrium model [1], we use the following general formulation by Aashtiani and Magnanti [AM81]:

**Definition 1** *A flow* $f = \cup_{P \in \mathcal{P}} f_P$ *is at* Wardrop equilibrium *for instance* $(G, r, F, l)$ *iff it satisfies the following constraints:*

$$\begin{aligned}
(T_P(f) - u_i) f_P &= 0 \quad \textit{for all } P \in \mathcal{P}_i, i = 1 \ldots k && (1)\\
(T_P(f) - u_M) f_P &= 0 \quad \textit{for all } P \in \mathcal{P}_M\\
T_P(f) - u_i &\geq 0 \quad \textit{for all } P \in \mathcal{P}_i, i = 1 \ldots k\\
T_P(f) - u_M &\geq 0 \quad \textit{for all } P \in \mathcal{P}_M\\
\sum_{P \in \mathcal{P}_i} f_P - r_i &= 0 \quad \textit{for all } i = 1 \ldots k\\
\sum_{P \in \mathcal{P}_M} f_P - F &= 0\\
f &\geq 0\\
u &\geq 0
\end{aligned}$$

[1] We use this classic (amongst the OR community) definition of Wardrop (or traffic) equilibria since we find it less cumbersome for our exposition (usually this definition is also formulated as a variational inequality- we use the equivalent formulation of [AM81] again because it facilitates our exposition.) Another definition of equilibria (as used for example in [RT02], [Rou02]) is the notion of *Nash equilibria*, where each commodity is comprised by individual users, each of them carrying an infinitesimal amount of the flow. Wardrop equilibria can be seen as the limit of Nash equilibria when this infinitesimal amount of flow tends to 0, as is shown in [HM85]. To avoid confusion, note that when we refer to 'users', we refer to commodities specified by their Origin-Destination pairs, their (fixed) demands, and the set of network paths they are allowed to use. We make no reference to Nash equilibria or infinitesimal users as defined in this kind of equilibria.

*where $T_P$ is the delay time or general disutility for path $P$, $f_P$ is the flow through path $P$, and $u = (u_1, \ldots, u_k, u_M)$ is the vector of shortest travel times (or generalized costs) for the commodities.*

Note that $T_P$ does not need to be the same function for all paths $P$ (and, indeed, it will be a different function for the good and the malicious users). Also we emphasize that $T_P$ need *not* be the path latency (the latter is given by function $l_P$). In what follows we define precisely the functions $T_P$ for all users, and thus we define completely the equilibrium model of Definition 1.

The first four equations are the conditions for the existence of a Wardrop traffic equilibrium. They require that the general disutility for all paths $P$ that carry flow $f_P > 0$ is the same and equal to $u$ for every user, and less or equal to the disutility of any path with zero flow. Any flow that complies with this definition of a Wardrop equilibrium, also satisfies the following alternative characterization:

**Lemma 1.** *A flow that is feasible for instance $(G, r, F, l)$ is a Wardrop equilibrium iff for every commodity $i$ ($i$ can be the malicious commodity $M$) and every pair of paths $P_1, P_2 \in \mathcal{P}_i$ with $f_{P_1} > 0$, $T_{P_1}(f) \leq T_{P_2}(f)$.*

### *2.1. Existence of Wardrop equilibrium*

The model of Definition 1 is very general. It turns out that the existence of a Wardrop equilibrium in this model can also be proved under very general assumptions. More specifically, the following theorem follows immediately from Theorem 5.4 in [AM81]:

**Theorem 1.** *Suppose that $T_P$ is a positive continuous function for all $P \in \mathcal{P}$. Then there is a flow that satisfies the conditions of Definition 1.*

A function is positive if its values are positive. In order to make sure that the disutility functions we use later are positive, and therefore a Wardrop equilibrium always exists, from now on we make the following natural assumption:

**Assumption 1** *We assume that the latency function for every edge is an increasing function of the total flow, i.e. as the good or malicious (or both) flows increase for an edge, its latency also increases.*

Assumption 1 will be used throughout this paper in order to make sure that Theorem 1 holds, and a Wordrop equilibrium exists.

## 3. Social optimum when malicious users are present

The existence of a malicious user forces us to redefine the notion of 'social optimum' [P20]. In addition to a set of users that collectively strive to *minimize* their collective cost, we have a user who strives to *maximize* this same cost. Therefore we define the 'socially best' flow in terms of a *minimax* problem. Note that in such a setting the notion of an "optimal flow" is replaced by the notion of a flow "in equilibrium". Therefore our work compares a Wardrop equilibrium to a minimax equilibrium (as opposed to the comparison of a Wardrop equilibrium to an optimal solution of a minimization problem, as in [RT02]).

In what follows, we denote the flow of the good users by $f^G$, and the flow of the malicious user by $f^M$ (recall that we denote by $f$ the total flow). We consider the following minimax formulation:

$$\max_{f^M} \min_{f^G} \sum_{e \in E} c_e(f_e^M, f_e^G) \quad \text{subject to:} \tag{MINMAX}$$

$$\begin{aligned}
\sum_{P \in \mathcal{P}_i} f_P^G &= r_i && \forall i \in \{1, \ldots, k\} \\
\sum_{P \in \mathcal{P}_M} f_P^M &= F && \\
f_e^G &= \sum_{P \in \mathcal{P} : e \in P} f_P^G && \forall e \in E \\
f_e^M &= \sum_{P \in \mathcal{P} : e \in P} f_P^M && \forall e \in E \\
f_P^G &\geq 0 && \forall P \in \mathcal{P} \\
f_P^M &\geq 0 && \forall P \in \mathcal{P}
\end{aligned}$$

where $c_e(f_e^M, f_e^G)$ is the cost of flow $(f_e^M, f_e^G)$ passing through edge $e$, i.e., the objective function is $C(f^M, f^G)$. We call this minimax formulation (MINMAX).

The solution(s) to (MINMAX) are called *saddle-points*. The saddle-points are defined as follows:

**Definition 2** *A flow* $(\bar{f}^G, \bar{f}^M)$ *is said to be a* saddle-point *of* $C$ *(with respect to maximizing in* $f^M$ *and minimizing in* $f^G$*) if*

$$C(\bar{f}^G, f^M) \leq C(\bar{f}^G, \bar{f}^M) \leq C(f^G, \bar{f}^M), \quad \forall f^M, \ \forall f^G. \tag{2}$$

We also refer to (MINMAX) saddle-points as *(MINMAX) equilibria.*

### *3.1. Characterization of saddle-points*

A saddle-point is not always guaranteed to exist. But under certain assumptions, we can show that (at least one) saddle-point exists. We assume the following for the cost function $C(f^M, f^G)$:

**Assumption 2** *The functions* $c_e(f_e^M, f_e^G)$ *are* continuous*,* differentiable*,* convex *with respect to* $f^G$*, and* concave *with respect to* $f^M$ *for all* $e \in E$*.*

An example of such functions is the linear latency functions case $l_e(f_e^G, f_e^M) := a_e(f_e^G + f_e^M) + b_e$ studied later. Following the methods of Dafermos and Sparrow [DS69], and under Assumption 2, we can prove the following theorem for the properties of saddle-points for (MINMAX).

**Theorem 2.** *Under Assumption 2, a feasible flow* $\bar{f} = (\bar{f}^M, \bar{f}^G)$ *is a solution (saddle-point) to the minimax problem (MINMAX) if and only if it has the following properties:*

$$\sum_{e\in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}) \le \sum_{e\in P'} \frac{\partial c_e}{\partial f_e^G}(\bar{f}), \quad \forall i = 1 \dots k,\ \forall P, P' \in \mathcal{P}_i \text{ with } \bar{f}_P^G > 0 \tag{3}$$

$$\sum_{e\in P} \frac{\partial c_e}{\partial f_e^M}(\bar{f}) \ge \sum_{e\in P'} \frac{\partial c_e}{\partial f_e^M}(\bar{f}), \quad \forall P, P' \in \mathcal{P}_M \text{ with } \bar{f}_P^M > 0 \tag{4}$$

*In particular, the above imply that for every 'good' user* $i = 1, \dots, k$*, and the malicious user we have:*

$$\sum_{e\in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}) = \sum_{e\in P'} \frac{\partial c_e}{\partial f_e^G}(\bar{f}) = A_i, \quad \forall P, P' \in \mathcal{P}_i \text{ with both } \bar{f}_P^G, \bar{f}_{P'}^G > 0 \tag{5}$$

$$\sum_{e\in P} \frac{\partial c_e}{\partial f_e^M}(\bar{f}) = \sum_{e\in P'} \frac{\partial c_e}{\partial f_e^M}(\bar{f}) = B, \quad \forall P, P' \in \mathcal{P}_M \text{ with both } \bar{f}_P^M, \bar{f}_{P'}^M > 0 \tag{6}$$

*Proof.* The proof is a straight-forward extension to the proof of Theorem 1.2 in [DS69]. First we prove the sufficiency of conditions (3), (4), i.e. we prove that if conditions (3), (4) are satisfied by a feasible flow $\bar{f} = (\bar{f}^G, \bar{f}^M)$, then $\bar{f}$ is a saddle point for (MINMAX). In order to show this, we have to show two things:

1. For every feasible flow $\bar{f} + \Delta\bar{f}^G = (\bar{f}^G + \Delta\bar{f}^G, \bar{f}^M)$, $C(\bar{f}) \leq C(\bar{f} + \Delta\bar{f}^G)$ (i.e. if we perturb the flow of the 'good' users by $\Delta\bar{f}^G$ by reallocation, so that the new flow is still *feasible*, the total cost cannot decrease).
2. For every feasible flow $\bar{f} + \Delta\bar{f}^M = (\bar{f}^G, \bar{f}^M + \Delta\bar{f}^M)$, $C(\bar{f}) \geq C(\bar{f} + \Delta\bar{f}^M)$ (i.e. if we perturb the flow of the malicious user by $\Delta\bar{f}^M$ by reallocation, so that the new flow is still *feasible*, the total cost cannot increase).

Here we show (1). Showing (2) is completely analogous.

The change of the cost because of the reallocation of the 'good' flow is

$$\Delta C = \sum_{e \in E} \left[ c_e(\bar{f}_e^G + \Delta\bar{f}_e^G, \bar{f}_e^M) - c_e(\bar{f}_e^G, \bar{f}_e^M) \right]$$

Assumption 2 implies that the functions $\frac{\partial c_e}{\partial f_e^G}(f_e^G, f_e^M)$ are non-decreasing functions of $f_e^G$ (because of the convexity of $c_e(f_e^G, f_e^M)$ with regard to $f_e^G$).

Therefore we can apply the Mean Value Theorem with respect to $f_e^G$ to get

$$\begin{aligned}
\Delta C &\geq \sum_{e\in E} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) \cdot \Delta \bar{f}_e^G \\
&= \sum_{i=1}^{k} \sum_{P\in\mathcal{P}_i} \sum_{e\in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) \cdot \Delta \bar{f}_e^G \\
&= \sum_{i=1}^{k} \sum_{P\in\mathcal{P}_i} \Delta \bar{f}_P^G \cdot \sum_{e\in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) \\
&\geq \sum_{i=1}^{k} \sum_{P\in\mathcal{P}_i} \Delta \bar{f}_P^G \cdot A_i \\
&= \sum_{i=1}^{k} A_i \sum_{P\in\mathcal{P}_i} \Delta \bar{f}_P^G = 0
\end{aligned}$$

where the second inequality is due to (3) and (5) together with the fact that if $\bar{f}_P^G = 0$ then $\Delta \bar{f}_P^G \geq 0$[2], and the last equality is due to the fact that the flow for user $i$ was reallocated, but its total value didn't change (it remained $r_i$), since it remained feasible.

By repeating the same argument for the case of reallocation of flow for the malicious user (and by using the concavity of $c_e$ with respect to $f_e^M$), we can show that the total cost cannot decrease. Therefore conditions (3) and (4) are sufficient.

In order to prove the necessity of (3) and (4), assume that $(\bar{f}^G, \bar{f}^M)$ is a saddle point for (MINMAX) and condition (3) doesn't hold, i.e., there are paths $P, Q$ with $\bar{f}_P^G > 0$ and such that

$$\sum_{e\in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) - \sum_{e\in Q} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) = \epsilon > 0 \qquad (7)$$

[2] Hence $\Delta \bar{f}_P^G \cdot \sum_{e\in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) \geq \Delta \bar{f}_P^G \cdot A_i$ for all $P \in \mathcal{P}_i$.

If a portion $\Delta f$ of $\bar{f}_P^G$ is reallocated to path $Q$, the change of the total cost will be

$$\Delta C = \sum_{e \in P \setminus Q} [c_e(\bar{f}_e^G - \Delta f, \bar{f}_e^M) - c_e(\bar{f}_e^G, \bar{f}_e^M)] + \sum_{e \in Q \setminus P} [c_e(\bar{f}_e^G + \Delta f, \bar{f}_e^M) - c_e(\bar{f}_e^G, \bar{f}_e^M)] \tag{8}$$

From the Mean Value Theorem for the convex function (on $f_e^G$) $c_e(f_e^G, f_e^M)$ we have that

$$\begin{aligned} c_e(\bar{f}_e^G - \Delta f, \bar{f}_e^M) - c_e(\bar{f}_e^G, \bar{f}_e^M) &\leq -\frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G - \Delta f, \bar{f}_e^M) \cdot \Delta f \\ c_e(\bar{f}_e^G + \Delta f, \bar{f}_e^M) - c_e(\bar{f}_e^G, \bar{f}_e^M) &\leq \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G + \Delta f, \bar{f}_e^M) \cdot \Delta f \end{aligned}$$

and therefore (8) implies that

$$\Delta C \leq \left[ - \sum_{e \in P \setminus Q} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G - \Delta f, \bar{f}_e^M) + \sum_{e \in Q \setminus P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G + \Delta f, \bar{f}_e^M) \right] \Delta f. \tag{9}$$

Since functions $\sum_{e \in P \setminus Q} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G - \Delta f, \bar{f}_e^M)$ and $\sum_{e \in Q \setminus P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G + \Delta f, \bar{f}_e^M)$ are continuous on $\Delta f$, we can choose $0 < \Delta f \leq f_P^G$ such that

$$\begin{aligned} \sum_{e \in P \setminus Q} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G - \Delta f, \bar{f}_e^M) &> \sum_{e \in P \setminus Q} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) - \frac{\epsilon}{3} \\ \sum_{e \in Q \setminus P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G + \Delta f, \bar{f}_e^M) &< \sum_{e \in Q \setminus P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) - \frac{\epsilon}{3} \end{aligned}$$

Hence we have that

$$\begin{aligned} \Delta C &< \left[ - \sum_{e \in P \setminus Q} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) + \sum_{e \in Q \setminus P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) + \frac{2\epsilon}{3} \right] \Delta f \\ &= \left[ - \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) + \sum_{e \in Q} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) + \frac{2\epsilon}{3} \right] \Delta f \\ &\overset{(7)}{<} -\frac{\epsilon}{3} \Delta f < 0 \end{aligned}$$

which contradicts the fact that $(\bar{f}^G, \bar{f}^M)$ is a saddle point. Hence (3) must hold. In exactly the same way, we can show that (4) must hold as well. □

Conditions (3) and (4) are simply the Kuhn-Tucker conditions for problem (MINMAX) [Roc70]. We sketched the proof of Theorem 2 here because the same proof techniques apply in the proof of Theorem 8 later.

## 4. Wardrop vs. Minimax equilibria

The saddle-points of (MINMAX) give us the social cost achieved in a system with both good and malicious users, provided there is a central authority that can direct the flow of each good user so that the total cost is the minimum possible, in the presence of an all-knowing malicious coordinator who wants to maximize the social cost. This cost, which is a quantitative estimate of the 'social cost' that can be achieved by a central coordinator, may be quite different to the total cost achieved by the lack of such a coordinator, i.e. by allowing each user (good or malicious) to act *selfishly*[3]. Here, we define natural selfish behaviors for both the good and malicious users, in accordance with the general model of Definition 1. Our aim will be to estimate how far can selfishness push the total cost from the optimal coordinated one (i.e. the best saddle-point of (MINMAX)). In order to do this, we use the new notion of 'social optimum' in the definition of the *price of*

[3] As was pointed out by an anonymous reviewer, one can see the good and malicious coordinators as two selfish players as well, playing at a level above the coordinated users. But the users under coordination are no longer allowed to act selfishly.

*anarchy* or *coordination ratio*, defined by Koutsoupias and Papadimitriou [KP99] and used by Roughgarden and Tardos [RT02].

**Definition 3 (Coordination ratio)** *Let $(G, r, F, l)$ be an instance of the routing problem on network $G$ with latency function $l_e(\cdot)$ for every edge $e$, with $k$ good users with demands $r_i,\ i = 1, \ldots, k$ and a malicious user with flow $F$. Then the* coordination ratio $\rho(G, r, F, l)$ *for this instance is defined as follows:*

$$\rho(G, r, F, l) = \frac{\textit{worst Wardrop equilibrium}}{\textit{best saddle-point of (MINMAX)}}. \tag{10}$$

In case the Wardrop equilibrium or/and the (MINMAX) equilibrium is/are unique, then the 'worst' or/and 'best' in the definition above can obviously be omitted. For the class of latency functions we study here, the Wardrop equilibria values need not be the same, since the bounds for the coordination ratio proven later also hold if 'worst' is replaced by 'any' in Definition 3. For the saddle-point values, we either prove(Theorem 5) or it follows from our assumptions (cf. discussion at the beginning of Section 7.2) that they all have the same cost. But in other cases, we should emphasize that it may be very difficult (or even impossible) to characterize the 'best saddle-point of (MINMAX)' (e.g. when the set of these equilibria is *not* convex, as is usually the case).

**Why is this definition of coordination ratio natural?** We view the coordination ratio as a measure of comparison between the network performance when there is an all-knowing central coordinating authority for the good as well as the malicious users, and the network performance when all users are selfish. Note that in both sides of this comparison we don't get rid of the malicious behavior. We consider this to be natural, in view of the fact that the existence of malicious behavior is independent of the level of coordination. This becomes somewhat clearer when one starts thinking what can happen if, for example, malicious behavior is not allowed in the coordinated version (note that in this case instead of (MINMAX) we get the usual minimization definition of social cost.) If this is the case, how should the flow of the malicious user be treated? Should it become the flow of one more good user, assuming that the malicious user will still want to route his flow regardless? Or should it disappear from the network, since in the case a central authority exists a malicious user has no incentive to send any flow? Moreover, one should see (MINMAX) not as another game (although it can also be seen as such), but as an *optimization problem* for two central coordinators: a good one that tries to minimize his/her cost, and a malicious one, that tries to maximize the good coordinator's cost. Hence our "coordination ratio" as defined above tries to capture exactly the coordination ratio *without removing the malicious behavior*, and therefore we believe it is a quite natural definition. As is mentioned in the open problems section, currently we do not have a persuading definition for the 'price of evil', i.e., we don't know what would be the meaning of a comparison between a selfish setting with malicious users and a centrally coordinated network without malicious users. In our opinion, this is the greatest modelling challenge this work leaves open.

**Effective malicious users** So far we have not specified exactly what the strategy of a malicious user should be. We have only assumed that intuitively a malicious user wants to hurt the performance of the good users. But what seems intuitively to be malicious behavior, may turn out to be *beneficial.* A simple example is the *Braess paradox*, shown in Figure 1. Suppose that there is one good user (commodity) with demand equal to 6, and edge latency functions that depend on the flow $x$ that goes through an edge as shown in the figure. Then if the network (a) is used, at equilibrium the flow will be split evenly between paths $1 \to 2 \to 4$ and $1 \to 3 \to 4$, with a latency of 83 for both paths, and a total cost of 498. But if network (b) is used, then at equilibrium the flow will split in three equal parts, using paths $1 \to 2 \to 4, 1 \to 3 \to 4$ and $1 \to 2 \to 3 \to 4$, each with a latency of 92, for an overall cost of $552 > 498$. This well known phenomenon can be reversed, if we consider also a malicious user for network (b), who (naively) believes that edge $(2, 3)$ actually helps the traffic, for the same reasons that the well-intentioned network builders put the edge there in the first place. Hence the malicious user appropriates the use of vertices $2, 3$ and routes his flow from 2 to 3. Suppose that he has 1000 units of flow at his disposal (so he has overwhelming power in his hands!) Routing this flow will force the good user to revert to the behavior he exhibited in network (a), which in fact is an improvement! Hence it is obvious that not all models of 'malicious' behavior are really malicious. Here we are interested only in malicious users that really hurt the performance of a network, i.e., they are *effective*:

**Definition 4** *A malicious user for* $(G, r, F, l)$ *is* effective *iff* $\rho(G, r, F, l) \geq 1$.

As part of our results, we show that the malicious user behavior we consider in the next section is effective.

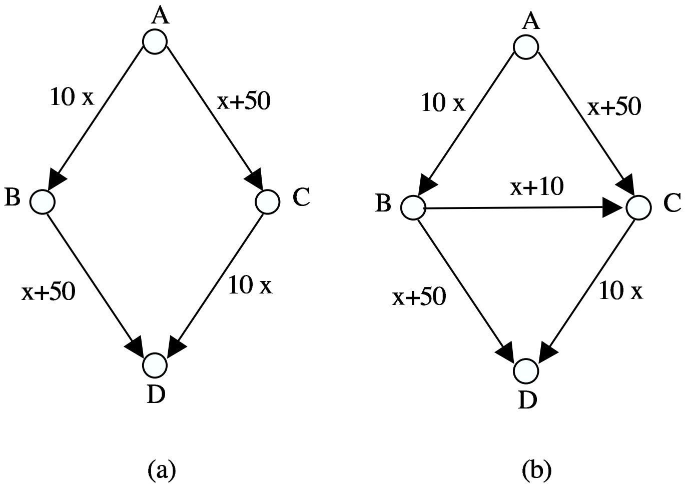

**Fig. 1.** The Braess paradox.

## 5. A specific model of good and malicious behavior

According to the model of Definition 1, the selfish users will base their decisions for picking flow paths on their individual notion of general disutility $T_P$, for every path $P$. This disutility is very easy to be defined for the 'good' users: it is simply the latency of the path, i.e.

$$T_P(f^G, f^M) := l_P(f^G, f^M) \ (= \sum_{e \in P} l_e(f_e)), \quad \forall i = 1, \ldots, k, \ \forall P \in \mathcal{P}_i \qquad (11)$$

For the malicious user though, the form of his general disutility in fact determines how powerful or weak this user can be. In this paper we study malicious players that base their decisions exclusively on the costs of *individual paths*. The malicious player exhibits a rather greedy behavior, and does not (or cannot[4]) take into account the impact of his decisions on the whole network (e.g. by solving (MINMAX) so that his allocation of flow will have the worst impact on the

[4] maybe because of lack of resources, e.g. time in an on-line scenario

'social cost' he might be able to achieve more damage than looking greedily at the costs of individual paths). Let $R := \sum_{i=1}^{K} r_i$. Define

$$M(r, F) := 1 + \sum_{e \in E} R \cdot \frac{\partial l_e}{\partial f_e^M} R, 0.$$

The definition of $M(\cdot)$ may seem cryptic, but we need it in order to be able to guarantee the existence of a Wardrop equilibrium (Theorem 1) when we define the disutility functions for the malicious user.

**Lemma 2.** $M(r, F) > \sum_{e \in E} f_e^G \cdot \frac{\partial l_e}{\partial f_e^M}(f_e^G, 0)$ *for any* $f^G$ *that satisfies the given (good) demands.*

*Proof.* Since the edge latency functions $l_e$ are concave with respect to $f_e^M$ (because the $c_e$'s are concave with respect to $f_e^M$ from Assumption 2), the function $\frac{\partial l_e}{\partial f_e^M}(f_e^G, f_e^M)$ is decreasing with respect to $f_e^M$, so its maximum value for a particular $f_e^G$ is $\frac{\partial l_e}{\partial f_e^M}(f_e^G, 0)$, which is also positive because $l_e$ increases when $f_e^M$ increases (Assumption 1.) Hence

$$\sum_{e \in E} f_e^G \cdot \frac{\partial l_e}{\partial f_e^M}(f_e^G, 0) \geq \sum_{e \in E} f_e^G \cdot \frac{\partial l_e}{\partial f_e^M}(f_e^G, f_e^M)$$

The same reasoning can be used to show that

$$\sum_{e \in E} f_e^G \cdot \frac{\partial l_e}{\partial f_e^M}(R, 0) \geq \sum_{e \in E} f_e^G \cdot \frac{\partial l_e}{\partial f_e^M}(f_e^G, 0)$$

and then the lemma follows easily. □

The general disutility for the malicious user paths is defined as follows:

$$T_P(f^G, f^M) := M(r, F) - \sum_{e \in P} f_e^G \cdot \frac{\partial l_e}{\partial f_e^M}(f_e^G, f_e^M), \quad \forall P \in \mathcal{P}_M \tag{12}$$

In other words, the malicious player always tries to send his flow through a path with the biggest possible congestion increase for every unit of flow he allocates to this path, i.e. the malicious player follows a “best value for your money” policy. Lemma 2 implies that $T_P(f^G, f^M) > 0, \ \forall P \in \mathcal{P}_M$.

## 6. Bicriteria Bound

As in the case of [RT02] we can prove a “bicriteria” result that gives an upper bound for the ratio between the cost at Wardrop equilibrium and the cost of the saddle-point solution.

**Theorem 3.** *If $f = (f^G, f^M)$ is a flow at Wardrop Equilibrium for $(G, r, F, l)$ and $\hat{f} = (\hat{f}^G, \hat{f}^M)$ is a saddle-point of (MINMAX) for $(G, 2r, F, l)$ then $C(f) \leq C(\hat{f})$.*

*Proof.* Recall that for any flow $f = (f^G, f^M)$

$$C(f) = \sum_e f_e^G \cdot l_e(f_e^G, f_e^M).$$

If $f$ is at Wardrop equilibrium, then the total latency along any flow path $P$ for good user $i$ from $s_i$ to $t_i$, $i = 1 \ldots, k$ is the same, denoted by $L_i(f)$[5], and the total cost can be expressed as $C(f) = \sum_i L_i(f) r_i$. Define a new latency function $\bar{l}_e(x, y)$ as follows:

[5] Recall that is the same as $T_P$ for user $i$.

$$\bar{l}_e(x,y) = \begin{cases} l_e(x,y) & \text{if} \quad x > f_e^G \quad \text{and} \quad y > f_e^M \\ l_e(x,f_e^M) & \quad\;\; x > f_e^G \quad \text{and} \quad y \le f_e^M \\ l_e(f_e^G,y) & \quad\;\; x \le f_e^G \quad \text{and} \quad y > f_e^M \\ l_e(f_e^G,f_e^M) & \quad\;\; x \le f_e^G \quad \text{and} \quad y \le f_e^M \end{cases} \tag{13}$$

Note that the difference $\bar{l}_e(x,f_e^M) - l_e(x,f_e^M)$ is zero for $x \ge f_e^G$. Therefore the following is true for all $x \ge 0$:

$$x(\bar{l}_e(x,f_e^M) - l_e(x,f_e^M)) \le l_e(f_e^G,f_e^M)f_e^G. \tag{14}$$

The new latency functions give a new cost (cost with respect to $\bar{l}$) that is not too far from the real cost:

$$\begin{aligned} \sum_e \bar{l}_e(\hat{f}_e^G,f_e^M)\hat{f}_e^G - C(\hat{f}^G,\hat{f}^M) &\le \sum_e \bar{l}_e(\hat{f}_e^G,f_e^M)\hat{f}_e^G - C(\hat{f}^G,f^M) \\ &= \sum_e \hat{f}_e^G(\bar{l}_e(\hat{f}_e^G,f_e^M) - l_e(\hat{f}_e^G,f_e^M)) \\ &\le \sum_e f_e^G \cdot l_e(f_e^G,f_e^M) \\ &= C(f) \end{aligned} \tag{15}$$

The first inequality is due to the fact that $\hat{f} = (\hat{f}^G,\hat{f}^M)$ is a saddle-point for $(G,2r,F,l)$, i.e. $C(\hat{f}^G,f^M) \le C(\hat{f}^G,\hat{f}^M)$ since $(\hat{f}^G,f^M)$ is a feasible solution for (MINMAX). The second inequality comes from (14) for $x := \hat{f}_e^G$.

Consider any path $P \in \mathcal{P}_i$. From the definition of $\bar{l}_e$ we have that

$$\sum_{e\in P} \bar{l}_e(0,f_e^M) = \sum_{e\in P} l_e(f_e^G,f_e^M) = L_i(f).$$

and from the fact that $\bar{l}_e(x, f_e^M)$ is an increasing function of $x$ we get

$$\sum_{e\in P} \bar{l}_e(\hat{f}_e^G, f_e^M) \geq \sum_{e\in P} \bar{l}_e(0, f_e^M).$$

Therefore:

$$\begin{aligned} \sum_{e\in E} \bar{l}_e(\hat{f}_e^G, f_e^M) \cdot \hat{f}_e^G &\geq \sum_i \sum_{P\in\mathcal{P}_i} \hat{f}_P^G \sum_{e\in P} \bar{l}_e(\hat{f}_e^G, f_e^M) \\ &\geq \sum_i \sum_{P\in\mathcal{P}_i} L_i(f)\hat{f}_P^G \qquad (16) \\ &= \sum_i 2L_i(f)r_i = 2C(f) \end{aligned}$$

By combining (15) with (16) we get $C(f) \leq C(\hat{f})$. □

The same proof also gives the following result:

**Theorem 4.** *If $f = (f^G, f^M)$ is a flow at Wardrop Equilibrium for $(G, r, F, l)$ and $\hat{f} = (\hat{f}^G, \hat{f}^M)$ is a saddle-point of (MINMAX) for $(G, (1+\gamma)r, F, l)$, $\gamma > 0$ then $C(f) \leq \frac{1}{\gamma}C(\hat{f})$.*

At a first glance, it seems rather surprising that the bicriteria bounds of [RT02] are quite robust against the existence of a malicious user. But if we look closer to the quantities compared in the theorems above, we see that while the demands of the good users are increased, the flow quantity at the disposal of the malicious user remained the same. Intuitively, the malicious user has the same power to disrupt the good users in both cases, and therefore if he settles with some strategy to do so for the initial good demands, this strategy should work about as well when the latter demands increase. The same goes for the good users' strategies as well.

## 7. Upper bounds

In this section we use the techniques of [RT02],[Rou02] to derive upper bounds for the coordination ratio. First (Section 7.1), we consider linear latency functions in detail, mainly to show that the techniques of [RT02],[Rou02] apply directly to our setting. This is extended to the more general setting of [Rou02] in Section 7.2. In both cases we prove that the malicious user as defined in Section 5 is effective.

### *7.1. Special case: linear latency functions*

In this section we deal with the special case of linear edge latency functions, i.e. for every edge $e \in E$, $l_e(f_e^G, f_e^M) = a_e(f_e^G + f_e^M) + b_e$ for some $a_e > 0, b_e > 0$. Note that we assume that the latency for an edge is positive even if no flow passes through it. This is a quite natural assumption (in all physical systems there is always some delay in moving from point A to point B, even if there is no congestion at all), and allows Theorem 1 to apply in this case. We modify our shorthand notation to $(G, r, F, a, b)$ to include the linear coefficient vectors.

In this special case we have

- $T_P(f^G, f^M) := \sum_{e \in P}(a_e f_e^G + a_e f_e^M + b_e),\ \forall i = 1, \ldots, k,\ \forall P \in \mathcal{P}_i$
- $T_P(f^G, f^M) := M(r, F) - \sum_{e \in P} a_e f_e^G,\ \forall P \in \mathcal{P}_M$

Lemma 1 and Theorem 2 take a more specific form for the linear case:

**Lemma 3.** *Let $l_e(f_e^G, f_e^M) = a_e(f_e^G + f_e^M) + b_e$ with $a_e > 0, b_e > 0$ be the latency function for every edge $e \in E$ of $G$.*

*(a) a flow $f = (f^G, f^M)$ is at Wardrop equilibrium iff*

*– for all users $i = 1, \ldots, k$ and paths $P, P' \in \mathcal{P}_i$ with $f_P > 0$*

$$\sum_{e \in P} \left(a_e f_e^G + a_e f_e^M + b_e\right) \leq \sum_{e \in P'} \left(a_e f_e^G + a_e f_e^M + b_e\right)$$

*– for all paths $P, P' \in \mathcal{P}_M$ with $f_P > 0$*

$$\sum_{e \in P} a_e f_e^G \geq \sum_{e \in P'} a_e f_e^G$$

*(b) a flow $\bar{f} = (\bar{f}^G, \bar{f}^M)$ is an equilibrium (saddle-point) for* (MINMAX) *iff*

*– for all commodities $i = 1, \ldots, k$ and paths $P, P' \in \mathcal{P}_i$ with $\bar{f}_P > 0$*

$$\sum_{e \in P} \left(2a_e \bar{f}_e^G + a_e \bar{f}_e^M + b_e\right) \leq \sum_{e \in P'} \left(2a_e \bar{f}_e^G + a_e \bar{f}_e^M + b_e\right)$$

*– for all paths $P, P' \in \mathcal{P}_M$ with $\bar{f}_P > 0$*

$$\sum_{e \in P} a_e \bar{f}_e^G \geq \sum_{e \in P'} a_e \bar{f}_e^G$$

Note that the conditions for the malicious user paths are exactly the same in both cases.

For this special form of the edge latency functions, we can prove that the saddle-point cost for (MINMAX) is unique:

**Theorem 5.** *If $f = (f^G, f^M)$ and $\bar{f} = (\bar{f}^G, \bar{f}^M)$ are two saddle-points of (MINMAX) with linear latency functions, then $f^G = \bar{f}^G$ and $C(f^G, f^M) = C(\bar{f}^G, \bar{f}^M)$.*

*Proof.* First we concentrate on a particular good user $i$ and a particular flow path $P \in \mathcal{P}_i$. Theorem 2 implies that the following two complementarity conditions hold:

$$f_P^G \left[ \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(f_e^G, f_e^M) - A_i \right] = 0$$
$$\bar{f}_P^G \left[ \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) - \bar{A}_i \right] = 0$$

Also, Theorem 2 implies that

$$\sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(f_e^G, f_e^M) \geq A_i$$
$$\sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) \geq \bar{A}_i$$

From the above, it is clear that

$$(f_P^G - \bar{f}_P^G) \left[ \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(f_e^G, f_e^M) - A_i - \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) + \bar{A}_i \right] \leq 0$$

By summing over all paths in $\mathcal{P}_i$ we get

$$\sum_{P \in \mathcal{P}_i} (f_P^G - \bar{f}_P^G) \left[ \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(f_e^G, f_e^M) - \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) \right] + (\bar{A}_i - A_i) \cdot \sum_{P \in \mathcal{P}_i} (f_P^G - \bar{f}_P^G) \leq 0$$

and therefore

$$\sum_{P \in \mathcal{P}_i} (f_P^G - \bar{f}_P^G) \left[ \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(f_e^G, f_e^M) - \sum_{e \in P} \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) \right] \leq 0$$

due to the fact that for both flows $\sum_{P \in \mathcal{P}_i} f_P^G = \sum_{P \in \mathcal{P}_i} \bar{f}_P^G = r_i$. By summing over all users $i = 1, \ldots, k$ we get

$$\sum_{e \in E} (f_e^G - \bar{f}_e^G) \left[ \frac{\partial c_e}{\partial f_e^G}(f_e^G, f_e^M) - \frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G, \bar{f}_e^M) \right] \leq 0 \tag{17}$$

We repeat the same arguments for the malicious user. More specifically, from Theorem 2 we get that

$$f_P^M\left[\sum_{e\in P}\frac{\partial c_e}{\partial f_e^M}(f_e^G,f_e^M)-B\right]=0$$

$$\bar{f}_P^M\left[\sum_{e\in P}\frac{\partial c_e}{\partial f_e^M}(\bar{f}_e^G,\bar{f}_e^M)-\bar{B}\right]=0$$

Also, Theorem 2 implies that

$$\sum_{e\in P}\frac{\partial c_e}{\partial f_e^G}(f_e^G,f_e^M)\le B$$

$$\sum_{e\in P}\frac{\partial c_e}{\partial f_e^G}(\bar{f}_e^G,\bar{f}_e^M)\le \bar{B}$$

Exactly as before we can show that

$$\sum_{e\in E}(f_e^M-\bar{f}_e^M)\left[\frac{\partial c_e}{\partial f_e^M}(f_e^G,f_e^M)-\frac{\partial c_e}{\partial f_e^M}(\bar{f}_e^G,\bar{f}_e^M)\right]\ge 0 \tag{18}$$

By substituting the cost function $c_e(f_e^G,f_e^M)=a_e{f_e^G}^2+a_ef_e^Gf_e^M+b_ef_e^G$ in (17), (18), we get

$$2\sum_{e\in E}a_e(f_e^G-\bar{f}_e^G)^2+\sum_{e\in E}a_e(f_e^G-\bar{f}_e^G)(f_e^M-\bar{f}_e^M)\le 0 \tag{19}$$

$$\sum_{e\in E}a_e(f_e^G-\bar{f}_e^G)(f_e^M-\bar{f}_e^M)\ge 0 \tag{20}$$

which implies that $f_e^G=\bar{f}_e^G$, $\forall e\in E$. But this implies that $C(f^G,f^M)=C(\bar{f}^G,\bar{f}^M)$, because otherwise, for example if $C(f^G,f^M)<C(\bar{f}^G,\bar{f}^M)$, we would also have $C(\bar{f}^G,f^M)=C(\bar{f}^G,\bar{f}^M)$, and $\bar{f}$ is not a saddle-point, contradiction. □

Lemma 3 implies the following

**Lemma 4.** *Let $(f^G, f^M)$ be a Wardrop equilibrium flow for instance $(G, r, F, a, b)$. Then the following are true:*

*(a) the flow $(f^G/2, f^M)$ is a* (MINMAX) *equilibrium for $(G, r/2, F, a, b)$*

*(b) the flow $(f^G, f^M/2)$ is a* (MINMAX) *equilibrium for $(G, r, F/2, 2a, b)$*

*(c) $\frac{\partial c_P}{\partial f^G}(f^G/2, f^M) = l_P(f^G, f^M)$ for all $P \in \mathcal{P}$.*

*Proof.* Parts (a), (b) follow directly from Lemma 3. For part (c), note that for each path $P$

$$l_P(f^G, f^M) = \sum_{e \in P}(a_e f_e^G + a_e f_e^M + b_e)$$

and

$$\frac{\partial c_P}{\partial f^G}(f^G/2, f^M) = \sum_{e \in P}(2a_e x + a_e y + b_e)\bigg|_{(x=f^G/2, y=f^M)}.$$

□

In what follows, let $\Delta_i^G(\bar{f}^G, \bar{f}^M) := \frac{\partial c_i}{\partial f^G}(\bar{f}^G, \bar{f}^M)$ be the minimum marginal cost of increasing $\bar{f}^G$ on an $s_i - t_i$ path, and $L_i(f^G, f^M)$ is the disutility for user $i$ in Wardrop equilibrium $(f^G, f^M)$.

**Lemma 5.** *Let $(\bar{f}^G, \bar{f}^M)$ be a* (MINMAX) *equilibrium for instance $(G, r, F, a, b)$. Then for any $\delta > 0$ a feasible flow for instance $(G, (1+\delta)r, F, a, b)$ has cost at least*

$$C(\bar{f}^G, \bar{f}^M) + \delta \sum_{i=1}^{k} \Delta_i^G(\bar{f}^G, \bar{f}^M) r_i.$$

*Proof.* Lemma 4.4 in [RT02]. □

We now prove our main theorem for the coordination ratio in the linear case.

**Theorem 6.** *For instance* $(G, r, F, a, b)$, $1 \leq \rho(G, r, F, a, b) \leq \frac{4}{3}$.

*Proof.* Let $(f^G, f^M)$ be a Wardrop equilibrium flow and $(\bar{f}^G, \bar{f}^M)$ a *(MINMAX)* equilibrium in $(G, r, F, a, b)$. Then, according to Lemma 4, flow $(f^G/2, f^M)$ is a *(MINMAX)* equilibrium for instance $(G, r/2, F, a, b)$ and flow $(f^G, f^M/2)$ is a *(MINMAX)* equilibrium for instance $(G, r, F/2, 2a, b)$. Therefore

$$\begin{aligned} C(\bar{f}^G, \bar{f}^M) &\geq C(\bar{f}^G, f^M) \\ &\geq C(f^G/2, f^M) + \sum_{i=1}^{k} \Delta_i^G(f^G/2, f^M)\frac{r_i}{2} \\ &= C(f^G/2, f^M) + \frac{1}{2}\sum_{i=1}^{k} L_i(f^G, f^M) r_i \\ &= C(f^G/2, f^M) + \frac{1}{2}C(f^G, f^M) \end{aligned} \tag{21}$$

where the first inequality comes from the fact that $(\bar{f}^G, \bar{f}^M)$ is a *(MINMAX)* equilibrium for $(G, r, F, a, b)$, the second inequality holds because of Lemmata 4(a) and 5, the third equality is due to Lemma 4(c), and the fourth equality holds because in a Wardrop equilibrium $(f^G, f^M)$ the latency for every $s_i - t_i$ path that carries some flow is equal to $L_i(f^G, f^M)$.

For the cost $C(f^G/2, f^M)$ of the *(MINMAX)* equilibrium for instance $(G, r/2, F, a, b)$, we have

$$\begin{aligned} C(f^G/2, f^M) &= \sum_{e \in E}(\frac{1}{4}a_e f_e^{G^2} + \frac{1}{2}a_e f_e^G f_e^M + \frac{1}{2}b_e f_e^G) \\ &\geq \frac{1}{4}\sum_{e \in E}(a_e f_e^{G^2} + a_e f_e^G f_e^M + b_e f_e^G) \\ &= \frac{1}{4}C(f^G, f^M) \end{aligned} \tag{22}$$

From inequalities (21), (22) we get $C(\bar{f}^G, \bar{f}^M) \geq \frac{3}{4}C(f^G, f^M)$, therefore $\rho(G, r, F, a, b) \leq \frac{4}{3}$.

For the lower bound of the ratio we have:

$$\begin{aligned}
C(\bar{f}^G, \bar{f}^M) &\leq C(f^G, \bar{f}^M) \\
&= \sum_{e\in E} (a_e (f_e^G)^2 + a_e f_e^G \bar{f}_e^M + b_e f_e^G) \\
&= \sum_{e\in E} (2a_e (f_e^G)^2 + 2a_e f_e^G \frac{\bar{f}_e^M}{2} + b_e f_e^G) - \sum_{e\in E} a_e (f_e^G)^2 \\
&= C(f^G, \frac{\bar{f}^M}{2}) - \sum_{e\in E} a_e (f_e^G)^2 \\
&\leq C(f^G, f^M/2) - \sum_{e\in E} a_e (f_e^G)^2 \\
&= C(f^G, f^M)
\end{aligned}$$

where $C(f^G, \frac{\bar{f}^M}{2})$ in the fourth line is the cost of flow $(f^G, \frac{\bar{f}^M}{2})$ for instance $(G, r, F/2, 2a, b)$, and the inequality in the fifth line is due to the fact that flow $(f^G, f^M/2)$ is a *(MINMAX)* equilibrium for instance $(G, r, F/2, 2a, b)$. Hence $1 \leq \rho(G, r, F, a, b)$. □

Note that the lower bound for the coordination ratio is tight, since $\rho(G, r, F, a, b) = 1$ if $G$ is just a path with the sources for all users in one end, and all the sinks in the other. Note that the upper bound is also tight, since [RT02] show that this bound is tight when there are no malicious users ($F = 0$).

### *7.2. More general latency functions*

The techniques used in the previous section for upper-bounding the coordination ratio were a straight-forward extension of the techniques in [RT02]. In fact, we can invoke the techniques of [Rou02] to obtain upper-bounds for much more general latency functions. But before we generalize the results of the previous section, we have to assume that a saddle-point of (MINMAX) *exists.* Also we don't give a direct proof of the uniqueness of the saddle-points value for (MINMAX) like Theorem 5, but we will assume that Assumption 2 holds, and this implies uniqueness of the saddle-value (cf. [Roc70] Chapter 36, especially Lemma 36.2.)

**Definition 5 ([Rou02], Def. 3.1)** *A collection $\mathcal{L}$ of latency functions is* standard *if it contains a non-zero function and if for each $l \in \mathcal{L}$, the function $x \cdot l(x)$ is convex on $[0, +\infty)$.*

In our setting, $f_e^G$ plays the role of $x$ and $l_e(f_e^G, f_e^M)$ plays the role of $l(x)$ in the definition above (as well as in the two definitions below.) For this very general family of latency functions $l$, Roughgarden [Rou02] defines the *anarchy value* $\alpha(l)$ of $l$, that turns out to capture exactly the notion of the price of anarchy for $l$. In the following definition, $\hat{l}$ is the *good marginal cost function* for $l$, i.e., for each edge $e$ $\hat{l}_e(f_e^G) = \frac{\partial}{\partial f_e^G} f_e^G l_e(f_e^G, f_e^M)$.

**Definition 6 ([Rou02], Def. 3.2)** *Let $l$ be a non-zero latency function such that $x \cdot l(x)$ is convex on $[0, +\infty)$. The* anarchy value $\alpha(l)$ *of $l$ is*

$$\alpha(l) = \sup_{r>0:l(r)>0} [\lambda\mu + (1-\lambda)]^{-1}$$

*where $\lambda \in (0,1)$ satisfies $\hat{l}(\lambda r) = l(r)$ and $\mu \in [0,1]$ is defined by $\mu = l(\lambda r)/l(r)$.*

**Definition 7 ([Rou02], Def. 3.3)** *The* anarchy value $\alpha(\mathcal{L})$ *of a standard class $\mathcal{L}$ of latency functions is*

$$\alpha(\mathcal{L}) = \sup_{0 \neq l \in \mathcal{L}} \alpha(l).$$

When our latency functions (as functions of $f^G$) belong to a latency functions class $\mathcal{L}$ as the ones above, then the proof of Roughgarden's [Rou02] main result carries over to our setting:

**Theorem 7. ([Rou02], Thm. 3.9)** *Let $\mathcal{L}$ be a standard class of latency functions with anarchy value $\alpha(\mathcal{L})$, and $l(f^G, f^M)$ a latency function that belongs to $\mathcal{L}$ as a function of $f^G$, and satisfies Assumptions 1,2. Let $(G, r, F, l)$ denote an instance of selfish routing with malicious user. Then $\rho(G, r, F, l) \leq \alpha(\mathcal{L})$.*

We also show that the malicious user defined by (12) is effective when the class of latency functions $\mathcal{L}$ satisfies Assumption 2:

**Theorem 8.** *If $\mathcal{L}$ satisfies Assumption 2, then $\rho(G, r, F, l) \geq 1$.*

*Proof.* Let $f, \bar{f}$ be a Wardrop equilibrium and a saddle point respectively for instance $(G, r, F, l)$. Then, if $C(\cdot)$ is the social cost function, and since $\bar{f}$ is a saddle point, we have

$$C(\bar{f}^G, \bar{f}^M) \leq C(f^G, \bar{f}^M)$$

Since $f$ is a Wordrop equilibrium, we have

$$T_P(f^G, f^M) \leq T_{P'}(f^G, \bar{f}^M), \quad \forall P, P' \in \mathcal{P}_M \text{ with } f_P^M > 0$$

which implies that

$$\sum_{e \in P} \frac{\partial c_e}{\partial f_e^M}(f^G, f^M) \geq \sum_{e \in P'} \frac{\partial c_e}{\partial f_e^M}(f^G, \bar{f}^M), \quad \forall P, P' \in \mathcal{P}_M \text{ with } f_P^M > 0$$

Hence, it follows that

$$C(f^G, \bar{f}^M) \leq C(f^G, f^M)$$

in a way completely analogous to the proof of Theorem 2. □

Again, this is a tight lower bound for the same reason as in the previous section. The tightness of the upper bound is also proven by Roughgarden [Rou02] (Theorem 4.4.)

## 8. Discussion and open problems

Our study concentrated on a single malicious user, but it is obvious from the above that it can be extended to many malicious users (each with his/her own flow rate and set of available paths $\mathcal{P}_{\mathcal{M}}$), just as we can have many good users. It should be clear though that by allowing a single malicious user, the malicious behavior is potentially more powerful, because the single malicious user can allocate any portion of his/her flow rate to a specific set of paths, which is a subset of the available paths. This is not the case when a particular subset is allocated to a separate malicious user who *must* route a specific amount of flow through them. Therefore, it is not surprising that our results extend to a setting with many malicious users, since they hold for the potentially more powerful setting with a single malicious user (and the same set of available paths).

The model presented in our work gives rise to many open problems. It would be very interesting to present a natural definition and results connecting the social cost of an equilibrium point in a network with malicious users and the cost in an equivalent instance without malicious users. This would give a clearer characterization of the negative impact of the presence of malicious flow. The model defined in our work gives rise to unique saddle-points and Wardrop equilibria. It would be interesting to consider a more general model that allows multiple equilibria (for example, by adding capacities for the edges in the network [SM03]) and analyze the performance of the system in the presence of malicious users.

## Acknowledgments

We are greatly indebted to an anonymous reviewer, whose thorough and detailed review helped us to greatly improve this manuscript. We would also like to thank Tamás Terlaky and Nicola Galesi for many helpful discussions.

## References

[AM81] H. Z. Aashtiani and T. L. Magnanti. Equilibria on a congested transportation network. *SIAM Journal on Algebraic and Discrete Methods*, 2(3):213–226, September 1981.

[BMW56] M. Beckmann, C. B. McGuire, and C. B. Winsten. Studies in the economics of transportation. *Yale University Press*, 1956.

[CV02] A. Czumaj and B. Vöcking. Tight bounds for worst-case equilibria. In *Proceedings of the 13th Annual ACM-SIAM Symposium On Discrete Mathematics*, pages 413–420, 2002.

[DS69] S. Dafermos and F. Sparrow. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards*, 73B:91–118, 1969.

[HM85] A. Haurie and P. Marcotte. On the relationship between Nash-Cournot and Wardrop equilibria. *Networks*, 15, pp. 295–308, 1985.

[KV03] G. Karakostas and A. Viglas. Equilibria for networks with malicious users. In *Proceedings of the 14th Annual International Symposium on Algorithms and Computation (ISAAC)*, pages 696–704, 2003.

[KP99] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science, LNCS 1563*, pages 404–413, 1999.

[MS01] M. Mavronicolas and P. Spirakis. The price of selfish routing. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 510–519, 2001.

[P20] A. Pigou. *The economics of Welfare.* Macmillan, London, 1920.

[Roc70] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[Rou01a] T. Roughgarden. Designing networks for selfish users is hard. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 472–481, 2001.

[Rou01b] T. Roughgarden. Stackelberg scheduling strategies. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 104–113, 2001.

[Rou02] T. Roughgarden. The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences*, 67(2): 341–364, 2003.

[RT02] T. Roughgarden and É. Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, March 2002.

[SM03] A. S. Schulz and N.E. Stier Moses. On the performance of user equilibria in traffic networks. In *14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 86–87, 2003.