Congestion Games with Malicious Players

Moshe Babaioff*
School of Information
University of California at Berkeley
Berkeley, CA, 94720 USA
moshe@ischool.berkeley.edu

Robert Kleinberg†
Department of Computer Science
Cornell University
Ithaca, NY 14853
rdk@cs.cornell.edu

Christos H. Papadimitriou‡
Computer Science Division
University of California at Berkeley
Berkeley, CA, 94720 USA
christos@cs.berkeley.edu

ABSTRACT

We study the equilibria of non-atomic congestion games in which there are two types of players: rational players, who seek to minimize their own delay, and malicious players, who seek to maximize the average delay experienced by the rational players. We study the existence of pure and mixed Nash equilibria for these games, and we seek to quantify the impact of the malicious players on the equilibrium. One counterintuitive phenomenon which we demonstrate is the "windfall of malice": paradoxically, when a myopically malicious player gains control of a fraction of the flow, the new equilibrium may be more favorable for the remaining rational players than the previous equilibrium.

Categories and Subject Descriptors

F.0 [**Theory of Computation**]: General

General Terms

Economics, Theory

Keywords

Selfish Routing, Malicious Behavior, Equilibrium, Congestion Games

1. INTRODUCTION

Game Theory is the study of strategic behavior of rational agents. Described this way, Game Theory sounds a little unrealistic, because much of what is going on in the real world is, intuitively, irrational. The standard game-theoretic response to this line of criticism is that what we intuitively call "irrational agents" are just rational players with very strange utility functions. This is a reasonable retort when it refers to generic games. However, much work in Game Theory, and especially in the interface between Game Theory, Networking, and Computation, is about standard types of games — such as auctions, congestion games, facility location games, network creation games, etc. Many of these are studied with the explicit ambition to apply the results to the real world. Since the utilities of these games cannot be arbitrarily "strange," the original criticism stands.

In this paper we model one type of what is usually meant by "irrationality" in the above argument, namely *malicious players*. We define malicious players in the context of a particular *symmetric game;* suppose that, in an n-player symmetric game, the utilities of $m < n$ players, henceforth called malicious, change from the common utility shared by all players *to the negative sum of the utilities of the* $n - m$ *non-malicious players.* We are interested in the effect such change has on the quality as well as nature (pure versus mixed) of the game's Nash equilibria. We define the *price of malice* to be the relative deterioration of the sum of the utilities of the non-malicious players when the remaining players turn malicious.

That malice has a price is not very surprising; what is somewhat unexpected is that *this price can be negative*, and malicious players may improve system performance, intuitively because their presence may incentivize the other players to forego antisocial selfish behavior. For example, consider a version of the prisoner's dilemma with three players and three strategies: collaborate, defect, and *inspect.* When a player inspects, her own utility is negative, but that of any defecting player deteriorates as well. It is easy to see that the numbers can be set in such a way that, if any of the three players turns malicious (and therefore inspects, sacrificing her own well-being in order to hurt the others), then the other players end up collaborating at equilibrium, for a net increase in their sum of utilities — in fact, a relative increase that can be arbitrarily high.

We study the effects of malicious agents on non-atomic congestion games [13]. In such games a continuum of players, comprising a flow of some specified value v, choose routes in a network with a single source and sink whose edges have load-dependent delays. It is known that such a game has a pure Nash equilibrium with equal delays for all; the quality of this equilibrium (compared to the "social optimum" minimum delay flow) has been studied extensively

*Research supported by NSF ITR Award ANI-0331659.
†Research supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship. Portions of this work were completed while the author was an NSF postdoctoral fellow at U.C. Berkeley.
‡Research supported by NSF grant CCF-0515259 and a grant from Yahoo Research and a MICRO grant.

EC'07, June 11–15, 2007, San Diego, California, USA.

[13, 12]. But suppose instead that some fraction of the flow becomes controlled by a *malicious player* whose utility is the total delay by each of the other players. Does this new game have a pure Nash equilibrium? And, if it does, how does it compare with the equilibrium with $\mu = 0$ (no malicious player)? We define the *price of malice* to be the limit of this deterioration per unit of malicious flow, as μ goes to zero.

If flows are allowed to contain cycles (and therefore a malicious player can make loads arbitrarily high by going around in circles indefinitely), then it is easy to construct situations in which a malicious player can wreak havoc on a network (strictly speaking, such situations do not have a Nash equilibrium, and so we cannot speak of a price of malice). We show (Theorem 2) that even in acyclic networks the price of malice can be significant; upper bounding the price of malice by the network parameters (such as the number of edges and the "relative slope" of the delays) is an important open problem.

Perhaps the heaviest price of malice may be the fact that the presence of a malicious player *upsets the Nash equilibrium regime of congestion games.* Ordinarily, congestion games are known to always have a pure Nash equilibrium. In contrast, in Section 4 we notice that, in the presence of a malicious player, pure Nash equilibria may not exist. However, we prove two compensating results: First, there is always a "semi-pure" Nash equilibrium, in which only the malicious player mixes strategies. Second, if the delays are weakly concave (and in particular if they are linear) then pure Nash equilibria exist. The existence proofs rely on Kakutani's Fixed Point Theorem and Prokhorov's Theorem, two powerful results that we have not seen before applied in the context of congestion games.

1.1 Related work

Several recent papers have considered agents that are not acting rationally, and while the general philosophical direction of our work is somewhat similar to these works, there are still significant differences between our work and the papers we describe below.

Two papers [7, 2] consider auctions with agents that derive utility from the disutility of others, and present similar results. Both papers derive symmetric Bayes Nash equilibria for spiteful agents in 1st-price and 2nd-price sealed bid auctions. A spiteful agent's value for an outcome is a convex combination of his own original profit and the total loss of the other agents (taken with coefficient α, the *spite* coefficient). The papers consider the equilibrium when *all* agents are spiteful with the same coefficient (unlike in our model in which only a small fraction of the flow is controlled by a malicious player, and this player is purely malicious, i.e. spiteful with coefficient 1). Interestingly they show that the revenue equivalence between second-price and first-price auctions breaks down with spiteful agents, with second-price outperforming first-price.

Eliaz [3] considers the problem of implementation when some agents are faulty, playing an arbitrary strategy, possibly in a malicious way. The paper presents a solution concept to allow implementation in case that up to k agents are faulty, but neither their identity nor their exact number are known. Unlike our model (which assumes that a malicious player will choose a strategy that maximizes the combined discontent of the non-malicious players, given the strategic choices of all other players) the non-rational agents in their model can play *arbitrarily* and the paper focuses on implementation issues, while we focus on quantifying the implications of malice on given systems.

Closest to our work is the paper by Karakostas and Viglas [6] (henceforth KV) which studies equilibria for network congestion games with malicious users. The KV model of malicious behavior corresponds to a continuum of infinitesimal malicious players, collectively controlling the malicious flow. We allow for a more powerful malicious behavior by allowing coordination (modeled as a single myopic malicious agent controlling all the malicious flow). We show that coordination indeed leads to a different equilibrium concept for some networks, but not when latency functions are concave. The difference between the two equilibrium concepts, for general networks, arises from the fact that a single malicious player can use mixed strategies that are unavailable to a continuum of uncoordinated malicious players. The focus of the KV paper is on generalizing the notion of Price of Anarchy to the case that there is also a malicious flow in the system, by comparing the case that the good users are controlled by a single entity, to the case that they are behaving selfishly. One of the main open problems in this paper is the connection between the social cost at an equilibrium point with and without malicious users. Our work addresses this issue by defining the notion of Price of Malice and studying it.

Closest in spirit to our study of the Price of Malice is the paper by Moscibroda et al. [8] which aims to study the implications of malicious behavior on systems consisting of selfish agents. The paper presents a concept of Price of Malice and says "The Price of Malice is a ratio that expresses how much the presence of malicious players deteriorates the social welfare of a system consisting of selfish players.". Yet, there are many differences between this paper and ours. Important differences exist in the definition of equilibrium in the presence of malice and the definition of the Price of Malice. First, selfish players in their game are extremely risk averse and basically each one perceives the malicious agents as if they are all attacking him or her. Second, the definition of the Price of Malice is very different, as they look at the ratio between two different worst-case ratios (the price of anarchy with b malicious agents and with 0 malicious agents) even though those worst-case ratios may arise on different problem instances. Instead of this type of indirect comparison, we directly compare the outcome of games with a mixture of rational and malicious agents to the outcome with only rational agents.

2. THE PRICE OF MALICE

2.1 Definitions

2.1.1 Non-atomic congestion games

The following definitions are standard from the theory of congestion games, e.g. [11], and readers familiar with this material are encouraged to proceed to Section 2.1.2. We use $\mathbb{R}_+$ to denote the set of non-negative real numbers.

DEFINITION 1 (CONGESTION GAME). *A* symmetric non-atomic congestion game *(henceforth, simply called a "congestion game") is specified by an ordered quadruple* $\mathcal{G} = (E, \vec{\ell}, \Pi, v)$, *whose components are called:*

- *the* edge set E*, a finite set;*
- *the vector of* latency functions $\vec{\ell}$*, a function from* $\mathbb{R}_+$ *to the vector space* $\mathbb{R}^E$*: for* $e \in E$*, the* e*-th component of* $\vec{\ell}$ *is denoted by* ℓ_e*, and is a non-decreasing function from* $\mathbb{R}_+$ *to* $\mathbb{R}_+$*;*
- *the* path set Π*, a subset of* 2^E*; and*
- *the* flow value v*, a non-negative real number which we will sometimes denote by* $v(\mathscr{G})$*.*

If E is the edge set of a (directed or undirected) graph G, and Π is a set of paths in G, then we call $\mathscr{G}$ a network congestion game*. We will use the terminology "edge" and "path" in describing abstract congestion games, though in the general case we do not expect E to be interpreted as a set of edges of a graph nor P as a set of paths in a graph.*

DEFINITION 2 (FLOW). *A* flow *in a congestion game $\mathscr{G}$ is a function f from Π to $\mathbb{R}_+$. (One interprets $f(P)$ as the amount of flow using path P.) The* flow value *is $v(f) = \sum_{P\in\Pi} f(P)$. The set of all flows in $\mathscr{G}$ is denoted by $F(\mathscr{G})$. The set of all flows whose flow value is equal to some number w is denoted by $F(\mathscr{G}, w)$. The set $F(\mathscr{G}, w)$ is a compact, convex set (in fact, a convex polytope) in $\mathbb{R}^\Pi$. We will abbreviate $F(\mathscr{G}, w)$ to F when $\mathscr{G}, w$ are understood from context.*

DEFINITION 3 (COST). *If f is a flow, the* load *on an edge $e \in E$ is*

$$x_e(f) = \sum_{P\in\Pi | e\in P} f(P).$$

The delay *on a path $P \in \Pi$ is*

$$L(P) = \sum_{e\in P} \ell_e(x_e(f)).$$

The cost *of f is*

$$C(f) = \sum_{P\in\Pi} f(P)L(P) = \sum_{e\in E} x_e(f)\ell_e(x_e(f)).$$

DEFINITION 4 (NASH FLOW). *If f is a flow, the set of* best responses *to f is the set* $\arg\min_{P\in\Pi} L(P)$*. A flow f in a congestion game $\mathscr{G}$ is a* Nash flow *if $v(f) = v(\mathscr{G})$ and every path $P \in \Pi$ which satisfies $f(P) > 0$ is a best response to f. The* Nash cost *and* Nash delay *of $\mathscr{G}$, denoted by $C(\mathscr{G})$ and $D(\mathscr{G})$, are the quantities $C(f)$ and $D(f) = C(f)/v(f)$, respectively, where f is any Nash flow of $\mathscr{G}$. We will see in Proposition 1 below that $C(\mathscr{G})$ and $D(\mathscr{G})$ do not depend on the choice of the Nash flow f.*

DEFINITION 5 (POTENTIAL FUNCTION). *For a congestion game $\mathscr{G}$, the* potential function $\Phi_{\mathscr{G}}$ *(denoted simply by Φ when the game $\mathscr{G}$ is understood from context) is a real-valued function on $F(\mathscr{G})$ defined by*

$$\Phi_{\mathscr{G}}(f) = \sum_{e\in E} \int_0^{x_e(f)} \ell_e(y)\, dy.$$

The following standard facts about the potential function will be useful to us.

PROPOSITION 1 ([13]). *The potential function $\Phi = \Phi_{\mathscr{G}}$ is a convex function on $F(\mathscr{G})$. It is strictly convex if all of the latency functions ℓ_e are strictly increasing. For a flow f of value $v(\mathscr{G})$, the following are equivalent:*

1. *f is a local minimum of Φ.*
2. *f is a global minimum of Φ.*
3. *f is a Nash flow.*

Moreover, for any two Nash flows $f, \tilde{f}$ we have $C(f) = C(\tilde{f})$, and furthermore the Nash delay $D(f)$ is equal to the delay $L(P)$ on any path $P \in \Pi$ satisfying $f(P) > 0$.

2.1.2 Congestion games with malicious players

DEFINITION 6 (MALICIOUS PLAYER). *A* congestion game with a malicious player *is specified by a congestion game $\mathscr{G}$ together with a real number $w(\mathscr{G})$ satisfying $0 \le w(\mathscr{G}) \le v(\mathscr{G})$. We interpret $w(\mathscr{G})$ as the amount of flow controlled by the malicious player.*

When the malicious player routes its flow using a particular (possibly randomized) flow g, the remaining flow (controlled by the rational players) is, in effect, participating in a modified congestion game whose latency functions have been changed to reflect the load imposed by the malicious player. We now define this notion precisely.

DEFINITION 7 (INDUCED GAME). *Let $\mathscr{G} = (E, \vec{\ell}, \Pi, v)$ be a congestion game with a malicious player, $w = w(\mathscr{G})$, and γ a probability measure on $F(\mathscr{G}, w)$. The* induced latency function ℓ_e^γ *on an edge e is defined by*

$$\ell_e^\gamma(x) = \mathbf{E}(\ell_e(x + x_e(g))),$$

where g is a random sample from the distribution γ. The induced game $\mathscr{G}^\gamma$ *is the congestion game $(E, \vec{\ell}^\gamma, \Pi, v - w)$. If f is a flow in $\mathscr{G}$, the* induced cost $C^\gamma(f)$ *is the cost of f in the induced game $\mathscr{G}^\gamma$. When γ is a point mass concentrated on a single flow $g \in F(\mathscr{G}, w)$, we will use the notation $\mathscr{G}^g$ (resp. C^g, ℓ_e^g) to mean the same thing as $\mathscr{G}^\gamma$ (resp. C^γ, ℓ_e^γ).*

DEFINITION 8 (MALICIOUS BEST RESPONSE). *If f is a flow in $\mathscr{G}$, the set of* malicious best responses *to f is the set*

$$MBR(f) = \arg\max_{g\in F(\mathscr{G},w)} C^g(f).$$

A probability measure on $F(\mathscr{G}, w)$ is a malicious best response to f if it is supported on the set $MBR(f)$.

We are now in a position to define the equilibria of a congestion game with a malicious player. Intuitively, a pair of flows (f, g) — with f representing the rational players and g representing the malicious player — is an equilibrium if none of the rational players can unilaterally improve their delay by switching to a different path, and if the malicious player can not inflict greater damage on the rational players by shifting from g to some other flow. In order to guarantee the existence of equilibria, it is necessary to allow the malicious player to use a mixed strategy, i.e. to sample a random flow from $F(\mathscr{G}, w)$. Thus an equilibrium is actually a pair (f, γ) where f is a flow of value $v - w$ and γ is a distribution on the set of flows of value w.

DEFINITION 9 (EQUILIBRIUM). *If $\mathscr{G}$ is a congestion game with a malicious player and $w = w(\mathscr{G})$ is the amount of malicious flow, then an* equilibrium *of $\mathscr{G}$ is an ordered pair (f, γ) such that f is a Nash flow in the induced game $\mathscr{G}^\gamma$, and γ is a malicious best response to f. An equilibrium is* pure *if γ is a point mass concentrated on a single flow $g \in F(\mathscr{G}, w)$.*

DEFINITION 10 (NASH DELAY). *Let $\mathscr{G}$ be a congestion game with a malicious player, $w = w(\mathscr{G})$, and $\mathcal{E}$ the set of equilibria of $\mathscr{G}$. The Nash delay $D(\mathscr{G}, w)$ is defined to be the supremum of the set $\{C^\gamma(f)/v(f) \mid (f,\gamma) \in \mathcal{E}\}$.*

Note that, in order for $D(\mathscr{G}, w)$ to be well-defined, it must be the case that the set of equilibria of $\mathscr{G}$ (with w units of malicious flow) is nonempty. We will see in Section 4 that this is indeed the case.

For a congestion game $\mathscr{G}$ with flow value $v = v(\mathscr{G})$, the price of malice measures the rate at which the Nash delay deteriorates as a small fraction of the flow comes under the control of a malicious player.

DEFINITION 11 (PRICE OF MALICE). *The* price of malice, $\mathsf{POM}(\mathscr{G})$, *is defined by*

$$\mathsf{POM}(\mathscr{G}) = \lim_{\varepsilon \to 0} \frac{D(\mathscr{G}, \varepsilon v) - D(\mathscr{G})}{\varepsilon D(\mathscr{G})} \tag{1}$$

when the limit exists.

Note that the price of malice quantifies the first order effect of a small fraction of malicious flow. Clearly one can change the definition to capture lower order effects (for example an $O(\varepsilon^2)$ increase in relative delay).

A counterintuitive phenomenon which we will explore later in this paper is the *windfall of malice*, whereby the presence of a malicious player in the game actually improves the delay experienced by the rational players. We say that a game exhibits windfall of malice if it has a negative price of malice.

2.2 A differential criterion for equilibrium

It is useful to relate the definition of a malicious best response given above (Definition 8) to a criterion which is based on the derivatives of the latency functions, and which says that the malicious player's flow should be distributed on paths which maximize the marginal cost (to the rational players) per unit of flow. Throughout this section we assume that $\mathscr{G}$ is a congestion game with differentiable latency functions.

DEFINITION 12 (DIFFERENTIAL MBR). *Let $\mathscr{G}$ be a congestion game with differentiable latency functions. Consider any two flows $f, g \in F(\mathscr{G})$. We say that g is a* differential malicious best response (DMBR) *to f if for every two paths $P, P' \in \Pi$ such that $g(P) > 0$, we have*

$$\sum_{e \in P} x_e(f) \ell_e'(x_e(f) + x_e(g)) \geq \sum_{e \in P'} x_e(f) \ell_e'(x_e(f) + x_e(g)).$$

This definition is closely related to the definition of malicious best response implied by equation (9) in [6]. Indeed, we will see that being a DMBR to f is always a necessary condition for being a malicious best response to f, and that when the latency functions are concave it is also a sufficient condition. Thus our definition of malicious best response (hence also our definition of equilibrium) is equivalent to the definition given by Karakostas and Viglas [6] in the special case when latency functions are concave.

LEMMA 1. *Every malicious best response to f is a DMBR to f.*

PROOF. Let g be a malicious best response to f, and let $P, P' \in \Pi$ be two paths such that $g(P) > 0$. For $t \geq 0$, consider the flow $g^{(t)}$ defined by

$$g^{(t)}(Q) = \begin{cases} g(Q) & \text{if } Q \neq P, P' \\ g(Q) - t & \text{if } Q = P \\ g(Q) + t & \text{if } Q = P' \end{cases}$$

If $w = v(g)$ then $g^{(t)} \in F(\mathscr{G}, w)$ for $t \in [0, g(P)]$. Since $g \in \arg\max_{h \in F(\mathscr{G},w)} C^h(f)$ we have

$$\frac{d}{dt}\left(C^{g^{(t)}}(f)\right)_{t=0} \leq 0.$$

The left side is equal to $\sum_{e \in P'} x_e(f) \ell_e'(x_e(f) + x_e(g)) - \sum_{e \in P} x_e(f) \ell_e'(x_e(f) + x_e(g))$. □

LEMMA 2. *If g is a DMBR to f, then for every flow h of value $v(g)$,*

$$\sum_{e \in E} x_e(f) \ell_e'(x_e(f) + x_e(g)) [x_e(g) - x_e(h)] \geq 0. \tag{2}$$

PROOF. For any path P, define $B(P)$ to be the sum

$$B(P) = \sum_{e \in P} x_e(f) \ell_e'(x_e(f) + x_e(g)).$$

The left side of (2) is equal to $\sum_{P \in \Pi} [g(P) - h(P)] B(P)$. Hence (2) is equivalent to

$$\sum_{P \in \Pi} g(P) B(P) \geq \sum_{P \in \Pi} h(P) B(P). \tag{3}$$

If $M = \max_{P \in \Pi} B(P)$ then by the definition of a DMBR, we have $B(P) = M$ for every path P such that $g(P) > 0$; hence the left side of (3) is equal to $v(g) \cdot M$. Similarly the right side is bounded above by $v(g) \cdot M$. □

THEOREM 1. *Assume that $\mathscr{G}$ is a congestion game with malicious players and for every edge e, ℓ_e is a differentiable, weakly concave function. Then a flow g is a DMBR to f if and only if g is a malicious best response to f.*

PROOF. By Lemma 1 every malicious best response to f is a DMBR to f, so we are left to show that every DMBR to f is a malicious best response to f.

For an edge e, let λ_e be the function

$$\lambda_e(x) = \ell_e(x_e(f)+x_e(g)) + \ell_e'(x_e(f)+x_e(g))(x - x_e(f) - x_e(g)).$$

This is a linear function of x which satisfies

$$\begin{aligned} \lambda_e(x_e(f) + x_e(g)) &= \ell_e(x_e(f) + x_e(g)) \\ \lambda_e'(x_e(f) + x_e(g)) &= \ell_e'(x_e(f) + x_e(g)). \end{aligned}$$

Since ℓ_e is concave and λ_e is a linear function whose value and first derivative agree with those of ℓ_e at $x_e(f) + x_e(g)$, we may conclude that $\lambda_e(x) \geq \ell_e(x)$ for all x. Now suppose that g is a DMBR to f, and h is any flow of value $v(g)$. By Lemma 2 we have

$$\begin{aligned} \sum_{e \in E} x_e(f) \ell_e'(x_e(f) + x_e(g)) [x_e(h) - x_e(g)] &\leq 0 \\ \sum_{e \in E} x_e(f) [\lambda_e(x_e(f) + x_e(h)) - \lambda_e(x_e(f) + x_e(g))] &\leq 0 \end{aligned}$$

$$\sum_{e \in E} x_e(f) \lambda_e(x_e(f) + x_e(h)) \leq \sum_{e \in E} x_e(f) \lambda_e(x_e(f) + x_e(g)).$$

Now using the fact that $\lambda_e(x) \geq \ell_e(x)$ for all x, with equality when $x = x_e(f) + x_e(g)$, we obtain

$$\sum_{e \in E} x_e(f)\ell_e(x_e(f) + x_e(h)) \leq \sum_{e \in E} x_e(f)\ell_e(x_e(f) + x_e(g)).$$

As h was an arbitrary flow of value $v(g)$, this confirms that g is a malicious best response to f. □

Our definition of malicious best response may be regarded as the appropriate definition for modeling a single (myopically) malicious player controlling w units flow, while the definition of differential malicious best response models a continuum of infinitesimal malicious players, collectively controlling w units of flow. (Definition 12 is tantamount to asserting that one cannot increase $C^g(f)$ by rerouting an infinitesimal amount of flow.) Since all malicious players experience the same payoff, it is plausible that w units of flow controlled by a continuum of such players will behave identically to the same amount of flow controlled by a single malicious player. Indeed, Lemma 1 shows that this is exactly what happens when the latency functions are concave. Interestingly, this is not what happens in general when latency functions can be non-concave (see Example 1). The reason is that a single malicious player has the power to play a mixed strategy, while this can never happen with a continuum of malicious players unless we allow them to correlate their random choices. (Even if each of the infinitesimal players uses a mixed strategy, if their random choices are independent then the law of large numbers ensures that their combined flow is equal to a single element of $F(\mathscr{G}, w)$ with probability 1.)

We next show that in any equilibrium with small enough malicious flow, the malicious flow uses only paths that maximize the per-unit cost at the Nash equilibrium.

Definition 13 (differentially malicious flow). *Let $\mathscr{G}$ be a congestion game with differentiable latency functions. Consider a Nash flow $f_N \in F(\mathscr{G})$. A path $P^* \in \Pi$ is a* differentially malicious path w.r.t. *f_N if for every $P \in \Pi$*

$$\sum_{e \in P^*} x_e(f_N)\ell'_e(x_e(f_N)) \geq \sum_{e \in P} x_e(f_N)\ell'_e(x_e(f_N))$$

A flow g is differentially malicious w.r.t. *f_N if for any $P \in \Pi$ such that $g(P) > 0$, P is a differentially malicious path w.r.t. f_N.*

Proposition 2. *Let $\mathscr{G}$ be a congestion game with a malicious player, with continuously differentiable latency functions. For any Nash flow $f_N \in F(\mathscr{G})$ there is an open set $U \subseteq F(\mathscr{G})$ containing f_N, such that for any $f \in U$, it holds that every $g \in MBR(f)$ (of value $v(f_N) - v(f)$) is also differentially malicious w.r.t. f_N.*

Proof. For any path $P \in \Pi$, define a two-variable function $h_P : F(\mathscr{G}) \times F(\mathscr{G}) \to \mathbb{R}$ as follows:

$$h_P(f, g) = \sum_{e \in P} x_e(f)\ell'_e(x_e(f) + x_e(g)).$$

Observe that h_P is a continuous function because ℓ_e is continuously differentiable for every edge e. Observe also that g is a DMBR to f if and only if $\{P : g(P) > 0\} \subseteq \arg\max_{P \in \Pi} h_P(f, g)$.

Now let Π_{mal} denote the set of all paths which are differentially malicious w.r.t. f_N, i.e. $\Pi_{mal} = \arg\max_{P \in \Pi} h_P(f_N, 0)$.

If $\Pi_{mal} = \Pi$ then every path is differentially malicious and the proposition follows trivially. Otherwise, the two-variable function $h(f, g)$ defined by

$$h(f, g) = \min\{h_{P^*}(f, g) - h_P(f, g) \,|\, P^* \in \Pi_{mal}, P \notin \Pi_{mal}\}$$

is continuous and satisfies $h(f_N, 0) > 0$, by the definition of Π_{mal}. Therefore the set $W = \{(f, g) : h(f, g) > 0\}$ is an open neighborhood of $(f_N, 0)$ in $F(\mathscr{G}) \times F(\mathscr{G})$. Let $W_1 \times W_2$ be an open subset of W such that $f_N \in W_1$, $0 \in W_2$. Without loss of generality (replacing W_1, W_2 with smaller open neighborhoods of $f_N, 0$ if necessary) we may assume that for some real number $\delta > 0$,

$$\begin{aligned} W_1 &\subseteq \{f \in F(\mathscr{G}) \,|\, v(f) > v(\mathscr{G}) - \delta\} \\ W_2 &= \{g \in F(\mathscr{G}) \,|\, v(g) < \delta\} \end{aligned}$$

We claim that $U = W_1$ satisfies the conclusion of the proposition. For any $f \in W_1$, if g is a malicious best response to f of value $v(\mathscr{G}) - v(f)$, then $v(g) < \delta$ hence $g \in W_2$. Thus $(f, g) \in W_1 \times W_2 \subseteq W$, which implies $h(f, g) > 0$. By the definition of $h(f, g)$, this implies that $\arg\max_{P \in \Pi} h_P(f, g) \subseteq \Pi_{mal}$. Recalling that g is a malicious best response (and hence, by Lemma 1, a DMBR) to f, we see that every path P with $g(P) > 0$ is an element of $\arg\max_{P \in \Pi} h_P(f, g)$, hence every such P belongs to Π_{mal}. □

2.3 Lower bound on the price of malice

In this section we construct network congestion games with a large price of malice. Intuitively, the price of malice can be large for at least two reasons:

1. The network contains some edges whose latency functions grow very rapidly, so that a small amount of additional flow can have a very large impact on the delay.

2. The network contains a very long path, so that the malicious player can send its flow on this path and thereby influence many of the paths being used by the rational players.

We capture the first property using the notion of *relative slope* of the latency functions, which has also been used elsewhere in the literature on selfish routing, e.g. [4]. We capture the second property by building a congestion game with a unique equilibrium, namely a pure equilibrium in which the rational players use many disjoint short paths and the malicious player uses a single long path that intersects all of the short paths.

Definition 14. *Let $\ell : [0, 1] \to \mathbb{R}_+$ be a continuous nondecreasing function that is continuously differentiable. The* relative slope *of ℓ is defined to be the number*

$$d = \sup_{x \in [0,1]} \frac{x\ell'(x)}{\ell(x)}.$$

Theorem 2. *Let $\ell : [0, 1] \to \mathbb{R}_+$ be a continuous nondecreasing function that is continuously differentiable, and let d be the relative slope of ℓ. For any m there exists a network congestion game with $O(m)$ edges, such that the latency function of each edge is either ℓ or 0, and such that the price of malice is $d(m - 1)$.*

Proof. The continuous function $\frac{x\ell'(x)}{\ell(x)}$ achieves its supremum, d, at some point of the interval $[0, 1]$ because $[0, 1]$

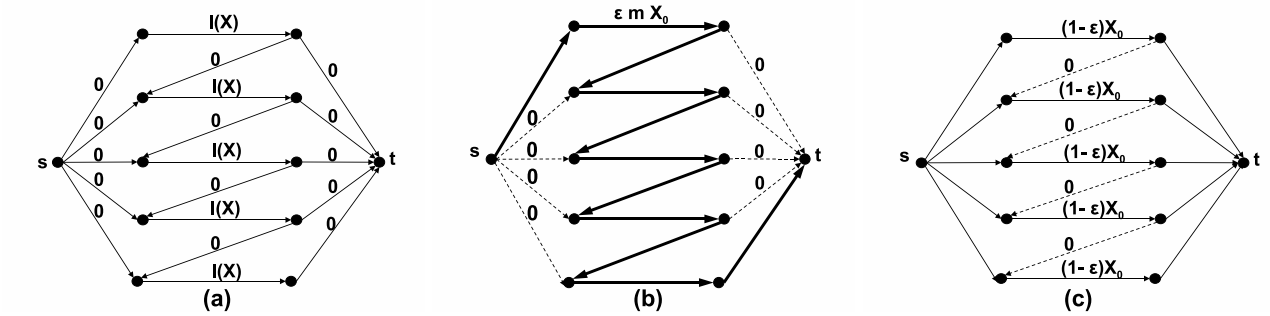


Figure 1: A network congestion game with a large price of malice. (a) The congestion game. (b) The malicious player's equilibrium strategy. (c) The rational players' equilibrium strategy.

is compact. Let X_0 be a point where the supremum is achieved. The network is illustrated in Figure 1(a) for $m = 5$. In this network congestion game the flow value is mX_0. The network has m parallel paths of length 3, all have the same latency function $\ell(x)$ on the middle edge. All other edges have a constant latency 0. Backward edges enable the malicious flow to travel all the edges with non-zero latency functions (a path of length $2m + 1$).

Figure 1(b) illustrates the path that the malicious flow of size ϵmX_0 takes. As the latency functions are the same on every one of the m paths, in equilibrium the rational flow of size $(1 - \epsilon)mX_0$ will be split equally on the m paths, thus in equilibrium on each path there is a rational flow of size $(1 - \epsilon)X_0$ (in case that $\epsilon = 0$ this means a flow of X_0). This is illustrated in Figure 1(c). The total flow on each of the middle edges of the m paths is $(1 - \epsilon)X_0 + \epsilon mX_0 = X_0 + \epsilon X_0(m - 1)$.

The latency with ϵ units of malicious flow is $\ell(X_0+\epsilon X_0(m-1))$, and the latency with no malicious flow is $\ell(X_0)$. Using the fact that $\lim_{\varepsilon\to 0} \frac{\ell(x+a\epsilon)-\ell(x)}{\epsilon} = a \cdot \ell'(x)$, for $a = X_0(m - 1)$ we obtain that the price of malice is

$$\begin{aligned}\mathsf{POM}(\mathscr{G}) &= \lim_{\varepsilon\to 0} \frac{\ell(X_0 + \epsilon X_0(m-1)) - \ell(X_0)}{\epsilon\ell(X_0)} \\ &= \frac{X_0(m-1)\ell'(X_0)}{\ell(X_0)} = (m-1)\cdot d\end{aligned}$$

□

3. THE WINDFALL OF MALICE

The following claims show that there exists a network with "windfall of malice", that is, replacing some of the rational flow with malicious flow causes the delay of the rational flow to *decrease*. At first look it seems surprising that there can be a *decrease* in the latency experienced by the rational agents, as the malicious agent is trying to *maximize* the latency of the rational agents. But this phenomenon is not too different from the well-known Braess' paradox, which gives an example of a network for which an *increase* in the latency function on an edge *improves* the Nash delay. While in the Braess' paradox network there is no windfall of malice, we are able to construct a network that is based on that network that does have a windfall. In the network that we construct, the malicious agent, by trying to do as much harm as possible, increases the latency on every possible edge, and by doing so it causes the rational agents to take alternative routes that are less harmful to the other rational agents.

Claim 1. *There exists a network congestion game for which the price of malice is negative.*

Proof. We construct a network congestion game $\mathscr{G}$ with flow value 1 and network with source s and target t as presented in Figure 2(a). The latency function on each edge is presented in the graph. (Some of the edges have a constant latency of either 0 or 1, and we just write the constant near the appropriate edge). δ is a parameter such that $1 > \delta > 0$.

Figure 2(b) presents the path that the malicious flow of value ϵ takes. As for this path the malicious flow goes on every edge with non-constant latency, it is clear that this flow is always a malicious best response, independent of the rational flow. This implies that there is a unique Nash delay in the induced game.

Figure 2(c) presents the rational flow. Near each edge we denote the value of flow on the edge. It is easy to verify that the Nash flow is defined by the following flows on the edges: $f_1 = 1 - \epsilon\frac{1+2\delta}{1+\delta}$, $f_2 = \epsilon\frac{\delta}{1+\delta}$, $f_3 = \frac{1}{2} - \epsilon\frac{1+3\delta}{2(1+\delta)}$.

These flow values correspond to a flow of value f_2 on the paths (s, u, t) and (s, d, t), and flow of value f_3 on the paths (s, u, n_1, m_1, d, t) and (s, u, n_2, m_2, d, t).

For $\epsilon = 0$ (no malicious flow) the rational delay is $2 - \delta$. For $\epsilon > 0$ the rational delay is $2 - \delta - \epsilon\frac{\delta(1-\delta)}{1+\delta}$. Thus, the price of malice for this network is

$$\mathsf{POM}(\mathscr{G}) = \lim_{\varepsilon\to 0} \frac{D(\mathscr{G},\varepsilon) - D(\mathscr{G})}{\varepsilon D(\mathscr{G})} = -\frac{\delta(1-\delta)}{1+\delta}$$

which is a negative constant for any δ such that $1 > \delta > 0$. □

Next we show that the above example can be generalized, and the windfall of malice grows at least linearly with the size of the graph.

Claim 2. *For any $m \geq 1$ there exists a network congestion game with $O(m)$ edges for which the price of malice is $-\frac{m^2}{2(m+2)}$ (i.e., a windfall of malice of order m).*

Proof. For a given m we construct a network congestion game based on the m-th Braess graph [12] which generalizes Braess's paradox, as defined below. The network for $m = 4$ is illustrated in Figure 3(a). The m-th network has $4m + 4$ nodes and a flow value of m. The set of nodes is $\{s, t, v_1, \ldots, v_{m+1}, w_1, \ldots, w_{m+1}, p_1, \ldots, p_m, q_1, \ldots, q_m\}$. For $i \in \{0, \ldots, m\}$ the graph has the following edges. Edges (s, v_{m+1-i}) and (w_{i+1}, t) both have latency function $D_i(X) =$

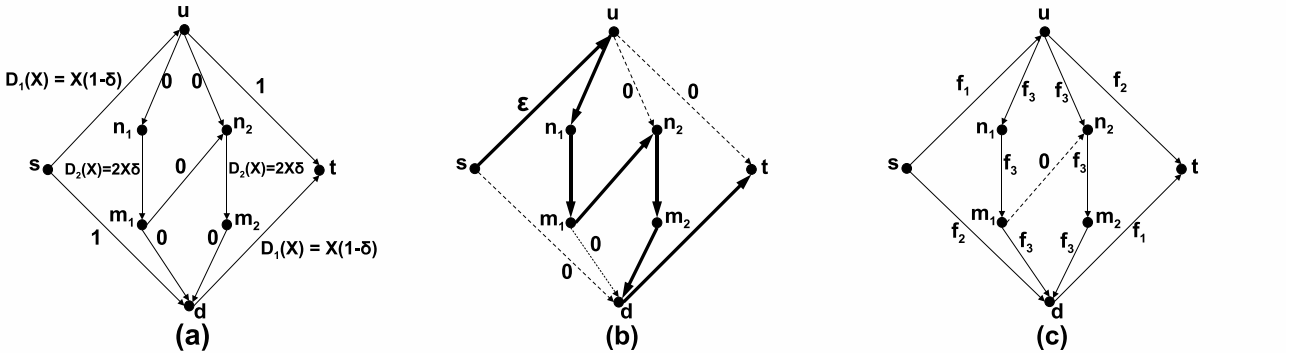


Figure 2: A network congestion game with a negative price of malice. (a) The congestion game. (b) The malicious player's equilibrium strategy. (c) The rational players' equilibrium strategy.

$X \cdot i/2$. Edges (v_{i+1}, w_{i+1}) and (w_{i+1}, v_{i+1}) both have constant latency of 1.[1]

For $i \in \{1, \ldots, m\}$ node v_i is connected to w_{i+1} by a "Z gadget": (v_i, q_i), (p_i, q_i) and (p_i, w_{i+1}) have constant latency 0, and edges (v_i, p_i) and (q_i, w_{i+1}) have a linear latency function X.

The equilibrium flow is as follows. The malicious flow is of size $m\epsilon$ and it travels the path $(s, v_1, p_1, q_1, w_2, \ldots, v_i, p_i, q_i, w_{i+1}, \ldots, v_m, p_m, q_m, w_{m+1}, t)$, see Figure 3(b) for illustration. The rational equilibrium flow is illustrated in Figure 3(c). There is a flow of f_2 on (s, v_1) and on (w_{m+1}, t). There is a flow of $f_2 - f_3$ on $(v_1, w_1), (w_1, t)$, (s, v_{m+1}) and on (v_{m+1}, w_{m+1}). For $i \in \{2, \ldots, m\}$ there is a flow of f_1 on (s, v_i) and on (w_i, t), and a flow of $f_1 - f_3$ on (v_i, w_i). Finally, there is a flow of f_3 on each of the Z gadgets; this flow equally splits on two paths as follows. For $i \in \{1, \ldots, m\}$ there is a flow of $f_3/2$ on $(v_i, q_i), (q_i, w_{i+1}), (v_i, p_i)$ and (p_i, w_{i+1}).

Let $f_1 = 1 - \epsilon\frac{m}{m+2}, f_2 = f_1 - m\epsilon = 1 - \epsilon m(\frac{1}{m+2} + 1)$ and $f_3 = 1 - \epsilon m(2 - \frac{1}{m+2})$. Observe that the flow value (out of s and into t) is $(m-1) \cdot f_1 + f_2 + (f_2 - f_3) + \epsilon m = m$.

We next prove that this is indeed an equilibrium flow. We first observe that rational flow has two types of paths, both with the same delay. The first is a path (s, v_i, w_i, t) for some $i \in \{1, \ldots, m+1\}$, the second is a path $(s, v_i, p_i, w_{i+1}, t)$ (or $(s, v_i, q_i, w_{i+1}, t)$), for some $i \in \{1, \ldots, m\}$. For $i \in \{2, \ldots, m\}$, the delay on the first type is $D_{m+1-i}(f_1) + 1 + D_{i-1}(f_1) = D_m(f_1) + 1 = f_1 \cdot \frac{m}{2} + 1 = \frac{m}{2} + 1 - \epsilon\frac{m^2}{2(m+2)}$. Note that there is a malicious flow of ϵm on (s, v_1) and (w_{m+1}, t) thus the paths (s, v_1, w_1, t) and (s, v_{m+1}, w_{m+1}, t) have delay $D_m(f_2 + \epsilon m) + 1 + 0 = D_m(f_1) + 1$ which is the same delay.

Next we consider paths of the second type, of the form $(s, v_i, p_i, w_{i+1}, t)$ (or $(s, v_i, q_i, w_{i+1}, t)$), for some $i \in \{1, \ldots, m\}$. The total flow on (w_{i+1}, t) is always f_1, even for $i = m$ as $f_2 = f_1 - \epsilon m$ and there is ϵm malicious flow on the edge. The delay on such a path is $D_{m+1-i}(f_1) + (\frac{f_3}{2} + \epsilon m) + D_i(f_1) = D_{m+1}(f_1) + \frac{f_3}{2} + \epsilon m = f_1 \cdot \frac{m+1}{2} + \frac{f_3}{2} + \epsilon m = f_1 \cdot \frac{m}{2} + \frac{f_3 + f_1 + 2\epsilon m}{2} = f_1 \cdot \frac{m}{2} + 1$.

Finally we prove that the malicious flow is indeed playing a best response to the rational flow. We look at the Nash flow with $f_1 = f_2 = f_3 = 1$ and 0 malicious flow. By Proposition 2 it is sufficient to show that the path taken by the malicious flow (as presented in Figure 3(b)) is the unique acyclic[2] path P which maximizes the expresion $q(P) = \sum_{e \in P} x_e(f_N)\ell'_e(x_e(f_N))$.

Let $P_1 = (s, v_1, p_1, q_1, w_2, t), P'_1 = (s, v_1, w_1, t)$, $P_2 = (s, v_m, p_m, q_m, w_{m+1}, t)$ and $P'_2 = (s, v_{m+1}, w_{m+1}, t)$. As $q(P_1) > q(P'_1)$ and $q(P_2) > q(P'_2)$ we do not need to consider the paths P'_1 and P'_2. As on each of the Z gadgets the malicious flow can travel all edges with non constant delay, it is clear that the malicious flow best response must be a path of the form $P_{j,r} = (s, v_j, p_j, q_j, w_{j+1}, v_{j+1}, p_{j+1}, q_{j+1}, w_{j+2}, \ldots, v_r, p_r, q_r, w_{r+1}, t)$ for some $j \geq 1$ and some r with $m \geq r \geq j$. It holds that $q(P_{j,r}) = (m+1-j)/2 + (r-j+1) + r/2 = 3(r-j+1)/2$ and this is an increasing function of r and decreasing function of j thus it is maximized at $r = m$ and $j = 1$.

□

4. THE EXISTENCE OF EQUILIBRIA

Congestion games *without* malicious players have pure Nash equilibria because they are potential games: the potential function Φ defined in Definition 5 decreases whenever a player shifts from one path to another one with lower delay, hence any flow which minimizes Φ must be a pure Nash equilibrium of the congestion game. But congestion games *with* malicious players are not potential games, and as such there is no guarantee that they will have pure Nash equilibria. In fact, a simple example illustrates that a pure Nash equilibrium may not exist even for network congestion games played on a pair of parallel links.

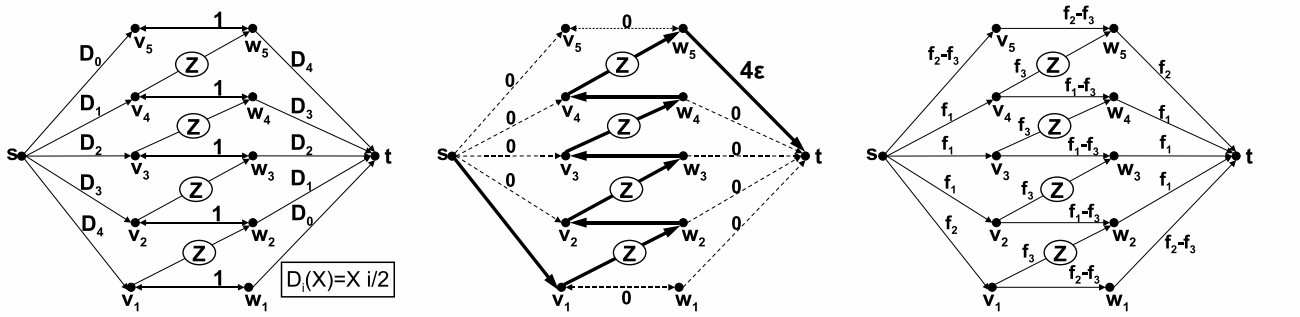
EXAMPLE 1. *Consider a network congestion game in a graph consisting of a source and sink joined by two parallel edges e, e' whose latency functions are $\ell_e(x) = \ell_{e'}(x) = x^2$. Let $v = 2$ and $w = 1$, so that the rational players control 1 unit of flow and the malicious player also controls 1 unit of flow. We claim that this game has no pure Nash equilibrium. To prove it, assume by contradiction that (f, g) is a pair of flows constituting a Nash equilibrium. Let $a = f(\{e\}), b = g(\{e\})$. Then $f(\{e'\}) = 1 - a$ and $g(\{e'\}) = 1 - b$, and*

$$C^g(f) = a(a+b)^2 + (1-a)(2-a-b)^2. \qquad (4)$$

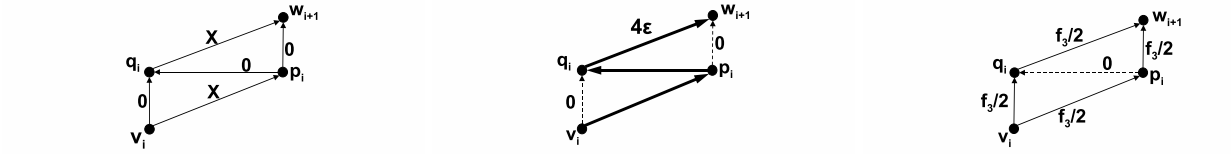
One consequence of (4) is that $C^g(f)$ is a strictly convex function of the parameter b, so its maximum is achieved

[1]This creates a graph with cycles, but the cycles play no role in the construction. One can easily modify the graph to create an acyclic graph with the same properties, see Appendix A.

[2]Although there are directed cycles in the graph (from v_i to w_i and back to v_i), the latency function on any such cycle is constant, thus there is a best response in which flow does not travel in cycles.



(a.1) The network (b.1) Malicious flow on the network (c.1) Rational flow on the network



(a.2) The Z gadget (b.2) Malcious flow on the Z gadget (c.2) Rational flow on the Z gadget

Figure 3: A network congestion game whose price of malice is negative and scales linearly with the network size, demonstrated for $m = 4$. (a) The congestion game. (b) The malicious player's equilibrium strategy. (c) The rational players' equilibrium strategy.

when $b = 0$ or $b = 1$ (or both) but not when $0 < b < 1$. Since we are assuming g is a malicious best response to f, it must be the case that $b = 0$ or $b = 1$. Assume without loss of generality that $b = 0$. Then the induced game $\mathscr{G}^g$ has latency functions $\ell_e^g(x) = x^2$, $\ell_{e'}^g(x) = (1+x)^2$. Since we are assuming f is a Nash flow for $\mathscr{G}^g$, we find that $a = 1$. But then the malicious best response to f is $b = 1$, contradicting our earlier assumption that $b = 0$.

At an intuitive level, the reason why the game constructed in this example has no pure Nash equilibrium is similar to the reason why there is no pure Nash equilibrium in the game "matching pennies". The strict convexity of the latency functions gives the malicious player an incentive to make the load on e, e' as unbalanced as possible, while the rational players have an incentive to make the load on e, e' as balanced as possible; no distribution of flow can simultaneously satisfies the objectives of both types of players.

In light of Example 1, we devote the rest of this section to proving two theorems: first, congestion games with malicious players have pure Nash equilibria as long as the latency functions are continuous and *weakly concave*[3]; second, congestion games with malicious players always have equilibria in the sense of Definition 9.

THEOREM 3. *If $\mathscr{G}$ is a congestion game with a malicious player, and for every edge e, ℓ_e is a continuous, weakly concave function, then there exists a pure equilibrium of $\mathscr{G}$.*

Given Theorem 1, which ensures that our definition of equilibrium is equivalent to the Karakostas-Viglas definition in the case of concave latency functions, it is possible to deduce this theorem from Theorem 1 of [6]. (Actually, our Theorem 3 makes slightly weaker hypotheses about the latency functions, but the proof technique used in [6] implies our theorem without much difficulty.) In the interest of making this paper self-contained, we present a simple alternative proof below.

[3]In particular, as a special case, pure Nash equilibria always exist when the latency functions are linear.

PROOF. Let $w = w(\mathscr{G})$. For a flow $g \in F(\mathscr{G}, w)$, let Φ^g denote the potential function of the game $\mathscr{G}^g$. Recall from Proposition 1 that a flow f is a Nash flow of $\mathscr{G}^g$ if and only if f is a minimizer of Φ^g. Thus a pair $(f, g) \in F(\mathscr{G}, v - w) \times F(\mathscr{G}, w)$ is a pure equilibrium of $\mathscr{G}$ if and only if f is a minimizer of $\Phi^g(f)$ and g is a maximizer of $C^g(f)$. In other words, a pure equilibrium of $\mathscr{G}$ is equivalent to a pure equilibrium of the two-player normal form in which the strategy sets of the two players are $F(\mathscr{G}, v - w)$ and $F(\mathscr{G}, w)$, respectively, and their payoff functions are $-\Phi^g(f)$ and $C^g(f)$, respectively.

Now let us recall the following easy consequence of Kakutani's Fixed Point Theorem. (See, for example, Proposition 20.3 of [9].)

PROPOSITION 3. *A normal form game with finitely many players has a pure Nash equilibrium provided that*

- *Each player's strategy set is a nonempty compact convex subset of a Euclidean space.*
- *For each i, the payoff function of player i is continuous and is a weakly concave function of player i's strategy.*

The first condition is satisfied because the sets $F(\mathscr{G}, v - w), F(\mathscr{G}, w)$ are nonempty convex polytopes. To verify the second condition, first recall that Φ^g is a continuous and

weakly convex function, so $-\Phi^g$ is continuous and weakly concave. Finally, recall that

$$C^g(f) = \sum_{e \in E} x_e(f)\ell_e(x_e(f) + x_e(g)).$$

The function $x_e(g)$ is a linear function of g, and ℓ_e is continuous and weakly concave, so $\ell_e(x_e(f)+x_e(g))$ is a continuous, weakly concave function of g. For fixed f, $C^g(f)$ is a nonnegative linear combination of such functions, so it is also continuous and weakly concave, as desired. □

PROPOSITION 4. *If $\mathscr{G}$ is a congestion game with a malicious player, and all the latency functions ℓ_e are strictly increasing, then $\mathscr{G}$ has an equilibrium.*

PROOF. We use the same two-player game $\widetilde{\mathscr{G}}$ introduced in the proof of Theorem 3. The strategy sets $F(\mathscr{G}, v-w)$ and $F(\mathscr{G}, w)$ are compact Hausdorff topological spaces, and the payoff functions $-\Phi^g(f)$ and $C^g(f)$ are continuous, so the existence theorem for mixed Nash equilibria of games with compact Hausdorff strategy sets [5] ensures that there exist Borel probability measures β_0, γ_0 on $F(\mathscr{G}, v-w)$ and $F(w)$, respectively, such that

$$\beta_0 \in \arg\min_{\beta} \Phi^{\gamma_0}(\beta) \quad (5)$$

$$\gamma_0 \in \arg\max_{\gamma} C^{\gamma}(\beta_0). \quad (6)$$

(Here $\Phi^\gamma(\beta)$ and $C^\gamma(\beta)$ denote the expected values of $\Phi^g(f)$ and $C^g(f)$ when f, g are sampled independently at random from distributions β, γ, respectively.) The only reason that (β_0, γ_0) may not constitute an equilibrium of $\mathscr{G}$ is that our definition of equilibrium requires the rational players to use a pure strategy, not a mixed strategy. In other words, we require the distribution β_0 to be a point mass concentrated at a single flow $f_0 \in F(\mathscr{G}, v-w)$.

Let f_0 denote the flow $f_0(P) = \mathbf{E}_{f \leftarrow \beta_0}[f(P)]$. Our assumption that the latency functions are strictly increasing implies that the function Φ^{γ_0} is strictly convex, so by Jensen's inequality,

$$\Phi^{\gamma_0}(f_0) \le \Phi^{\gamma_0}(\beta_0), \quad (7)$$

with equality if and only if the distribution β_0 is a point mass concentrated at f_0. The left and right sides of (7) are in fact equal, by (5). Consequently β_0 is a point mass concentrated at f_0. By (5) and (6), we may now conclude that (f_0, γ_0) is an equilibrium of $\mathscr{G}$. □

THEOREM 4. *Every congestion game with a malicious player has an equilibrium.*

PROOF. We have seen that the theorem holds when the latency functions are strictly increasing, so the idea of the proof is to approximate an arbitrary congestion game $\mathscr{G} = (E, \vec{\ell}, \Pi, v)$ by games with strictly increasing latency functions. For every positive integer n, let $\ell_e^{(n)}$ denote the latency function $\ell_e^{(n)}(x) = \ell_e(x) + x/n$, and let $\mathscr{G}^{(n)}$ denote the congestion game $(E, \vec{\ell}^{(n)}, \Pi, v)$. Proposition 4 ensures the existence of an equilibrium (f_n, γ_n) for $\mathscr{G}^{(n)}$. We next argue that this sequence of equilibria has a convergent subsequence, under a suitable definition of convergence.

For a separable compact metric space X, we may topologize the set $\Delta(X)$ of Borel probability measures on X using the *weak topology*, in which a sequence $\mu_1, \mu_2, \ldots$ converges to a probability measure μ if and only if $\int f\, d\mu_n \to \int f\, d\mu$ for every bounded continuous function f on X. The space $\Delta(X)$ is compact in the weak topology by Prokhorov's Theorem [1]. Since both $F(\mathscr{G}, v-w)$ and $F(\mathscr{G}, w)$ are separable compact metric spaces, we conclude that the space $F(\mathscr{G}, v-w) \times \Delta(F(\mathscr{G}, w))$ is compact and therefore the sequence (f_n, γ_n) has a convergent subsequence. Replacing the sequence $\mathscr{G}^{(1)}, \mathscr{G}^{(2)}, \ldots$ with a proper subsequence if necessary, we may assume from now on that we have a sequence of games $\mathscr{G}^{(n)} = (E, \vec{\ell}^{(n)}, \Pi, v)$ with equilibria (f_n, γ_n) such that $\ell_e^{(n)}(x) = \ell_e(x) + \alpha_n x$ for some sequence of constants $\alpha_1, \alpha_2, \ldots$ converging to zero, and such that the sequence $(f_1, \gamma_1), (f_2, \gamma_2), \ldots$ converges to a point $(f, \gamma) \in F(\mathscr{G}, v-w) \times \Delta(F(\mathscr{G}, w))$. We must now prove that (f, γ) is an equilibrium of $\mathscr{G}$.

It turns out that

$$\Phi^{\gamma_n}(f_n) \to \Phi^{\gamma}(f) \quad (8)$$

$$C^{\gamma_n}(f_n) \to C^{\gamma}(f). \quad (9)$$

The proofs of these two facts are omitted for space reasons. Assuming them for now, consider any $f' \in F(\mathscr{G}, v-w)$ and $\gamma' \in \Delta(F(\mathscr{G}, w))$. The function $g \mapsto \Phi^g(f')$ is a bounded continuous function of $g \in F(\mathscr{G}, w)$; by the definition of weak convergence this implies $\Phi^{\gamma_n}(f') \to \Phi^{\gamma}(f')$. Combining this with (8) we obtain

$$\Phi^{\gamma}(f) - \Phi^{\gamma}(f') = \lim_{n\to\infty} \left(\Phi^{\gamma_n}(f_n) - \Phi^{\gamma_n}(f')\right) \le 0,$$

hence f is a best response to γ. The functions $g \mapsto C^g(f_n)$, for $n = 1, 2, \ldots$, are uniformly bounded measurable functions of $g \in F(\mathscr{G}, w)$, and $\lim_{n\to\infty} C^g(f_n) = C^g(f)$ for all g. By Lebesgue's dominated convergence theorem, $C^{\gamma'}(f_n) \to C^{\gamma'}(f)$. Combining this with (9) we obtain

$$C^{\gamma}(f) - C^{\gamma'}(f) = \lim_{n\to\infty} \left(C^{\gamma_n}(f_n) - C^{\gamma'}(f_n)\right) \ge 0,$$

hence γ is a best response to f. Thus (f, γ) is an equilibrium as claimed. □

5. CONCLUSIONS AND OPEN PROBLEMS

This paper raises many more questions than it answers. We believe that our definition of malice can be productive in many other contexts; but even if one focuses on congestion games, as we did, there are many open problems to consider.

- We have only derived lower bounds on the price of malice, as well as on its windfall. We conjecture an $O(\ell d)$ upper bound for both, where ℓ is the length of the longest path in the network, and d is an upper bound on the relative slope of the delay functions. (See Definition 14.)

- Given that the presence of malicious players can affect networks in totally different ways, ranging, as we have seen, from disastrous to beneficial, it becomes imperative to understand the circumstances under which these conditions prevail. That is, we are interested in the *characterization* problem of networks for which, say, there is a positive windfall of malice; similarly for a positive price of malice.

  This seems a tall order; even the following simple problem of this sort is currently open: Is there a windfall of malice (of any order of magnitude, not necessarily

first order) in the case of a network consisting of parallel edges (with arbitrary delays)? We conjecture that there isn't, but we have been unable to prove it yet. More generally, we conjecture that there is no windfall of malice in abstract congestion networks in which the set of paths is a *matroid*. Another possible relevant condition for the absence of a windfall might be the absence of some sort of *generalized Braess' paradox*. It would be interesting to explore the extent to which these important properties of networks imply one another.

- In the same spirit as the characterization problem, it would be equally interesting to be able to determine algorithmically the price of malice for individual networks. This brings up the following suite of problems: Given a network with a fraction of malicious flow, find a semi-pure Nash equilibrium, as guaranteed by Theorem 4. Or, given such a network with weakly concave (or even linear) delays, find a pure Nash equilibrium (Theorem 3). Are these problems PPAD-complete (our proof establishes that they are in the class PPAD [10]), or is there an alternative algorithmic way of establishing existence? Since uniqueness of equilibria is no longer guaranteed, perhaps the most useful problem is to find an equlibrium with the largest (or smallest) possible price of malice; this problem may even be NP-complete.

6. ACKNOWLEDGEMENTS

We thank Tim Roughgarden for helpful conversations about this work. We thank Nicolás E. Stier-Moses for pointing us to [6].

7. REFERENCES

[1] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley, 1999.

[2] Felix Brandt, Tuomas Sandholm, and Yoav Shoham. Spiteful bidding in sealed-bid auctions. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[3] Kfir Eliaz. Fault tolerant implementation. *Review of Economic Studies*, 69(3):589–610, July 2002.

[4] S. Fischer, H. Räcke, and B. Vöcking. Fast convergence to wardrop equilibria by adaptive sampling methods. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC 2006)*, pages 653–662, 2006.

[5] I. L. Glicksberg. A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points. *Proc. American Math. Soc.*, 3(1):170–174, 1952.

[6] George Karakostas and Anastasios Viglas. Equilibria for networks with malicious users. In Toshihide Ibaraki, Naoki Katoh, and Hirotaka Ono, editors, *ISAAC*, volume 2906 of *Lecture Notes in Computer Science*, pages 696–704. Springer, 2003.

[7] John Morgan, Ken Steiglitz, and George Reis. The spite motive and equilibrium behavior in auctions. *Contributions to Economic Analysis & Policy*, 2(1):1102–1102, 2003.

[8] Thomas Moscibroda, Stefan Schmid, and Roger Wattenhofer. When Selfish Meets Evil: Byzantine Players in a Virus Inoculation Game. In *25th Annual Symposium on Principles of Distributed Computing (PODC), Denver, Colorado, USA*, July 2006.

[9] Martin J. Osborne and Ariel Rubinstein. *A course in game theory*. MIT Press, 1994.

[10] C. H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *J. Comput. Syst. Sci.*, 48(3):498–532, 1994.

[11] R. W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

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

[13] T. Roughgarden and É. Tardos. How bad is selfish routing? *J. ACM*, 49(2):236–259, 2002.

APPENDIX

A. ACYCLIC CONSTRUCTION FOR WINDFALL OF MALICE

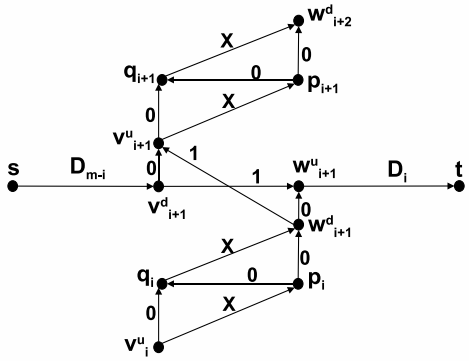


Figure 4: The modification to the network congestion game of Figure 3 which creates an acyclic graph.

In Claim 2 we have presented a construction of a graph that contains directed cycles and has linear windfall of malice. Next we explain how one can modify the graph to create an acyclic graph with the same properties.

The modification is presented in Figure 4. Each node v_i is split to two nodes, v_i^d and v_i^u with a 0 latency edge (v_i^d, v_i^u). Similarly, each node w_i is split to two nodes, w_i^d and w_i^u with a 0 latency edge (w_i^d, w_i^u). The cycle is replaced by edges (v_{i+1}^d, w_{i+1}^u) and (w_{i+1}^d, v_{i+1}^u) both has constant latency of 1. The incoming and outgoing edges of each of the original nodes are split between the two nodes.

The rational and malicious flows in this modified graph are derived from the corresponding flows in the original graph in the obvious manner. The rational flow in the modified graph is uniquely determined by the property that the flow value on each edge (v_i^d, w_i^u) is equal to the flow value on the original edge (v_i, w_i). The malicious flow in the modified graph is uniquely determined by the property that the flow value on each edge (w_i^d, v_i^u) is equal to the flow value on the original edge (w_i, v_i). The proof that these flows still constitute an equilibrium in the modified graph is a straightforward modification of the proof of Claim 2.