Malicious Bayesian Congestion Games*

Martin Gairing

International Computer Science Institute, Berkeley, CA, USA
gairing@icsi.berkeley.edu

Abstract. In this paper, we introduce *malicious Bayesian congestion games* as an extension to congestion games where players might act in a malicious way. In such a game each player has two *types*. Either the player is a rational player seeking to minimize her own delay, or – with a certain probability – the player is *malicious* in which case her only goal is to disturb the other players as much as possible.

We show that such games do in general not possess a Bayesian Nash equilibrium in pure strategies (i.e. a *pure Bayesian Nash equilibrium*). Moreover, given a game, we show that it is NP-complete to decide whether it admits a pure Bayesian Nash equilibrium. This result even holds when resource latency functions are linear, each player is malicious with the same probability, and all strategy sets consist of singleton sets of resources. For a slightly more restricted class of malicious Bayesian congestion games, we provide easy checkable properties that are necessary and sufficient for the existence of a pure Bayesian Nash equilibrium.

In the second part of the paper we study the impact of the malicious types on the overall performance of the system (i.e. the *social cost*). To measure this impact, we use the *Price of Malice*. We provide (tight) bounds on the Price of Malice for an interesting class of malicious Bayesian congestion games. Moreover, we show that for certain congestion games the advent of malicious types can also be beneficial to the system in the sense that the social cost of the worst case equilibrium decreases. We provide a tight bound on the maximum factor by which this happens.

1 Introduction

Motivation and Framework. Over the last decade, the study of strategic behavior in distributed systems has improved our understanding of modern computer artifacts such as the Internet. Normally, the users of such distributed systems are modeled as rational, utility optimizing players. However, in many real world scenarios, users do not necessarily act rationally, but rather *irrationally*. In this paper, we address one form of irrationality, namely, we allow that players act in a *malicious* way. In this case, the only goal of a malicious player is to disturb the (non-malicious) players as much as possible. The presence of *Denial of Service attacks* in the Internet is an example showing that such systems

* This work was supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD).

E. Bampis and M. Skutella (Eds.): WAOA 2008, LNCS 5426, pp. 119–132, 2009.

are quite realistic. In many such systems with malicious players, the players have only *incomplete information* about the set of malicious players. A standard approach for modeling games with incomplete information uses the Harsanyi transformation [14], which converts a game with incomplete information to a game where players have different *types*. The type of a player represents its private information that is not common knowledge to all players. In the resulting *Bayesian game*, each player's uncertainty about each other's type is described by a probability distribution.

One aspect of Game Theory that was studied extensively in recent years is the *Price of Anarchy* as introduced by Koutsoupias and Papadimitriou [16]. The Price of Anarchy is the worst case ratio between the value of the *social cost* in an equilibrium state of the system and that of some social optimum. Usually, the equilibrium state is defined as *Nash equilibrium* – a state in which no player can unilaterally improve her private objective function, also coined as *private cost*. A Nash equilibrium is *pure* if all players choose a pure strategy and *mixed* if players choose probability distributions over pure strategies.

While the celebrated result of Nash [20] guarantees the existence of a mixed Nash equilibrium for every finite game, pure Nash equilibria are not guaranteed to exist (see e.g. [9,12,17,18]). A natural question to ask, is whether a given game possesses a pure Nash equilibrium or not. We address this question by asking about the complexity of this decision problem.

A class of games that always possess pure Nash equilibria is the class of congestion games as introduced by Rosenthal [21]. Here, the strategy set of each player is a subset of the power set of given resources, the latency on each resource is described by a latency function in the number of players sharing this resource, and the private cost of each player is the sum of the latencies of its chosen resources. Milchtaich [18] considered weighted congestion games as an extension to congestion games in which the players have weights and thus different influence on the latency of the resources.

To measure the influence of malicious behavior, Moscibroda et al. [19] introduced the *Price of Byzantine Anarchy* as the worst case ratio between the social cost in an equilibrium state of the system under some assumption on the malicious players and the social cost of some social optimum without malicious players. They further define the *Price of Malice* as the ratio between the Price of Byzantine Anarchy and the Price of Anarchy. We will use a similar definition and define the equilibrium state as a Bayesian Nash equilibrium.

Contribution. In this paper, we introduce *malicious Bayesian congestion games* as an extension to congestion games where players might act in a malicious way. Following Harsanyi's transformation [14], we allow each player to be of two *types*. Either the player is a rational player seeking to minimize her own delay, or – with a certain probability – the player is *malicious* in which case her only goal is to disturb the other players as much as possible. For such games we study the complexity of deciding whether a given game has a pure Bayesian Nash equilibrium. Moreover, we study the impact of the malicious types on the

overall performance of the system (i.e. the *social cost*). To measure this impact, we use the *Price of Malice*, which we define similarly as Moscibroda et al. [19].

As our main result, we show that it is NP-complete to decide whether a given malicious Bayesian congestion game admits a pure Bayesian Nash equilibrium even if resource latency functions are linear and all strategy sets are singleton sets. We show that this result holds already for the very restricted case that each player is malicious with the same probability, and for the case that only one player is malicious with positive probability (Theorem 1). The same result even holds if we further restrict to the case that each player has at most four strategies and at most three players can be assigned to each resource (Theorem 2). For *symmetric* Bayesian congestion games with identical type probability, identical latency functions and strategy sets that consist only of singletons, we provide easy checkable properties that are necessary and sufficient for the existence of a pure Bayesian Nash equilibrium (Theorem 3).

We then shift gears and present results related to the Price of Malice. For general malicious Bayesian congestion games with linear latency functions, we show an upper bound on the Price of Byzantine Anarchy (Theorem 4). Moreover, we prove a lower bound on the same ratio that already holds for the case of identical type probabilities (Theorem 5). As a corollary, we get an asymptotic tight bound on the Price of Malice (Corollary 2). We close the paper with a tight lower bound on the maximum factor by which the social cost of a worst case (Bayesian) Nash equilibrium of a congestion game might decrease by introducing malicious types (Theorem 6).

Related Work. Congestion games and variants thereof have long been used to model non-cooperative resource sharing among selfish players. Rosenthal [21] showed that congestion games always possess pure Nash equilibria. The complexity of computing such a pure Nash equilibrium has been settled for arbitrary latency functions by Fabrikant et al. [8] and later for linear latency functions by Ackermann et al. [1]. On the other hand, for weighted congestion games, Libman and Orda [17], Fotakis et al. [9] and Goemans et al. [12] provide examples that do not allow for a pure Nash equilibrium. Dunkel and Schulz [7] showed that it is NP-complete to decide the existence of a pure Nash equilibrium for a given weighted congestion game.

The Price of Anarchy for weighted congestion games has been studied extensively (see e.g. [3,2,5]). In case of linear latency functions, the Price of Anarchy is exactly $\frac{5}{2}$ for unweighted congestion games [5] and $1 + \Phi$ for weighted congestion games [3], where $\Phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio. The exact value of the Price of Anarchy is also known for the case of polynomial latency functions [2]. For bounds on the Price of Anarchy of (weighted) congestion games with each strategy set being a singleton set of resources, we refer to [11] and references therein.

Several recent papers considered games allowing for malicious player behavior [4,15,19]. Moscibroda et al. [19] introduced the Price of Malice and gave bounds on the Price of Malice for a virus inoculation game where some of the players are malicious. In fact, our definition of Price of Malice is motivated by the

corresponding definition from this paper. Karakostas et al. [15] and Babaioff et al. [4], study malicious player behavior in *non-atomic* congestion games. Here, each player from a continuum of infinitely many players controls only an infinitesimally small amount of weight and a fraction of those players is malicious. In contrast to those papers, our games are atomic, and thus have only finitely many players. This yields different results.

For general Bayesian games, questions concerning the complexity of deciding the existence of a pure Bayesian Nash equilibrium have been addressed in two recent works [6,13]. On the one hand, if the game is given in *standard normal form*, i.e. the utility functions and the type probability distribution are represented extensively as tables, then deciding the existence of a pure Bayesian Nash equilibrium is NP-complete [6]. On the other hand, if both – the utility functions and the type probability distribution – are succinctly encoded, then the problem becomes PP-complete [13]. In contrast to [6], malicious Bayesian congestion games are succinctly represented but they are more structured as the games considered by Gottlob et al. [13].

A certain class of Bayesian congestion game has been introduced in [10]. Here, players act completely rationally but they are uncertain about each other's weight. Among other results, the authors show that such games always possess pure Bayesian Nash equilibria if latency functions are linear.

Roadmap. The rest of the paper is organized as follows. In Section 2, we introduce malicious Bayesian congestion games. In Section 3, we present our results on the complexity of deciding for pure Bayesian Nash equilibria, while Section 4 comprises our findings related to the Price of Malice.

2 Model

2.1 Congestion Games

Instance. A *congestion game* Γ is a tuple $\Gamma = (\mathcal{N}, E, (S_u)_{u\in\mathcal{N}}, (f_e)_{e\in E})$. Here, $\mathcal{N}$ is the set of *players* and E is the finite set of *resources*. Throughout, we denote $n = |\mathcal{N}|$ and $r = |E|$ and assume $n \geq 2$ and $r \geq 2$. For every player $u \in \mathcal{N}$, $S_u \subseteq 2^E$ is the *strategy set* of player u. Denote $S = S_1 \times \ldots \times S_n$. For every resource $e \in E$, the *latency function* $f_e : \mathbb{N} \to \mathbb{R}$ is a non-negative, non-decreasing function that describes the *latency* on resource e. For most of our results, we consider *affine latency functions* with non-negative coefficients, that is, for all resources $e \in E$, the latency function is of the form $f_e(\delta) = a_e \cdot \delta + b_e$ with $a_e, b_e \geq 0$. Affine latency functions are *linear* if $b_e = 0$ for all $e \in E$. A congestion game is called *symmetric*, if $S_u = S'_u$ for any pair of players u, u'.

Strategies and Strategy Profiles. A *pure strategy* for player u is some specific strategy $s_u \in S_u$, while a *mixed strategy* $Q_u = (q(u, s_u))_{s_u\in S_u}$ is a probability distribution over S_u, where $q(u, s_u)$ denotes the probability that player u chooses the pure strategy s_u.

A *pure strategy profile* is an n-tuple $\boldsymbol{s} = (s_1, \ldots, s_n)$ whereas a *mixed strategy profile* $\mathbf{Q} = (Q_1, \ldots, Q_n)$ is represented by an n-tuple of mixed strategies. For a

mixed strategy profile $\mathbf{Q}$, denote by $q(\boldsymbol{s}) = \prod_{u \in \mathcal{N}} q(u, s_u)$ the probability that the players choose the pure strategy profile $\boldsymbol{s}$.

Load and Private Cost. For a pure strategy profile $\boldsymbol{s}$, denote by $\delta_e(\boldsymbol{s}) = |\{u \in \mathcal{N} : e \in s_u\}|$ the *load* on resource $e \in [m]$, i.e. the number of players assigned to e. In the same way, for a partial strategy profile $\boldsymbol{s}_{-i}$, denote $\delta_e(\boldsymbol{s}_{-i}) = |\{u \in \mathcal{N} \setminus \{i\} : e \in s_u\}|$ the *load* on resource $e \in [m]$ without player i.

Fix a pure strategy profile $\boldsymbol{s}$. The *private cost* $\mathsf{PC}_u(\boldsymbol{s})$ of player $u \in \mathcal{N}$ is defined by the *latency* of the chosen resources. Thus $\mathsf{PC}_u(\boldsymbol{s}) = \sum_{e \in s_u} f_e(\delta_e(\boldsymbol{s}))$. For a mixed strategy profile $\mathbf{Q}$, the *private cost* of player $u \in \mathcal{N}$ is $\mathsf{PC}_u(\mathbf{Q}) = \sum_{\boldsymbol{s} \in S} q(\boldsymbol{s}) \cdot \mathsf{PC}_u(\boldsymbol{s})$.

Social Cost. Associated with a congestion game Γ and a mixed strategy profile $\mathbf{Q}$ is the *social cost* $\mathsf{SC}(\Gamma, \mathbf{Q})$ as a measure of social welfare. In particular we use the expected average latency. That is,

$$\mathsf{SC}(\Gamma, \mathbf{Q}) = \frac{1}{n} \sum_{u \in \mathcal{N}} \mathsf{PC}_u(\mathbf{Q}) = \frac{1}{n} \sum_{\boldsymbol{s} \in S} q(\boldsymbol{s}) \sum_{e \in E} \delta_e(\boldsymbol{s}) \cdot f_e(\delta_e(\boldsymbol{s})).$$

Observe, that this measure differs from the *total latency* [22] only by the factor n.

The *optimum* associated with a congestion game Γ is the least possible social cost, over all pure strategy profiles $\boldsymbol{s} \in S$. Thus, $\mathsf{OPT}(\Gamma) = \min_{\boldsymbol{s} \in S} \mathsf{SC}(\Gamma, \boldsymbol{s})$.

Nash Equilibria. Given a congestion game and an associated mixed strategy profile $\mathbf{Q}$, player $u \in \mathcal{N}$ is *satisfied* if the player cannot improve its private cost by unilaterally changing its strategy. Otherwise, player u is *unsatisfied*. The mixed strategy profile $\mathbf{Q}$ is a *Nash equilibrium* if and only if all players $u \in \mathcal{N}$ are satisfied, that is, $\mathsf{PC}_u(\mathbf{Q}) \leq \mathsf{PC}_u(\mathbf{Q}_{-u}, s_u)$ for all $u \in \mathcal{N}$ and $s_u \in S_u$.

Depending on the type of strategy profile we distinguish between *pure* and *mixed* Nash equilibria.

Price of Anarchy. Let $\mathcal{G}$ be a class of congestion games. The *Price of Anarchy*, also called *coordination ratio* and denoted by PoA, is the supremum, over all instances $\Gamma \in \mathcal{G}$ and Nash equilibria $\mathbf{Q}$, of the ratio $\frac{\mathsf{SC}(\Gamma, \mathbf{Q})}{\mathsf{OPT}(\Gamma)}$. Thus, $\mathsf{PoA} = \sup_{\Gamma \in \mathcal{G}, \mathbf{Q}} \frac{\mathsf{SC}(\Gamma, \mathbf{Q})}{\mathsf{OPT}(\Gamma)}$.

2.2 Malicious Bayesian Congestion Games

Instance. A *malicious Bayesian congestion game* Ψ is an extension of congestion games, where each player is malicious with a certain probability. Following Harsanyi's approach, we model such a game with incomplete information as a Bayesian game, where each player $u \in \mathcal{N}$ can be of two types: Either u is *selfish* or *malicious*. For each type of player $u \in \mathcal{N}$ we introduce two independent type-agents u^s and u^m, denoting the *selfish* and *malicious type-agent* of player u, respectively.

Let p_u be the probability that player $u \in \mathcal{N}$ is malicious and call p_u the *type probability* of player u. Define the *type probability vector* $\boldsymbol{p} = (p_1, \ldots, p_n)$ in the

natural way. Denote $p_{\min} = \min_{u \in \mathcal{N}} p_u$. In the case of identical type probabilities $p_u = p$ for all player $u \in \mathcal{N}$. Define $\Delta = \sum_{u \in \mathcal{N}} p_u$ as the *expected number of malicious players*. Observe, that for identical type probabilities $\Delta = p \cdot n$. Denote by Γ_Ψ the congestion game that arises from the malicious Bayesian congestion game Ψ by setting $p_u = 0$ for all player $u \in \mathcal{N}$.

Summing up, a malicious Bayesian congestion game Ψ is given by a tuple $\Psi = (\mathcal{N}, E, (S_u)_{u \in \mathcal{N}}, (p_u)_{u \in \mathcal{N}}, (f_e)_{e \in E})$.

Strategies and Strategy Profiles. A pure strategy σ_u for player $u \in \mathcal{N}$ is now a tuple $\sigma_u = (\sigma(u^s), \sigma(u^m)) \in S_u^2$, where $\sigma(u^s)$ and $\sigma(u^m)$ denote the strategy of the selfish type-agent and malicious type-agent of player u, respectively. Denote $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$. A *mixed strategy* Q_i is now a probability distribution over $S_i \times S_i$. Define $\mathbf{Q}$ and $q(\boldsymbol{\sigma})$ as before.

Private Cost. For any type probability vector $\boldsymbol{p}$ and pure strategy profile $\boldsymbol{\sigma}$, denote the *expected selfish load* on resource $e \in E$ by $\delta_e(\boldsymbol{\sigma}) = \sum_{u \in \mathcal{N}: e \in \sigma(u^s)} (1 - p_u)$ and the *expected malicious load* by $\kappa_e(\boldsymbol{\sigma}) = \sum_{u \in \mathcal{N}: e \in \sigma(u^m)} p_u$. For a partial assignment $\boldsymbol{\sigma}_{-u}$ define $\delta_e(\boldsymbol{\sigma}_{-u})$ and $\kappa_e(\boldsymbol{\sigma}_{-u})$ accordingly, by disregarding player u.

Fix any type probability vector $\boldsymbol{p}$ and pure strategy profile $\boldsymbol{\sigma}$. The *private cost* of player $u \in \mathcal{N}$ is defined by $\mathsf{PC}_u(\boldsymbol{p}, \boldsymbol{\sigma}) = \sum_{e \in \sigma(u^s)} f_e\left(\delta_e(\boldsymbol{\sigma}_{-u}) + \kappa_e(\boldsymbol{\sigma}_{-u}) + 1\right)$. In other words $\mathsf{PC}_u(\boldsymbol{p}, \boldsymbol{\sigma})$ is the expected latency that player u experiences if player u is selfish. For each player $u \in \mathcal{N}$, type-agent u^s aims to minimize $\mathsf{PC}_u(\boldsymbol{p}, \boldsymbol{\sigma})$. Observe, that $\mathsf{PC}_u(\boldsymbol{p}, \boldsymbol{\sigma})$ does not depend on $\sigma(u^m)$. For a mixed strategy profile $\mathbf{Q}$, define $\mathsf{PC}_u(\boldsymbol{p}, \mathbf{Q})$ accordingly.

Social Cost. Let Ψ be a malicious Bayesian congestion game with type probability vector $\boldsymbol{p}$ and let $\mathbf{Q}$ be a mixed strategy profile for Ψ. We generalize the definition of *social cost* $\mathsf{SC}(\Psi, \mathbf{Q})$ to the weighted average latency of the selfish type-agents. That is, $\mathsf{SC}(\Psi, \mathbf{Q}) = \frac{\sum_{u \in \mathcal{N}} (1 - p_u) \cdot \mathsf{PC}_u(\boldsymbol{p}, \mathbf{Q})}{n - \Delta}$.

Bayesian Nash equilibria. A selfish type-agent is *satisfied* if she cannot unilaterally decrease her private cost, that is, $\mathsf{PC}_u(\mathbf{Q}) \leq \mathsf{PC}_u(\mathbf{Q}_{-u^s}, \sigma(u^s))$ for all $u \in \mathcal{N}$ and $\sigma(u^s) \in S_u$. In contrast to the selfish type-agents, each malicious type-agent aims to maximize the social cost. So, a malicious type-agent is *satisfied* if she cannot increase the social cost by unilaterally changing her strategy.

For a malicious Bayesian congestion game, a mixed strategy profile $\mathbf{Q}$ is a *Bayesian Nash equilibrium* if and only if both type-agents of all players $u \in \mathcal{N}$ are satisfied. Depending on the type of strategy profile we again differ between *pure* and *mixed* Bayesian Nash equilibria.

Price of Byzantine Anarchy and Price of Malice. For a fixed expected number of malicious players Δ, let $\mathcal{G}(\Delta)$ be the class of malicious Bayesian congestion games where $\sum_{u \in \mathcal{N}} p_u = \Delta$. Similarly to [19], we define the *Price of Byzantine Anarchy*, denoted by PoB, as the supremum, over all instances $\Psi \in \mathcal{G}(\Delta)$ and Bayesian Nash equilibria $\mathbf{Q}$, of the ratio between the social cost in $\mathbf{Q}$ and the optimum social cost of the corresponding congestion game

Γ_Ψ. Thus, $\mathsf{PoB}(\Delta) = \sup_{\Psi \in \mathcal{G}(\Delta), \mathbf{Q}} \frac{\mathsf{SC}(\Psi, \mathbf{Q})}{\mathsf{OPT}(\Gamma_\Psi)}$. Observe that for $\Delta = 0$, the Price of Byzantine Anarchy $\mathsf{PoB}(0)$ reduces to the Price of Anarchy PoA as defined in Section 2.1.

Again similarly to [19], we define the *Price of Malice* by $\mathsf{PoM}(\Delta) = \frac{\mathsf{PoB}(\Delta)}{\mathsf{PoB}(0)}$.

3 Existence and Complexity of Pure Bayesian Nash Equilibria

In this section, we study the complexity of deciding whether a given malicious Bayesian congestion game possesses a pure Bayesian Nash equilibrium or not.

Theorem 1. *The problem of deciding whether a malicious Bayesian congestion game with linear latency functions possesses a pure Bayesian Nash equilibrium is NP-complete, even if all strategy sets consist of singletons and either of the following properties holds:*

(a) All players are malicious with the same probability p for any $0 < p < 1$.
(b) Only one player is malicious with positive probability p for any $0 < p \leq 1$.

Proof. Our proof uses a reduction from a restricted version of 3-SAT. Here, 3-SAT is restricted to instances where each clause is a disjunction of 2 or 3 variables and each variable occurs at most three times. Tovey [23] showed that it is NP-complete to decide the satisfiability of such instances. Consider an arbitrary instance of 3-SAT with set of variables $X = \{x_1, \dots x_\ell\}$ and set of clauses $C = \{c_1, \dots c_k\}$. Without loss of generality, we may assume that each variable occurs at most twice unnegated and at most twice negated.

Part (a): We will construct a malicious Bayesian congestion game with singleton strategy sets and identical type probability p. Our construction imposes one player u_c for each clause $c \in C$, one player u_x and two resources e_x^0, e_x^1 for each variable $x \in X$, 3 additional players u_0, u_1, u_2, and 5 additional resources e_0, e_1, e_2, e_3, e_4. Our construction is summarized in Figure 1. Resources

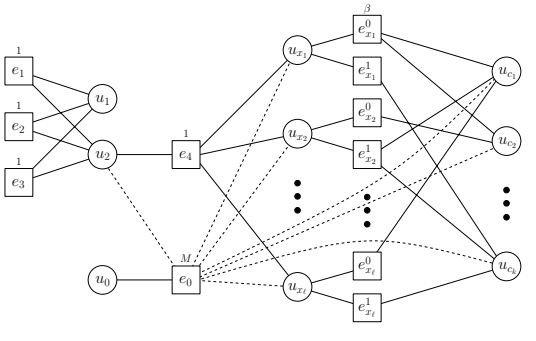

Fig. 1. Construction for the proof of Theorem 1

are depicted as squares and players as circles and an edge (solid or dotted) between a resource e and a player u indicates that $\{e\}$ is in u's strategy set. A number α above a resource e defines the slope of the corresponding linear latency function $f_e(\delta) = \alpha \cdot \delta$. Denote $E_v = \{e^0_{x_1}, e^1_{x_1}, \ldots, e^0_{x_\ell}, e^1_{x_\ell}\}$. For the proof of part (a), let $\beta = 2 - p$. So, all resources $e \in E_v$ share the latency function $f_e(\delta) = (2-p) \cdot \delta$.

Player u_0 can only be assigned to e_0. Both u_0 and e_0 are used to collect the malicious type-agents of all players except player u_1. Thus all those players have e_0 in their strategy set and M is chosen sufficiently large, such that for all those malicious type-agents e_0 is a dominant strategy and no selfish type other than u^s_0 will ever prefer to choose e_0. Choosing $M = \ell + 1$ suffices. Player u_1 and u_2 are connected to e_1, e_2, and e_3, while u_2 can also choose e_0 and e_4. For each variable $x \in X$, the corresponding *variable player* u_x is connected to e_0, e_4, e^0_x and e^1_x. Assigning the selfish type-agent u^s_x to e^0_x (resp. e^1_x) will be interpreted as setting x to true (resp. false). For each clause $c \in C$, the corresponding *clause player* u_c is connected to e_0 and to all resources e^0_x (e^1_x) with $x \in X$ and x appears *negated* (*unnegated*) in c. For the example in Figure 1, $c_1 = (\overline{x}_1 \vee x_2 \vee \overline{x}_\ell)$, $c_2 = (\overline{x}_1 \vee \overline{x}_2)$, and $c_k = (x_1 \vee x_2 \vee x_\ell)$. Observe that by the structure of our 3-SAT instance, no more than two clause players are connected to each resource in E_v. This completes the construction of the malicious Bayesian congestion game.

We will first show that if the 3-SAT instance is satisfiable then the corresponding Bayesian congestion game possesses a pure Bayesian Nash equilibrium. Given a satisfying truth assignment, we define a strategy profile $\boldsymbol{\sigma}$ of the malicious Bayesian congestion game as follows:

- Both type-agents of player u_0 can only be assigned to e_0.
- All malicious type-agents except u^m_1 are assigned to resource e_0. By the choice of M, none of those malicious type-agents can improve.
- Both type-agents of player u_1 are assigned to e_1 and no type-agent of any player is assigned to e_2 or e_3. It is easy to see that neither u^m_1 nor u^s_1 have an incentive to switch.
- Type agent u^s_2 is the only type-agent assigned to e_4. So, u^s_2 cannot improve.
- For each $x \in X$, the selfish type-agent u^s_x of variable player u_x is assigned to resource e^0_x if $x =$ true in the satisfying truth assignment, and to e^1_x otherwise. Each of these selfish type-agents is the only type-agent assigned to her resource. So, they all experience an expected latency of $\beta = 2 - p$ and changing to e_4 would yield the same expected latency. Thus, the selfish type-agents of all variable players are satisfied.
- Denote by E'_v the subset of resources from E_v to which no selfish type-agent of a variable player is assigned. Since we have a satisfying truth assignment, each clause player is connected to some resource from E'_v. For each $c \in C$, the selfish type-agent u^s_c is assigned to some resource in E'_v as follows:

  Consider the sub-game that consists only of the selfish type-agents of the clause players u_c, $c \in C$ and the set of resources E'_v. Observe that this sub-game is a (non-malicious) congestion game and thus admits a pure Nash equilibrium [21]. Assign the selfish type-agents of each clause player

according to this Nash equilibrium. So, none of these selfish type-agents can improve by changing to some other resource in E'_v. Moreover, at most two selfish type-agents are assigned to each resource in E'_v and there is exactly one selfish type-agent of a variable player assigned to each resource in $E_v \setminus E'_v$. Thus, the selfish type-agents of all clause players are satisfied.

Since no type-agent can improve be changing her strategy, it follows that $\boldsymbol{\sigma}$ in a pure Bayesian Nash equilibrium.

For the other direction observe that any pure Bayesian Nash equilibrium $\boldsymbol{\sigma}$ fulfills the following structural properties:

(I) All malicious type-agents except u_1^m are assigned to resource e_0 and u_0^s is the only selfish type-agent assigned to e_0.
(II) The selfish type-agent u_2^s is assigned to e_4 and no other type-agent is assigned to e_4.

Property (I) follows immediately by the choice of M. We will now prove property (II).

By way of contradiction assume that u_2^s is assigned to a resource in $\{e_1, e_2, e_3\}$ in a pure Bayesian Nash equilibrium $\boldsymbol{\sigma}$. In this case u_1^m will always choose the same resource as u_2^s. However, then there must be an empty resource in $\{e_1, e_2, e_3\}$ and u_2^s can improve by choosing this empty resource. This contradicts our assumption that $\boldsymbol{\sigma}$ is a pure Bayesian Nash equilibrium. Thus, u_2^s is assigned to e_4. If some other type-agent is also assigned to e_4, then u_2^s experiences an expected latency of at least $2 - p$ and u_2^s could decrease her expected latency to 1 by switching to the empty resource in $\{e_1, e_2, e_3\}$. Again a contradiction to $\boldsymbol{\sigma}$ being a pure Bayesian Nash equilibrium. It follows that u_2^s is the only type-agent assigned to e_4 in $\boldsymbol{\sigma}$. This completes the proof of property (II).

Since u_2^s is the only type-agent assigned to e_4 it follows that for each variable $x \in X$ the corresponding selfish type-agent u_x^s is either assigned to e_x^0 or to e_x^1. If u_x^s is not the only type-agent on that resource then her expected latency is at least $(2-p)^2$ while changing to e_4 would improve her expected latency to $2 - p$, a contradiction to $\boldsymbol{\sigma}$ being a pure Bayesian Nash equilibrium. It follows that the selfish type-agents of all clause players are only assigned to resources in E_v to which no selfish type-agent of a variable player is assigned. This is only possible if the strategies of the selfish type-agents $u_x^s, x \in X$ correspond to a satisfying truth assignment. This completes the proof of part (a).

Part (b): To see that (b) holds we alter the construction depicted in Figure 1 slightly by deleting player u_0 and resource e_0. Furthermore, in the new construction player u_1 is the only player that is malicious with positive probability p. For the slope of the latency functions of resources in E_v, let $\beta = \frac{3}{2}$ (in fact any $1 < \beta < 2$ would also do). The rest of the construction does not change. The proof now follows the same line of arguments as in part (a) with only minor changes. □

Theorem 2. *The results from Theorem 1 hold, even if $|S_u| \leq 4$ for all players $u \in \mathcal{N}$ and for each resource $e \in E$ there are at most three players $u \in \mathcal{N}$ with $\{e\} \in S_u$.*

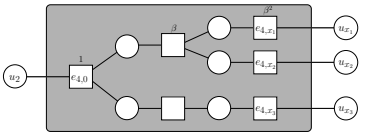

Fig. 2. Tree for $\ell = 3$

Proof (Sketch). We will slightly alter the construction from Figure 1. First observe that we have already $|S_u| \leq 4$ for all players $u \in \mathcal{N}$. Furthermore, the only resources that are in the strategy set of more than three players are e_4 and for part (a) also e_0.

To resolve this for e_4, disconnect all players from e_4 and replace the single resource e_4 with a binary tree of resources with root $e_{4,0}$ that has ℓ leaves $e_{4,x_1}, \dots e_{4,x_\ell}$, all with depth $\lceil \log(\ell) \rceil$. For a resource e at level j the latency function is defined by $f_e(\delta) = \beta^j \cdot \delta$. So $f_{e_{4,0}}(\delta) = 1$ and $f_{e_{4,x}}(\delta) = \beta^{\lceil \log(\ell) \rceil} \cdot \delta$ for all leaves $x \in X$. For each pair of resources from two consecutive levels, we introduce a new player to connect them. Call those players *tree players*. Figure 2 shows the construction for $\ell = 3$. Player u_2 is connected to resource $e_{4,0}$ and each variable player $x \in X$ is connected to $e_{4,x}$. We also change the latency function of all resources $e \in E_v$ (cf. Theorem 1) to $f_e(\delta) = \beta^{\lceil \log(\ell) \rceil} \cdot \delta$.

Moreover, for part (a) we have to resolve that more than three players are connected to e_0. To do so, we simply copy resource e_0 together with player u_0 multiple times and connect all players (including the tree players) except player u_1 to the new set of resources that evolve from e_0. By having sufficiently many copies, this can be done, such that no more than three players are connected to each new resource. Again, M is chosen sufficiently large, e.g. $M = 2^{\lceil \log(\ell) \rceil + 1}$.

Observe that u_2^s will only selfishly choose $e_{4,0}$ if all tree players choose the strategy that is closer to the leaves. The rest of the proof now simply follows the proof of Theorem 1. □

For the more restricted class of symmetric malicious Bayesian congestion game with singleton strategy sets, identical type probability p and identical latency functions we can easily decide whether a pure Bayesian Nash equilibrium exists or not.

Theorem 3. *A symmetric malicious Bayesian congestion game with singleton strategy sets, identical type probability p and identical (not necessarily linear) latency functions possesses a pure Bayesian Nash equilibrium if and only if either (a) $p \leq \frac{1}{2}$ and $r = 2$, or (b) $p \leq \frac{1}{2}$ and $r|n$.*

Observe, that the proof of Theorem 3 is constructive. So, if the requirements for the existence of a pure Bayesian Nash equilibrium are fulfilled, then this equilibrium can also be easily constructed in linear time.

4 Price of Malice

We now shift gears and present our results that are related to the Price of Malice. We start with a general upper bound on the Price of Byzantine Anarchy. The proof of this upper bound uses a technique from [5] adapted to the model of malicious Bayesian congestion games.

Theorem 4. *Consider the class of malicious Bayesian congestion games $\mathcal{G}(\Delta)$ with affine latency functions. Then, $\mathsf{PoB}(\Delta) \leq \frac{n}{n-\Delta}(1-p_{\min})\left(\Delta + \frac{3+\sqrt{5+4\Delta}}{2}\right)$.*

For the case of identical type probabilities we can provide a better upper bound on the Price of Byzantine Anarchy. Observe that for identical type probabilities, $\Delta = p \cdot n$ and $p_{\min} = p$. As an immediate corollary to Theorem 4, we get:

Corollary 1. *Consider the class of malicious Bayesian congestion games $\mathcal{G}(\Delta)$ with affine latency functions and identical type probability. Then, $\mathsf{PoB}(\Delta) \leq \Delta + \frac{3+\sqrt{5+4\Delta}}{2}$.*

We proceed by introducing a malicious Bayesian congestion game that is parameterized by a parameter α. In the remainder of the paper, we will make use of this construction twice, each time with a different parameter α.

Example 1. Given some $\alpha > 0$, construct a malicious Bayesian congestion game $\Psi(\alpha)$ with linear latency functions, $n \geq 3$ players and identical type probability p and $|E| = 2n$ as follows: Let $E = E_1 \cup E_2$ with $E_1 = \{g_1, \ldots, g_n\}$ and $E_2 = \{h_1, \ldots, h_n\}$. Each player $u \in \{1, \ldots n\}$ has three strategies in her strategy set. So, $S_u = \{s_u^1, s_u^2, s_u^3\}$ with $s_u^1 = \{g_u, h_u\}, s_u^2 = \{g_{u+1}, h_{u+1}, h_{u+2}\}$ and $s_u^3 = E_1 \cup E_2$, where $g_j = g_{j-n}$ and $h_j = h_{j-n}$ for $j > n$.

Each resource $e \in E_1$ has a latency function $f_e(\delta) = \alpha \cdot \delta$ whereas the resources $e \in E_2$ share the identity as their latency function, i.e. $f_e(\delta) = \delta$.

We make use of Example 1 to show a lower bound on the Price of Byzantine Anarchy:

Theorem 5. *Consider the class of malicious Bayesian congestion games $\mathcal{G}(\Delta)$ with linear latency functions and identical type probability p. Then, $\mathsf{PoB}(\Delta) \geq \Delta + 2$.*

Proof. Consider the malicious Bayesian congestion game $\Psi = \Psi(\alpha)$ given in Example 1 with $\alpha = \frac{1+(n-1)p}{1-p}$. Observe that $\Delta = n \cdot p$.

Obviously , the optimum allocation $\boldsymbol{s}^*$ for the corresponding non-malicious game Γ_Ψ is for each player $u \in \mathcal{N}$ to choose strategy s_u^1. This yields $\mathsf{SC}(\Gamma_\Psi, \boldsymbol{s}^*) = 1 + \alpha = \frac{2+(n-2)p}{1-p}$.

On the other hand, if $\sigma(u^m) = s_u^3$ and $\sigma(u^s) = s_u^2$ for all player $u \in \mathcal{N}$, then $\boldsymbol{\sigma}$ is a (pure) Bayesian Nash equilibrium for Ψ, with

$$\begin{aligned}\mathsf{SC}(\Psi, \boldsymbol{\sigma}) &= 2(1 + (1-p) + (n-1)p) + (1 + (n-1)p)\alpha \\ &= \frac{2(1-p)(2+(n-2)p) + (1+(n-1)p)^2}{1-p}\end{aligned}$$

It follows that

$$\begin{aligned}\frac{\mathsf{SC}(\Psi,\boldsymbol{\sigma})}{\mathsf{SC}(\Gamma_\Psi,\boldsymbol{s}^*)} &= 2(1-p)+\frac{(1+(n-1)p)^2}{2+(n-2)p}\\ &= 2(1-p)+\frac{1+(n-1)p(2+(n-1)p)}{2+(n-2)p}\\ &> 2-3p+n\cdot p\\ &= \Delta+2-3p.\end{aligned}$$

The theorem follows for $p \to 0$, which implies $n \to \infty$. □

Recall that the Price of Anarchy of (non-malicious) congestion games with affine latency functions is $\frac{5}{2}$ [5]. By combining this with Corollary 1 and Theorem 5 we get:

Corollary 2. *Consider the class of malicious Bayesian congestion games $\mathcal{G}(\Delta)$ with affine latency functions and identical type probability p. Then, $\mathsf{PoM}(\Delta) = \Theta(\Delta)$.*

For certain congestion games, introducing malicious types might also be beneficial to the system, in the sense that the social cost of the worst case equilibrium (one that maximizes social cost) decreases. To capture this, we define the *Windfall of Malice*. The term Windfall of Malice is due to [4]. For a malicious Bayesian congestion game Ψ, denote $\mathsf{WoM}(\Psi)$ as the ratio between the social costs of the worst case Nash equilibrium of the corresponding congestion game Γ_Ψ and the worst case Bayesian Nash equilibrium of Ψ. We show:

Theorem 6. *For each $\epsilon > 0$ there is a malicious Bayesian congestion game Ψ with linear latency functions and identical type probability p, such that $\mathsf{WoM}(\Psi) \geq \frac{5}{2} - \epsilon$.*

Proof (Sketch). Define $\Psi = \Psi(\alpha)$ as in Example 1 with $n = 3$ and $\alpha = 1$. For the congestion game Γ_Ψ that corresponds to Ψ, all players u choosing s_u^2 is a Nash equilibrium $\boldsymbol{s}$ that maximizes social cost and $\mathsf{SC}(\Gamma_\Psi, \boldsymbol{s}) = 5$.

For the malicious congestion game Ψ (where $p > 0$), there is a unique (pure) Bayesian Nash equilibrium $\boldsymbol{\sigma}$ where $\sigma(u^s) = s_u^1$ and $\sigma(u^m) = s_u^3$ for all players $u \in \mathcal{N}$. For its social cost we get $\mathsf{SC}(\Psi, \boldsymbol{\sigma}) = 2 + 4p$.

So, for each $\epsilon > 0$ there is a sufficiently small p, such that

$$\mathsf{WoM}(\Psi) = \frac{\mathsf{SC}(\Gamma_\Psi,\boldsymbol{s})}{\mathsf{SC}(\Psi,\boldsymbol{\sigma})} = \frac{5}{2+4p} \geq \frac{5}{2} - \epsilon.$$

This completes the proof of the theorem. □

We remark that this is actually a tight result, since for the considered class of malicious Bayesian congestion games the Windfall of Malice cannot be larger than the Price of Anarchy of the corresponding class of congestion games which was shown to be $\frac{5}{2}$ in [5].

5 Conclusion and Open Problems

In this paper, we have introduced and studied a new extension to congestion games, that we call malicious Bayesian congestion games. More specifically, we have studied problems concerned with the complexity of deciding the existence of pure Bayesian Nash equilibria. Furthermore, we have presented results on the Price of Malice.

Although we were able to derive multiple interesting results, this work also gives rise to many interesting open problems. We conclude this paper by stating those, that we consider the most prominent ones.

- Our NP-completeness result in Theorem 1 holds even for linear latency functions, identical type probabilities, and if all strategy sets are singleton sets of resources. However, if such games are further restricted to symmetric games and identical linear latency functions, then deciding the existence of a pure Bayesian Nash equilibrium becomes a trivial task. We believe that this task can also be performed in polynomial time for *non-identical* linear latency functions and symmetric strategy sets.
- Although the upper bound in Corollary 1 and the corresponding lower bound in Theorem 5 are asymptotically tight, there is still potential to improve. We conjecture that in this case $\mathsf{PoB}(\Delta) = \Delta + O(1)$.
- We believe that the concept of malicious Bayesian games is very interesting and deserves further study also in other scenarios. We hope, that our work will encourage others to study such malicious Bayesian games.

Acknowledgments

We are very grateful to Christos Papadimitriou and Andreas Maletti for many fruitful discussions on the topic. Moreover, we thank Florian Schoppmann for his helpful comments on an early version of this paper.

References

1. Ackermann, H., Röglin, H., Vöcking, B.: On the Impact of Combinatorial Structure on Congestion Games. In: Proc. of the 47th Annual Symposium on Foundations of Computer Science (FOCS 2006), pp. 613–622 (2006)
2. Aland, S., Dumrauf, D., Gairing, M., Monien, B., Schoppmann, F.: Exact price of anarchy for polynomial congestion games. In: Durand, B., Thomas, W. (eds.) STACS 2006. LNCS, vol. 3884, pp. 218–229. Springer, Heidelberg (2006)
3. Awerbuch, B., Azar, Y., Epstein, A.: The Price of Routing Unsplittable Flow. In: Proc. of the 37th Annual ACM Symposium on Theory of Computing (STOC 2005), pp. 57–66 (2005)
4. Babaioff, M., Kleinberg, R., Papadimitriou, C.H.: Congestion Games with Malicious Players. In: Proc. of the 8th ACM Conference on Electronic Commerce (EC 2007), pp. 103–112 (2007)

5. Christodoulou, G., Koutsoupias, E.: The Price of Anarchy of Finite Congestion Games. In: Proc. of the 37th Annual ACM Symposium on Theory of Computing (STOC 2005), pp. 67–73 (2005)
6. Conitzer, V., Sandholm, T.: New Complexity Results about Nash Equilibria. In: Proc. of 18th International Joint Conference on Artificial Intelligence (IJCAI 2003), pp. 765–771 (2003)
7. Dunkel, J., Schulz, A.S.: On the Complexity of Pure-Strategy Nash Equilibria in Congestion and Local-Effect Games. In: Spirakis, P.G., Mavronicolas, M., Kontogiannis, S.C. (eds.) WINE 2006. LNCS, vol. 4286, pp. 62–73. Springer, Heidelberg (2006)
8. Fabrikant, A., Papadimitriou, C.H., Talwar, K.: The Complexity of Pure Nash Equilibria. In: Proc. of the 36th Annual ACM Symposium on Theory of Computing (STOC 2004), pp. 604–612 (2004)
9. Fotakis, D.A., Kontogiannis, S.C., Spirakis, P.G.: Symmetry in network congestion games: Pure equilibria and anarchy cost. In: Erlebach, T., Persinao, G. (eds.) WAOA 2005. LNCS, vol. 3879, pp. 161–175. Springer, Heidelberg (2006)
10. Gairing, M., Monien, B., Tiemann, K.: Selfish Routing with Incomplete Information. Theory of Computing Systems 42(1), 91–130 (2008)
11. Gairing, M., Schoppmann, F.: Total Latency in Singleton Congestion Games. In: Deng, X., Graham, F.C. (eds.) WINE 2007. LNCS, vol. 4858, pp. 381–387. Springer, Heidelberg (2007)
12. Goemans, M.X., Mirrokni, V., Vetta, A.: Sink Equilibria and Convergence. In: Proc. of the 46th Annual Symposium on Foundations of Computer Science (FOCS 2005), pp. 142–154 (2005)
13. Gottlob, G., Greco, G., Mancini, T.: Complexity of Pure Equilibria in Bayesian Games. In: Proc. of 20th International Joint Conference on Artificial Intelligence (IJCAI 2007), pp. 1294–1299 (2007)
14. Harsanyi, J.C.: Games with Incomplete Information Played by Bayesian Players, I, II, III. Management Science 14, 159–182, 320–332, 468–502 (1967)
15. Karakostas, G., Viglas, A.: Equilibria for networks with malicious users. In: Ibaraki, T., Katoh, N., Ono, H. (eds.) ISAAC 2003. LNCS, vol. 2906, pp. 696–704. Springer, Heidelberg (2003)
16. Koutsoupias, E., Papadimitriou, C.H.: Worst-case equilibria. In: Meinel, C., Tison, S. (eds.) STACS 1999. LNCS, vol. 1563, pp. 404–413. Springer, Heidelberg (1999)
17. Libman, L., Orda, A.: Atomic Resource Sharing in Noncooperative Networks. Telecommunication Systems 17(4), 385–409 (2001)
18. Milchtaich, I.: Congestion Games with Player-Specific Payoff Functions. Games and Economic Behavior 13(1), 111–124 (1996)
19. Moscibroda, T., Schmid, S., Wattenhofer, R.: When Selfish Meets Evil: Byzantine Players in a Virus Inoculation Game. In: Proc. of the 25th Annual ACM Symposium on Principles of Distributed Computing (PODC 2006), pp. 35–44 (2006)
20. Nash, J.F.: Non-Cooperative Games. Annals of Mathematics 54(2), 286–295 (1951)
21. Rosenthal, R.W.: A Class of Games Possessing Pure-Strategy Nash Equilibria. International Journal of Game Theory 2, 65–67 (1973)
22. Roughgarden, T., Tardos, É.: How Bad Is Selfish Routing? Journal of the ACM 49(2), 236–259 (2002)
23. Tovey, C.A.: A Simplified NP-complete Satisfiability Problem. Discrete Applied Mathematics 8, 85–89 (1984)