Adversarial Behavior in Network Mechanism Design

Anil Kumar Chorppath
Technical University of Berlin
Deutsche Telekom Laboratories
10587 Berlin, Germany
anil.chorppath@sec.t-labs.tu-berlin.de

Tansu Alpcan
Technical University of Berlin
Deutsche Telekom Laboratories
10587 Berlin, Germany
alpcan@sec.t-labs.tu-berlin.de

ABSTRACT

This paper studies the effects of and countermeasures against adversarial behavior in network resource allocation mechanisms such as pricing and auctions. It models the heterogeneous behavior of users, which ranges from altruistic to selfish and even to malicious, using game theory. The paper adopts a mechanism design approach to quantify the effect of adversarial behavior and modify the mechanisms to respond. First, the *Price of Malice* of the existing network mechanisms to adversarial behavior, which ranges from extreme selfishness to destructive maliciousness, is analyzed. Then, two methods are discussed to counter such adversarial behavior: one is a differentiated pricing to punish the malicious users and another is a detection method based on the expected utility functions of the "regular" users on the network. Finally, the results obtained are illustrated with multiple examples and numerical simulations.

Keywords: Adversarial behavior, mechanism design, game theory, detection and counter measures, interference management, rate control.

1. INTRODUCTION

The behavior of different users (players) on networks may range from altruistic on the one end to malicious (adversarial) on the other end of a wide spectrum (see Figure1). While altruistic users aim to improve the overall network performance, a selfish player strategise to maximize her throughput by getting the proportional share of resources. A malicious user, however, tries to get a disproportionate share of network resource, and in addition may disrupt the whole network. Well-known examples of this adversarial behavior in networks include jamming in wireless networks and denial-of-service (DoS) attacks [1]- [2].

In this paper, we model the coexistence of altruistic, selfish and malicious players using a noncooperative game theoretic formulation and adopt a mechanism design approach. Here, we assume that malicious users mainly stay within the rules of of the system but exhibit adversarial behavior. We model them by assigning different utility functions than selfish players, such as own selfish utility minus the sum of utility of other users in the system or a convex one in contrast to the usually concave utility functions of selfish users. Thus, we map their destructive behavior such as jamming other players and launching Denial-of-Service(DoS) attacks to rational incentives.

To analyze the effects of adversarial behavior, we quantify the robustness of some known network mechanisms with respect to the adversarial behavior of (some of) their participants. A modified version the metric called Price of Malice [3]- [4] is defined suitable for games in network resource allocation and applied to two different network problems dealt here. In the cases analyzed, the malicious players are assumed to take the maximum resource share possible without detection and that way try to disrupt others.

To counter the adversarial behavior, we design mechanisms in which the prices are varied differentially to punish the malicious players after detecting them using any threshold detection technique based on the bids of users. Clearly, when the malicious users does not abide by the rules and vandalize the system, harder responses such as blocking the malicious users after detection are required. We employ a differentiated pricing scheme in which both aggressively selfish and malicious players with disproportional usage of resources are made to pay higher prices than regular selfish players. The effectiveness of this method is quantified using a specific trade-off metric defined. In addition, we discuss an approach for detecting malicious players using the fact that their utility functions do not belong to the same class of utility functions expected for selfish players. After detection the mechanism designer or other good players can ban or punish these malicious players.

Figure 1: Different behavior of users in networks

We consider two different types of network problems in this paper, which differ in coupling of users, i.e. how their actions affect each other, and resource sharing methods. The first one is rate (congestion) control with additive resource sharing, e.g. sharing of bandwidth at a link with fixed capacity. The second one is interference management, e.g. uplink power control in CDMA wireless networks with interference coupling. While allocating these divisible resources to selfish users, a loss in social welfare is caused at the resulting Nash equilibrium due to the selfish nature often referred as

GAMECOMM 2011, May 16, Paris, France

DOI 10.4108/icst.valuetools.2011.245776

Price of Anarchy. Mechanisms such as auctions and pricing mechanisms are proposed to shift the Nash equilibrium point to efficient point. In these mechanisms and underlying games, the selfish nature of rational users were modeled with concave utility functions. But in practical situations, there are altruistic users who care for the welfare of all the users and adversarial users who may deviate from equilibrium point even if it causes loss to them or will show extreme selfishness, i.e. they behave 'irrationally' if modeled using this class of utility functions. We retain the rationality assumption by associating them with different utility functions. In the presence of the these altruistic and adversarial agents, the mechanisms employed will have Nash equilibrium different from the efficient point and this deviation is captured in the metric Price of Malice. In this paper, Price of Malice is quantified for some specific network mechanisms and these mechanisms are modified to punish the adversarial users to make them come back to regular selfish behavior, which brings the system to the efficient Nash equilibrium point.

1.1 Related works

In networked systems with selfish users, a loss in overall social welfare was identified and referred as *Price of Anarchy* [5, 6]. In the presence of malicious users this concept was extended and *Price of Byzantine Anarchy* and *Price of Malice* was first introduced in [3]. They obtained bounds on these metrics which are parameterized by the number of malicious users for a virus inoculation game for social networks. A modified definition was proposed in [4] for congestion games based on the delay experienced at Nash equilibrium point with and without the presence of a malicious player. Both of these works have observed a *Windfall of Malice*, where malicious behavior actually improves the social welfare of non-oblivious selfish users due to the better cooperation resulting because of the 'fear factor' or effects similar to Braess's paradox [4]. In [7] a more general definition of Price of Malice is given with weaker assumptions than above mentioned works in the presence of Byzantine players and using a no-regret analysis. A game theoretic model for the strategic interaction of legitimate and malicious players is introduced in [8], where the authors have derived a bound on the damage caused by the malicious players. In [9], partial altruism of some of the users is analyzed and a bound on Price of Anarchy was obtained as a function of the altruism parameter. To get around with *Price of Anarchy*, pricing for price taking users [10–12] and auctions for price anticipating users [13, 14] are employed. In [15], the effect of spiteful behavior of some of the users is analyzed in the context of first and second price auctions and the revenue obtained is compared. In this paper, we quantify the Price of Malice of the mechanisms proposed for network resource allocation and modify the rules of these mechanisms to counter the malicious behavior.

To counter the adversarial behavior, Micali & Valiant in [16], have developed a modified Vickrey-Clarke-Groves(VCG) mechanism, taking into account collusive, irrational, and adversarial behavior for combinatorial auctions. In the proposed mechanism, the price charged to an agent is increased from VCG price by a scaled factor of the maximum social welfare of other agents. In spirit of this, we also modify the pricing in the proportional fair allocation mechanisms to punish the malicious users and incentivise them to come to regular selfish behavior.

The main contributions of this paper include:

1. Quantifying the Price of Malice in various network mechanisms with adversarial users.
2. Design of differentiated pricing scheme to punish adversarial users and definition of a trade-off parameter.
3. Detection of malicious users by comparing their (observed) utility function to those of regular (selfish) users.

The rest of the paper is organized as follows. The next section presents the underlying mechanism design model. Subsequently, Section 3 quantifies the Price of Malice of the network mechanisms with respect to the adversarial behavior. In Section 4, a differentiated pricing scheme to counter the adversarial behavior is introduced and a method to detect malicious agents is presented. Numerical simulations and their results are shown in Section 5. The paper concludes with remarks of Section 6.

2. MECHANISM DESIGN AND GAME MODEL WITH HETEROGENEOUS USERS

At the center of the mechanism design model is the *designer* $\mathcal{D}$ who influences N *users*, denoted by the set $\mathcal{A}$, and participating in a **strategic (noncooperative) game**. These players are autonomous and rational decision makers, who share and compete for limited resources of the network under the given constraints of the environment. Let us define an N-player strategic game, $\mathcal{G}$, where each player $i \in \mathcal{A}$ has a respective **decision variable** x_i such that

$$x = [x_1, \ldots, x_N] \in \mathcal{X} \subset \mathbb{R}^N,$$

where $\mathcal{X}$ is the decision space of all players. Let

$$x_{-i} = [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_N] \in \mathcal{X}_{-i} \subset \mathbb{R}^{N-1},$$

be the profile of decision variable of players other than i^{th} player and $\mathcal{X}_{-i}$ is the respective decision space. As a starting point, this paper assumes scalar decision variables and a compact and convex decision space. The decision variables may represent, in network resource allocation problems, player flow rate, power level or Signal to Interference Ratio (SINR). Due to the inherent coupling between the players, the decisions of players directly affect each other's performance as well as the aggregate allocation of limited resources.

The **preferences** of the players are captured by utility functions

$$U_i(x) : \mathcal{X} \to \mathbb{R}, \quad \forall i \in \mathcal{A},$$

which are chosen to be continuous and differentiable for analytical tractability. In this paper, the selfishness nature of users are captured by continuous and differentiable concave utility functions.

We consider here a mechanism design having heterogeneous users in the induced game, in which one subset of users have 'abnormal' utility function compared to the class of regular selfish users. The utility function of the class of malicious users can be very different depending on their nature and goals. The disrupting nature of malicious users where they want to create loss to other users even at the cost of their benefit and the altruistic nature of some users who want to care for the social welfare can be captured with a modified utility function. One such modified utility function can be obtained by a convex combination of user utilities

$$U_i^m = (1 - \theta_i)U_i + \theta_i \sum_j U_j, \tag{1}$$

where θ_i is the parameter between -1 and 1 which captures the range of behavior of a user given in Figure 1. This utility function can be modified by taking the average of the utilities of all the

users in the second term [9]. The table below lists the values of θ and corresponding user behavior.

θ	Behavior
$0 < \theta < 1$	altruistic
$\theta = 0$	selfish
$-1 < \theta < 0$	malicious

Let us define the set of selfish users be $\mathcal{S} \subset \mathcal{A}$. Also, the set of malicious and altruistic users, i.e. users with $\theta_i \neq 0$ is defined as $\mathcal{B}$ and $\mathcal{B} = \mathcal{A} \backslash S$. When the set $\mathcal{B}$ has only malicious users, the utility function of malicious users can be modified as

$$U_i^m = U_i + \theta_i \sum_{j \in \mathcal{S}} U_j, \ \forall i \in \mathcal{B}. \tag{2}$$

The extreme selfishness or greedy nature of malicious users can be also captured with a convex utility function. In this case they will take the maximum possible share of the resource constrained above by either physical limits or a level that leads to immediate detection.

The designer $\mathcal{D}$ devises a **mechanism** M, which can be represented by the mapping $M : \mathcal{X} \rightarrow \mathbb{R}^N$, implemented by introducing incentives in the form of *rules and prices* to players. The latter can be formulated by adding it as a cost term such that the player i has the quasi-linear cost function

$$J_i(x) = c_i(x) - U_i(x). \tag{3}$$

where $c_i(x)$ is the price payed by i^{th} user to the mechanism.

We differentiate between two kinds of mechanisms, auctions and pricing, which differ in the assumption on nature of the users and the interaction rules. In auction mechanisms, the designer $\mathcal{D}$ imposes on a player $i \in \mathcal{A}$ a user-specific

- resource allocation rule, $Q_i(x)$,
- resource pricing, $c_i(x)$,

based on their bids x. The price anticipating users decide on their bid, minimizing their individual cost.

In pricing mechanisms, the price taking users decide on their allocation as best response to the user specific price P_i induced by the designer and there is no explicit allocation rule dictated by the designer. In this case, the cost function is

$$J_i(x) = P_i x_i - U_i(Q(x)).$$

Similar to player preferences, the **designer objective**, e.g. maximization of aggregate user utilities or social welfare, can be formulated using a smooth objective function V for the designer:

$$V(x, U_i(x), c_i(x)) : \mathcal{X} \rightarrow \mathbb{R},$$

where $c_i(x)$ and $U_i(x)$, $i = 1, \ldots, N$ are user-specific pricing terms and player utilities, respectively. Hence, the global optimization problem of the designer is simply $\max_x V(x, U_i(x), c_i(x))$, which it solves *indirectly* by setting rules and prices. The different properties of mechanisms analyzed in this paper are attached in the appendix.

In this paper, for tractability purposes, we model user i's utility function as logarithmic, parameterized by her private value α_i. In this case, the aim of the designer in the auction setting will be to make the users report their true private value, i.e, $x_i = \alpha_i$ and carry out an efficient allocation based on that.

The players share and compete for limited resources in the given environment under its information and communication constraints. We focus on two basic types of resource sharing and coupling, which are often encountered in a variety of problems in networking:

1. *Additive resource sharing:* the players share a finite resource C such that

   $$\sum_{i=1}^{N} x_i = C.$$

   This type of coupling is encountered in bandwidth sharing and rate control in networks.

2. *Interference coupling* (linear interference): the resource allocated to player i, γ_i, is inversely proportional to interference generated others such that

   $$\gamma_i(x) = \frac{h_i x_i}{\sum_{j \neq i} h_j x_j + \sigma},$$

   where $h_i \ \forall i$ and σ denote some system parameters. Interference coupling occurs in wireless networks where γ represents signal-to-interference ratio.

We assume that the malicious users have information about the utility function of other selfish users but the regular selfish users do not have the information about the existence of malicious users and their identities. We consider the case where the selfish users cannot collaborate, detect and punish malicious users themselves since it will require a lot of common information and communication for coordination. Therefore, we use a designer who anticipates and detects malicious behavior of any user and modifies the pricing appropriately to counter the malicious behavior. In the next sections, the effect of malicious users to system social welfare is quantified and some counter measures are proposed.

3. PRICE OF MALICE IN MECHANISMS

In this section, we quantify the robustness of mechanisms described in the above setting, against malicious players. For this purpose, we first redefine the metric Price of Malice ($PoM(M)$) of mechanism M suitable for mechanisms in resource sharing setting. A similar metric called Price of Byzantine Anarchy is used in [3] to quantify the social welfare loss at Nash equilibrium point in the presence of malicious users compared to the optimal point, but in a virus inoculation game scenario. For congestion games with malicious flow concentrated on one malicious player Price of Malice was redefined, based on the delay experienced at Nash equilibrium point with and without the malicious player in [4].

DEFINITION 1. *The metric Price of Malice of a mechanism M*

is defined as:

$$PoM(M) := \frac{\sum_{j\in\mathcal{S}} U_j(Q_j(x)) - \sum_{j\in\mathcal{S}} U_j(Q'_j(x))}{\sum_{j\in\mathcal{S}} U_j(Q_j(x))},$$

where Q is the allocation at the Nash equilibrium when none of the users are malicious and Q' is the allocation at the Nash equilibrium in the presence of malicious users.

Now, we estimate the value of Price of Malice parameter for two networks which differ in user coupling and resource sharing as described in the previous section.

3.1 Price of Malice in Auction Mechanisms

We present auction mechanisms [17] for two network coupling schemes, rate control in wired networks and power allocation in interference coupled wireless networks, and quantify the Price of Malice for both cases. The adversarial behavior considered in this section is that the malicious players take maximum possible share of the resources and hence try to disrupt others by denying them their fair share of resources.

Additive Sharing (Rate Control in Networks)

We consider the rate sharing problem with users having separable utility in networks and quantify the effect of the adversarial behavior on it. Let users with utilities $U_i(Q_i) = \alpha_i \log Q_i(x)$ share a fixed bandwidth C such that $\sum_{i=1}^{N} Q_i(x) = C$, where $x_i \in (0, x_{max})$. The vector x in this case denotes player flow rates and Q the capacity allocated to them [19, 20]. Consider the utility function given in (2) and the cost of i^{th} user is given by,

$$J_i^m = c_i - U_i - \theta_i \sum_{j\in\mathcal{S}} U_j.$$

The designer solves the constrained optimization problem

$$\max_Q V(Q) \Leftrightarrow \max_Q \sum_i U_i(Q_i) \text{ such that } \sum_i Q_i = C, \quad (4)$$

in order to find a globally optimal allocation Q that satisfies this **efficiency criterion**. The associated Lagrangian function is then

$$L(Q) = \sum_i U_i(Q_i) + \lambda \left(C - \sum_i Q_i \right),$$

where $\lambda > 0$ is a scalar Lagrange multiplier. Under the convexity assumptions made, this leads to

$$\frac{\partial L}{\partial Q_i} \Rightarrow U_i'(Q_i) = \lambda, \ \forall i \in \mathcal{A}, \quad (5)$$

and the efficiency constraint

$$\frac{\partial L}{\partial \lambda} \Rightarrow \sum_i Q_i = C. \quad (6)$$

and $Q_i = 0$ for users with $U_i'(Q_i) < \lambda$.

Let the designer employ the total payment to i^{th} user as the one obtained in [14] assuming all the users are just selfish in an efficient auction mechanism M_a with proportional allocation which is defined based on the bid of player i,

$$Q_i := \frac{x_i}{\sum_j x_j + \omega} C. \quad (7)$$

The total payment of i^{th} user is

$$c_i = \log(1 + \frac{x_i}{\sum_{j\neq i} x_j}) \sum_{j\neq i} x_j,$$

as given by [14].

Using this payment function the best response of user having the modified utility function becomes

$$\frac{\partial J_i^m}{\partial x_i} = 0 \Longrightarrow x_i = \frac{\alpha_i \sum_{j\neq i} x_j}{\sum_{j\neq i} x_j + \theta_i \sum_{j\in\mathcal{S}} \frac{\alpha_j}{x_j}}.$$

We can observe that malicious users having $-1 \leq \theta < 0$, will report $x_i > \alpha_i$ and altruistic users having $0 < \theta \leq 1$, will report $x_i < \alpha_i$.

The allocation for the regular selfish users, i.e., users with $\theta_i = 0$ in the presence of altruistic or malicious users can be written as

$$Q_i' = \frac{\alpha_i C}{\sum_{j\in\mathcal{S}} \alpha_j + \sum_{k\in\mathcal{B}} \frac{\alpha_k \sum_{j\neq k} x_j}{\sum_{j\neq k} x_j + \theta_k (N - |\mathcal{B}|)}}. \quad (8)$$

Let

$$r_i = \frac{Q_i}{Q_i'} = \frac{\sum_{j\in\mathcal{S}} \alpha_j + \sum_{k\in\mathcal{B}} \frac{\alpha_k \sum_{j\neq k} x_j}{\sum_{j\neq k} x_j + \theta_k (N - |\mathcal{B}|)}}{\sum_{j\in\mathcal{S}} \alpha_j + \sum_{k\in\mathcal{B}} \alpha_k}. \quad (9)$$

For this additive resource sharing case, the Price of Malicious $PoM(M_a)$ is

$$PoM(M_a) = \frac{\sum_{j\in\mathcal{S}} \alpha_j \log(r_j)}{\sum_{j\in\mathcal{S}} \alpha_j \log(\frac{\alpha_j C}{\sum_i \alpha_i})}.$$

For the case where users are symmetric $\alpha_i = \alpha, \ \forall i$, and only one user is malicious or all the malicious user coordinate to form one entity, this simplifies to

$$PoM(M_a) = \frac{\log(\frac{N-1+\frac{\alpha}{\alpha+\theta_k}}{N})}{\log(\frac{C}{N})}.$$

From the above equations, the Price of Malice of the mechanism can be obtained knowing system parameters and user preferences and can be bounded above and below depending on the range and distribution of these values for the specific setting. Note that for the system with altruistic users and selfish users PoM will be negative, which can be clearly observed from the equation for symmetric case above, which means that the altruistic users are improving social welfare of other selfish users by taking lesser share of resource. In the case of malicious users, as θ decreases from 0 to -1, we can see that the $PoM(M_a)$ increases. We can also observe that when both malicious and altruistic users are present along with selfish ones, their effect on PoM depends on the number and degree of value of θ. If they are of same number and degree,i.e., if $|\theta_i| = |\theta_j|$ for $i \neq j$, then their effect on PoM cancel each other. The variation of values of $PoM(M_a)$ for different values of θ is given in the simulation section.

Interference Coupled Systems (CDMA Power Control)

Consider an auction mechanism in the context of a wireless network and uplink power control setting ([18, 21]) where due to the

interference coupling the resource sharing is inherently competitive. Let the user utilities be defined as $U_i(x) = \alpha_i \log \gamma_i(Q(x))$ and the individual power levels, Q, satisfy $\sum_{i=1}^{N} Q_i \leq C$, where the signal-to-interference ratio (SINR) received by the base station is

$$\gamma_i = \frac{Q_i(x)}{\sum_{j \neq i} Q_j(x) + \sigma},$$

and $x_i \in (0, x_{max})$.

An auction-based mechanism, $\mathcal{M}_b$, can be defined based on the bid of player i, with the resource allocation rule

$$Q_i := \frac{x_i}{\sum_j x_j} C, \tag{10}$$

which is proportional allocation as first analyzed in [14]. We can see that using this proportional allocation, full utilization of resource is attained, i.e. $\sum_i Q_i = C$. Now we decouple the user utilities by rewriting γ_i as

$$\gamma_i(Q_i) = \frac{Q_i(x)}{C - Q_i(x) + \sigma}, \tag{11}$$

using the full utilization property. For the allocation given in (10), the SINR is

$$\gamma_i(x) = \frac{x_i}{\sum_j x_j (C + \sigma) - x_i}. \tag{12}$$

In [18], it is observed that in systems with sufficiently high SINR assumption $U_i(x) = \alpha_i \log \gamma_i(Q_i(x))$ is concave in Q_i, where $\gamma_i(Q_i)$ is given by (11).A pricing mechanism given in [14] will make the selfish users to report $x_i = \alpha_i$.

In the presence of malicious and altruistic users the SINR obtained by the regular users will be,

$$\gamma_i'(x) = \frac{\alpha_i}{(\alpha^s + \sum_{k \in \mathcal{B}} \frac{\alpha_k \sum_{j \neq k} x_j}{\sum_{j \neq k} x_j + \theta_k (N - |\mathcal{B}|)})(C + \sigma) - \alpha_i} \tag{13}$$

where $\alpha^s = \sum_{j \in \mathcal{S}} \alpha_j$.

Then, we obtain $PoM(M_b)$ as

$$PoM(M_b) = \frac{\sum_{j \in \mathcal{S}} \alpha_j \log(\frac{\gamma_j}{\gamma_j'})}{\sum_{j \in \mathcal{S}} \alpha_j \log(\frac{\alpha_j C}{\sum_k \alpha_k (C+\sigma) - \alpha_j})}.$$

In the symmetric case and only one user is malicious, the PoM becomes

$$PoM(M_b) = \frac{\log(\frac{(N-1+\frac{\alpha}{\alpha+\theta_k})(C+\sigma)-1)}{N(C+\sigma)-1})}{\log(\frac{C}{N(C+\sigma)-1})}.$$

A similar behavior of $PoM(M_b)$ as in the case of additive sharing can be observed for different values of θ. The variation of $PoM(M_b)$ for different values of θ is given in the simulation section for a specific set of parameters.

3.2 Price of Malice in Pricing Mechanisms

A counterpart of the Price of Malice metric in Definition 1 for pricing mechanisms [10], which differ from auctions by their lack of an explicit resource allocation scheme, can be obtained by replacing $Q(x)$ and $Q'(x)$ with the action vector without malicious users x and with malicious users x', respectively.

In the case of additive resource sharing, the users with utilities $U_i(x_i) = \alpha_i \log x_i$ share the fixed resource $\sum_{i=1}^{N} x_i = C$, and $x_i \in (0, x_{max})$. Consider an efficient mechanism M_c, which can be implemented in an iterative way. The efficient allocation is

$$x_i = \frac{\alpha_i}{\lambda},$$

where λ is the Lagrange multiplier. In the case of all selfish users $\lambda = \sum_i \alpha_i / C$ and it will be set as price to the users.

Let each malicious user take a share x_m which can be x_{max}, the maximum share they can use without detection, according to their utility function, in order to disrupt others. Let λ' be the Lagrange multiplier in this case which will be a different point than $\lambda = \sum_i \alpha_i / C$. The remaining resource $(C - \sum_{\mathcal{B}} x_m)$ will be shared among good users, under the efficient mechanism M_c. In the additive sharing case $PoM(M_c)$ is,

$$PoM(M_c) = \frac{\sum_{j \in \mathcal{S}} \alpha_j \log(\frac{C\lambda'}{\sum_i \alpha_i})}{\sum_{j \in \mathcal{S}} \alpha_j \log(\frac{\alpha_j C}{\sum_i \alpha_i})}.$$

For symmetric case, where $\alpha_i = \alpha \ \forall i$, it becomes

$$PoM(M_c) = \frac{\log(\frac{C\lambda'}{N\alpha})}{\log(\frac{C}{N})}.$$

The counterpart of auction in the interference-coupled case for pricing can be obtained in a similar way and the mechanism can be denoted as M_d. The variation of values of $PoM(M_c)$ and $PoM(M_d)$ for different number of users is given and compared with each other in the simulation section.

4. RESPONSE MECHANISMS TO MALICIOUS USERS

The robustness analyses in the previous sections only measure the effect of selfish and malicious users but does not provide a way to encounter them. In this section, we consider two possible response schemes to adversarial behavior: one based on softer punishment scheme using differentiated pricing and the other relying on detection of malicious users observing their utility function. The latter has to be supported by a separate punishment mechanism which should follow the detection phase.

4.1 Differentiated Pricing

We consider a softer response scheme than blocking towards malicious users after explicit detection based on any well known (threshold) detection scheme. The response mechanism is implemented by the designer by deploying a differentiated pricing. First, we define a trade-off metric $T(M)$ for quantifying the effectiveness of a pricing-based response to a mechanism M. This metric provides a way to measure the trade-off between the damage due to malicious users and how much effort (price) it costs them to create this damage.

DEFINITION 2. *A metric for quantifying effectiveness of a pricing-based response mechanism against a set of malicious users $B \subset \mathcal{A}$*

is defined as:

$$T(M) \geq \frac{\sum_{j\in\mathcal{S}} U_j(Q'_j(x)) - \sum_{j\in\mathcal{S}} U_j(Q_j(x))}{\sum_{k\in\mathcal{B}} c_k(x)},$$

and the lower bound is achieved in the best case scenario of perfect differentiation in terms of pricing.

Now we utilize this metric to evaluate the properties of the differentiated pricing scheme on example networks. A necessary assumption we make in this subsection is that malicious users stay within the system and do not have any means to evade the pricing mechanisms imposed by the designer. This assumption is relaxed in the next subsection.

Auctions for Additive Sharing

We derive now a differentiated payment function to counter the malicious behavior of users. It is assumed here that the designer knows the value of θ of malicious user. In practical problems, this is not realistic and the designer needs to make the decision on payment function entirely based on user bids. Therefore, we assume that after detecting the malicious user using a threshold detection scheme based on the bids, the designer punishes the malicious users with a price function assuming $\theta = -1$, i.e, extreme maliciousness. Alternatively, once can couple this parameter with the confidence of the detection scheme used, i.e. low θ values for high probability of malicious behavior and vice versa. The best response of the i^{th} user who tries to minimize her cost in terms of the signal or bid to be sent is obtained by computing

$$\frac{\partial J_i}{\partial x_i} = \frac{\partial c_i}{\partial x_i} - \frac{\partial U_i}{\partial Q_i}\frac{\sum_{j\neq i} x_j}{(\sum_k x_k)^2} + \theta_i \sum_{j\neq i} \frac{\alpha_j}{x_j \sum_k x_k} = 0. \quad (14)$$

This condition is necessary and sufficient for optimality. Then,

$$\frac{\partial U_i(Q_i)}{\partial Q_i} = \frac{(\sum_k x_k)^2}{\sum_{j\neq i} x_j}\left(\frac{\partial c_i}{\partial x_i} + \theta_i \sum_{j\neq i} \frac{\alpha_j}{x_j \sum_k x_k}\right).$$

Let us denote $t = \sum_j x_j$, then $x_i = \dfrac{tQ_i}{C}$ and

$$\sum_{j\neq i} x_j = t - x_i = t\left(1 - \frac{Q_i}{C}\right).$$

Doing the substitutions,

$$\begin{aligned}\frac{\partial U_i(Q_i)}{\partial Q_i} &= \frac{t}{1-\dfrac{Q_i}{C}}\left(\frac{\partial c_i(Q_i,t)}{\partial x_i} + \theta_i \sum_{j\neq i}\frac{1}{t}\right) \\ &:= f(Q_i,t). \end{aligned} \quad (15)$$

When we compare (15) and (5), we can see that $f(Q_i,t)$ is equal to the Lagrange multiplier λ. Since $f(Q_i,t)$ is a function of Q_i, there will be unequal marginal valuations at equilibrium. For efficient allocation we need to obtain a price function which will induce a $f(Q_i,t)$ which will give identical marginal valuations at equilibrium [14]. For this we make $f(Q_i,t)$ independent of Q_i and derive corresponding price function. Let $f(Q_i,t) = g(t)$ where $g(t)$ is the generator function and

$$\frac{\partial c_i}{\partial x_i} = \frac{\sum_{j\neq i} x_j g(t)}{(\sum_k x_k)^2} - \theta_i \frac{1}{\sum_k x_k}\sum_{j\neq i}\frac{\alpha_j}{x_j}.$$

By integrating over x_i, we obtain

$$\begin{aligned} c_i(x) &= \int_0^{x_i} \frac{g(s+\sum_{j\neq i} x_j)}{(s+\sum_{j\neq i} x_j)^2} ds \sum_{j\neq i} x_j \\ &- \theta_i \int_0^{x_i} \frac{ds}{s+\sum_{k\neq j} x_k} \sum_{j\neq i}\frac{\alpha_j}{x_j}. \end{aligned} \quad (16)$$

For $g(t) = t$, we obtain

$$\begin{aligned} c_i(x) &= \log\left(1+\frac{x_i}{\sum_{j\neq i} x_j}\right)\sum_{j\neq i} x_j \\ &- \theta_i \log\left(1+\frac{x_i}{\sum_{j\neq i} x_j}\right)\sum_{j\neq i}\frac{\alpha_j}{x_j}. \end{aligned} \quad (17)$$

Let us assume that the users except i^{th} user are merely selfish due to the payment function of the mechanism they report $x_i = \alpha_i$. If the designer punishes the users who are detected as malicious with a payment in which $\theta_i = -1$, then the final pricing function becomes

$$c_i(x) = \log\left(1+\frac{x_i}{\sum_{j\neq i} x_j}\right)\left(\sum_{j\neq i} x_j + (N-1)\right). \quad (18)$$

Now we can define a mechanism M_m which is defined by the allocation rule (7) and pricing rule given by (18). Note that in this differentiated pricing scheme, the malicious users who will try to bid something higher than its private value will have to pay an additional amount proportional to their bid. The tradeoff-parameter of mechanism M_m is given by,

$$T(M_m) \geq \frac{\sum_{j\in\mathcal{S}} \alpha_j \log(r_j)}{\sum_{i\in\mathcal{B}} \log\left(1+\dfrac{x_i}{\sum_{j\neq i} x_j}\right)\left(\sum_{j\neq i} x_j + (N-1)\right)}.$$

Such a differentiated pricing scheme is widely used today in various settings, such as network access. For example, if some users of an Internet Service Provider (ISP) are creating burden to the network by using much higher amount of resources above a predetermined cap, they are priced differentially higher compared to other users. This reality is captured in our model since the higher usage above a threshold is punished even if it is not coming from the disproportionate use due to malicious nature.

In a similar way, a differentiated pricing mechanism can be also derived for interference coupled CDMA systems.

Pricing Mechanism for Additive Sharing

Let us consider the counterpart of pricing mechanism in additive sharing given in the previous section and study the effect of the differentiated pricing in that case. As one possibility we model the utility function of a malicious player as $U_i(x_i) = e^{\beta_i x_i}$, which reflects aggressive behavior in terms of resource demand. Note that this is still private information unknown to the designer. As a result, A malicious user takes a share of $x_m \in (\overline{x} + \epsilon, x_{max})$, where $\overline{x}$ is the mean and ϵ is some integer multiple of standard deviation of the demand vector x.

In order to counter the malicious behavior, the designer deploys differentiated pricing as part of a new mechanism M_e, which is a

modified version of M_c. It is characterized by the pricing function

$$P_i^d = \begin{cases} f(\kappa_i(x_i - (\overline{x} + \epsilon))) & \text{for } x_i \geq b \\ P_i & \text{for } x_i \leq b \end{cases},$$

where b is determined by a statistical method, for example $b = \overline{x} + k\sigma_x$, where $\overline{x}$ is the mean and σ_x is standard deviation and P_i is the pricing function in the original mechanism. The function $f(.)$ is selected suitably depending on the utility functions of selfish and malicious users. If it is assumed that selfish users have continuous and differentiable concave utility function and malicious users have convex utility functions, then $f(.)$ can be a continuous and differentiable convex function. For the logarithmic utility function assumed here for selfish users, we take $f(.)$ as exponential function. The value of b can be obtained alternatively from a clustering method or another Maximum-likelihood algorithm. Note that, the designer punishes the malicious players by employing a price function which increases exponentially with the share of resource taken by them, i.e. if they deviate too much from the mean behavior and create a significant burden on the system.

For the case of exponential pricing function, $T(M_e)$ is obtained as,

$$T(M_e) \geq \frac{\sum_{j \in \mathcal{S}} \alpha_j \log(\frac{C\lambda'}{\sum_i \alpha_i})}{\sum_{i \in \mathcal{B}} e^{\kappa_i(x_i - (\overline{x}+\epsilon))}}.$$

In the symmetric and only one malicious user case, it becomes

$$T(M_e) \geq \frac{\log(\frac{C\lambda'}{N\alpha})}{e^{\kappa_i(x_i - (\overline{x}+\epsilon))}}.$$

Pricing Mechanism for Interference Coupled Systems

Consider the case of pricing in interference coupled systems given in Section 3. To counter the malicious behavior, the designer introduces a new mechanism M_f which employs the differentiated pricing given by

$$P_i^d = \begin{cases} f(\kappa_i(\gamma_i(x_i, x_{-i}) - \gamma_i(\overline{x} + \epsilon, x_{-i}))) & \text{for } x_i \geq b \\ P_i & \text{for } x_i \leq b \end{cases},$$

In the case of logarithmic utility, $P_i = \lambda + \sum_{j \neq i} \frac{\alpha_j}{I_j}$, where λ is the Lagrange multiplier of the associated optimization problem and $I_i := \sum_{j \neq i} x_j + \sigma$ is the interference affecting player i [11, 12]. For the mechanism M_f, the trade-off metric $T(M_f)$ can be obtained in similar way as for additive sharing case. The variation of values of $T(M_e)$ and $T(M_f)$ for different number of users is given and compared with each other in the simulation section.

4.2 Malicious User Detection based on Utility Functions

It is not realistic in some scenarios to assume that the malicious users follow the rules of the mechanism. Therefore, it will not be possible to handle them by modifying the rules of the mechanism as done in the differentiated pricing case. In this section, we discuss one detection technique which can be used in addition to the classical techniques developed so far. All of the mechanisms discussed in this paper provide a way for the designer to infer the utility function of the users through the observations. The aggressive or malicious user behavior can be captured using utility functions that fundamentally differ from those of regular users. Let us assume that the designer expects that utility functions of the regular users belong to a certain class such as the class of concave functions. Then, if certain users are observed to take a bigger share of resources than the one indicated by the expected class of utility, they can be detected as malicious.

Let us consider the setting given in Section 3 and assume that utility of a malicious player is $U_i(x_i) = e^{\beta_i Q_i}$, for some parameter β_i. The designer can estimate the value of utility parameter α_i of each player i from the share of resource taken Q_i. Let Q_i^* be the intersection of utility function of malicious player and the logarithmic utility function expected from her if she was just a regular selfish player. Then, we obtain

$$e^{\beta_i Q_i^*} = \alpha_i \log Q_i^*,$$

and

$$\alpha_i = \frac{e^{\beta_i Q_i^*}}{\log Q_i^*}.$$

Clearly, the value of α_i obtained for a malicious user will be much higher than those of regular users. In other words, this utility parameter does not fit to any possible logarithmic curve. This is taken as an indication of adversarial behavior. Notice that, this detection scheme is more widely applicable than differentiated pricing since it does not require the malicious users abide by the pricing scheme. After detection, the malicious users can be blocked or punished with an additional method.

5. SIMULATIONS

In this section, computer simulation results are presented to show the different parameters of the proposed mechanisms. First, the Price of Malice $PoM(M_a)$ and $PoM(M_b)$ of pricing mechanism for additive sharing M_a and interference coupling M_b, respectively, using the setup in Section 3 by varying the value of θ from -1 to 0. The number of users $N = 50$ out of which 10 users are taken to be malicious with same θ value. The other system parameters are $C = 30$ and $\sigma = 1$. The simulations are done by generating the player preferences α's according to a uniform distribution on the support set $[0, 10]$ and plotted in Figure 2. It can be observed that value of $PoM(M_a)$ and $PoM(M_b)$ decreases as θ varies from -1 to 0 as expected.

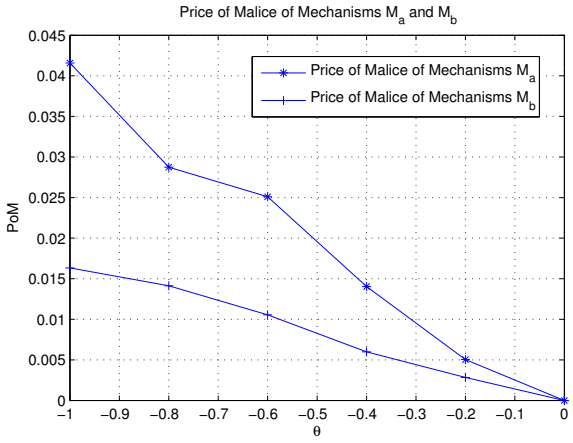

Figure 2: Price of Malice $PoM(M)$ of the pricing mechanism for additive coupling $\mathcal{M}_a$ and interference coupling $\mathcal{M}_b$ for varying values of θ.

We next compute the Price of Malice $PoM(M_c)$ and $PoM(M_d)$ for the auction mechanism for additive sharing M_c and interference coupling M_d, respectively, using the setup in Example 3 by varying the number of users from 8 to 15. The simulations are done by generating the player preferences α's according to a uniform distribution on the support set $[0, 2]$ and repeated 100 times. Then, the mean and standard deviation of the obtained $R(M)$ values are plotted in Figure 3. The number of malicious users is fixed at 3, $C = 5$, $\sigma = 0.5$ and $x_{max} = 1$. The malicious users take an allocation x_{max} and remaining share is allocated using respective iterative algorithms among good users. The quantities $PoM(M_c)$ and $PoM(M_d)$ are plotted in Figure 3. It can be observed that, for a fixed number of malicious users, as number of users increases the mechanisms become more robust, as expected. Next, the trade-off

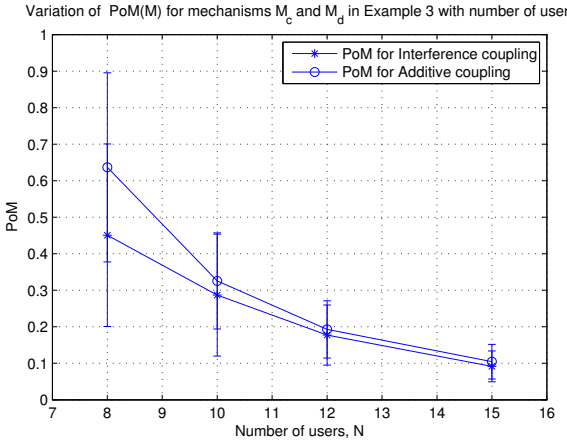

Figure 3: Price of Malice $PoM(M)$ of the auction mechanisms for additive coupling $\mathcal{M}_c$ and interference coupling $\mathcal{M}_d$ in Example 3 for varying number of users.

parameter $T(M)$ is plotted for auction mechanism M_m for additive sharing for different values of θ in Figure 4. The users having $x > \overline{x} + 2\,\sigma_x$ are priced differentially as described in Section 4.

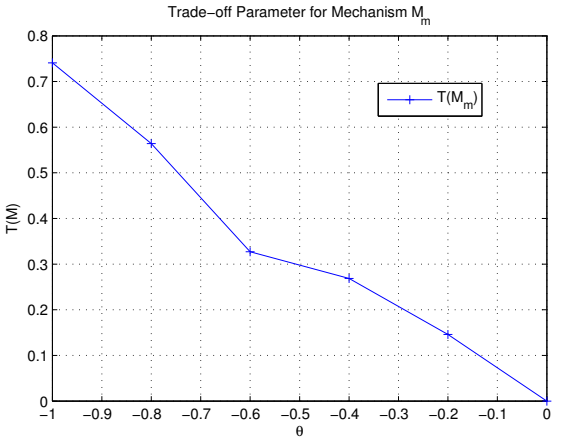

Figure 4: Trade off parameter $T(M)$ of auction mechanism M_m for additive sharing for varying values of θ.

Finally, the trade-off parameter $T(M)$ is plotted for pricing mechanisms M_e and M_f in Figure 5. An iterative algorithm as given in [12] is used to obtain allocation and prices. The other parameters remain the same as those used to generate the Figure 3. It can be seen from Figure 5 that mechanism M_f performs better than M_e in this case, possibly due to the coupling involved.

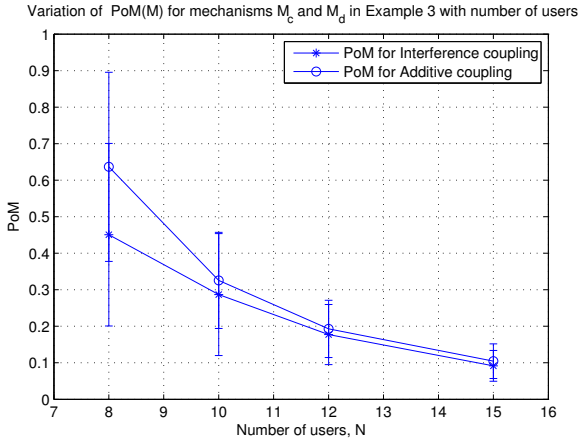

Figure 5: Trade off parameter $T(M)$ of pricing mechanisms for additive coupling $\mathcal{M}_e$ and interference coupling $\mathcal{M}_f$ in Examples 3 and 4 respectively with varying number of users.

6. CONCLUSION

We have studied adversarial behavior in network resource allocation schemes including pricing and auctions by adopting a mechanism design approach to measure and counter it. First, we have analyzed the robustness of the existing network mechanisms to adversarial behavior, which ranges from extreme selfishness to destructive maliciousness, using a quantitative metric Price of Malice. Next, we have presented two methods to counter such adversarial behavior: one is a differentiated pricing to punish the aggressive players and another is a detection method based on the expected utility functions of the "regular" users on the network. Finally, the results obtained have been illustrated with multiple examples and numerical simulations.

Future research directions include obtaining bounds on the parameters dealt in this paper and a study of collusion and related tradeoffs, as well as behavioral detection schemes. It is also an interesting direction to analyze the effect of altruism or partial altruism of some of the users in this context, as in the work [9].

7. ACKNOWLEDGEMENT

This work has been supported by Deutsche Telekom Laboratories.

8. REFERENCES

[1] W. Xu, W. Trappe, Y. Zhang and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks ", MobiHoc '05 Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing , pp. 47-56, 2005.

[2] E. Altman, K. Avrachenkov and A. Garnaev, " A Jamming Game in Wireless Networks with Transmission Cost,"Lecture Notes in Computer Science,Springer Berlin / Heidelberg, pp.1-12, vol.4465, 2007.

[3] T. Moscibroda and S. Schmid and R. Wattenhofer, "When selfish meets evil: byzantine players in a virus inoculation game", Proceedings of the twenty-fifth annual ACM

symposium on Principles of distributed computing, 2006, Denver, Colorado, USA.
[4] M. Babaioff and R. Kleinberg and C. H. Papadimitriou "Congestion games with malicious players", Proceedings of the 8th ACM conference on Electronic commerce, pp. 103-112, 2007, San Diego, California, USA.
[5] E. Koutsoupias and C. Papadimitriou," Worst-Case Equilibria," Lecture Notes in Computer Science, pp. 404âĂŞ413, 1999.
[6] T. Roughgarden, " The Price of Anarchy is Independent of the Network Topology", Proceedings of the 34th Annual ACM Symposium on the Theory of Computing, May 2002.
[7] A. Roth,"The Price of Malice in Linear Congestion Games," In WINE '08: Proceedings of the 4th International Workshop on Internet and Network Economics, Vol. 5385 (2008), pp. 118-125.
[8] S.Theodorakopoulos and J. S. Baras, "Game Theoretic Modeling of Malicious Users in Collaborative Networks, ", IEEE Journal on selected areas in communications, vol.26, no.7, pp.1317-1327, August, 2008.
[9] P.-A. Chen and D. Kempe,"Altruism, Selfishness, and Spite in Traffic Routing", Electronic Commerce, ECâĂŹ08, July 8âĂŞ12, 2008, Chicago, Illinois, USA.
[10] F. P. Kelly and A. K. Maulloo and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," Journal of the Operational Research Society, 1998, volume 49, pp. 237-252.
[11] T. Alpcan and L. Pavel, "Nash Equilibrium Design and Optimization", Proc. of Intl. Conf. on Game Theory for Networks (GameNets 2009),Istanbul, Turkey, May, 2009.
[12] T. Alpcan and H. Boche and S. Naik, "A Unified Mechanism Design Framework for Networked Systems," arxiv.org, September,2010, arXiv:1009.0377.
[13] R. Johari and J. N. Tsitsiklis, "Efficiency of Scalar-Parameterized Mechanisms", Operations Research, pp. 823-839, vol.57, no. 4,July 2010.
[14] R. T. Maheswaran and T. Basar, "Social Welfare of Selfish Agents: Motivating Efficiency for Divisible Resources", 43rd IEEE Conf. on Decision and Control (CDC), , pp. 1550-1555, Paradise Island, Bahamas, December 2004.
[15] F. Brandt and T. Sandholm and Y. Shoham, "Spiteful Bidding in Sealed-Bid Auctions", IJCAI'07 Proceedings of the 20th international joint conference on Artifical intelligence , pp. 1207-1214, 2007, Hyderabad, India.
[16] S. Micali and P. Valiant, "Revenue in Truly Combinatorial Auctions and Adversarial Mechanism Design", MIT-Computer Science and Artificial Intelligence Laboratory http://dspace.mit.edu/handle/1721.1/41872, June 2008.
[17] V. Krishna, Auction theory (2nd ed.), Academic Press, 2010.
[18] J. Huang and R. Berry and M. Honig, "Auction–based Spectrum Sharing", ACM Mobile Networks and Applications Journal, pp.405-418, 2006.
[19] T. Alpcan and T. Başar, "A Utility-Based Congestion Control Scheme for Internet-Style Networks with Delay", IEEE Trans. on Networking, April, 2010.
[20] R. Srikant, "The Mathematics of Internet Congestion Control", Systems & Control: Foundations & Applications, 2004.
[21] A. K. Chorppath and S. Bhashyam and R. Sundaresan, "A Convex Optimization Framework for Almost Budget Balanced Allocation of a Divisible Good"IEEE Transactions on Automation Science and Engineering, no.99, January,2011.
[22] H. Boche and S. Naik, "Mechanism Design and Implementation Theoretic Perspective of Interference Coupled Wireless Systems", Proc. of 47th Annual Allerton Conf. on Communication, Control, and Computing, September, 2009, Monticello, IL, USA.

9. APPENDIX

Definitions:

The properties of mechanisms considered in this paper can be formally defined as follows.

DEFINITION 1. ***Efficiency:*** *Efficient mechanisms maximize designer objective, i.e. they solve the problem* $\max_x V(x, U_i(x), c_i(x))$.

DEFINITION 2. ***Nash Equilibrium:*** *The strategy profile* $x^* = [x_1^*, \ldots, x_N^*]$ *is in Nash Equilibrium if the cost of each player is minimized at the equilibrium given the best strategies of other players.*

$$J_i(x_i^*, x_{-i}^*) \leq J_i(x_i, x_{-i}^*), \forall i \in \mathcal{A}, x_i \in \mathcal{X}_i$$

DEFINITION 3. ***Dominant Strategy Equilibrium:*** *The strategy profile* $\tilde{x} = [\tilde{x}_1, \ldots, \tilde{x}_N]$ *is in Dominant Strategy Equilibrium if the cost of each player is minimized at the equilibrium irrespective of the strategies of other players.*

$$J_i(\tilde{x_i}, x_{-i}) \leq J_i(x_i, x_{-i}), \forall i \in \mathcal{A}, x_i \in \mathcal{X}_i, x_{-i} \in \mathcal{X}_{-i}$$

DEFINITION 4. ***Strategy-proofness or Incentive Compatibility:*** *If the players do not gain anything by reporting a value other than their true value, i.e.*

$$J_i(x_i, x_{-i}) \leq \tilde{J}_i(\tilde{x}_i, x_{-i}), \forall i \in \mathcal{A}, \tilde{x}_i \in \mathcal{X}_i, x_{-i} \in \mathcal{X}_{-i}$$

where x *is the original value vector, and* $\tilde{x_i}$ *is the "misrepresented" value or action, then the mechanism is strategy-proof.*