

Closing the Price of Anarchy Gap in the Interdependent Security Game

Parinaz Naghizadeh and Mingyan Liu

Department of Electrical Engineering and Computer Science

University of Michigan, Ann Arbor, Michigan, 48109-2122

Email: {naghizad, mingyan}@umich.edu

Abstract—The reliability and security of a user in an interconnected system depends on all users' collective effort in security. Consequently, investments in security technologies by strategic users is typically modeled as a public good problem, known as the Interdependent Security (IDS) game. The equilibria for such games are often inefficient, as selfish users free-ride on positive externalities of others' contributions. In this paper, we present a mechanism that implements the socially optimal equilibrium in an IDS game through a message exchange process, in which users submit proposals about the security investment and tax/price profiles of one another. This mechanism is different from existing solutions in that (1) it results in socially optimal levels of investment, closing the Price of Anarchy gap in the IDS game, (2) it is applicable to a general model of user interdependencies. We further consider the issue of individual rationality, often a trivial condition to satisfy in many resource allocation problems, and argue that with positive externality, the incentive to stay out and free-ride on others' investment can make individual rationality much harder to satisfy in designing a mechanism.

I. INTRODUCTION

As a result of the rapid growth of the Internet, networks of all kinds, and file sharing systems, the security of a user, or entity, or a network¹, in the context of a bigger system of connected users, entities or networks, is no longer solely determined by that user's own investment in security, but becomes increasingly dependent on the effort exerted by the collection of interconnected users. Accordingly, the security and reliability of the interconnected system is viewed as a *public good*, for which the investments in security exhibit a *positive externality* effect: the investment of one user on security technologies will also improve the security posture of the other users interacting with it. Consequently, strategic users can choose to free-ride on others' effort, resulting in an overall under-investment in security.

This problem of (under-)investment in security by an interconnected group of selfish users, both in general as well as in the context of computer security, has been extensively studied in the framework of game theory, see e.g. [1], [2], [3], [4], [5], [6], [7], [8], and is often referred to as the Interdependent Security (IDS) game. IDS games were first presented by Kunreuther and Heal [1] to study the incentive of airlines to invest in baggage checking systems, and by Varian [2] in the context of computer system reliability. In the majority of

these papers, under-investment in security is verified by finding the levels of effort exerted in a Nash equilibrium of the IDS game, and comparing them with the socially optimal levels of investment.

The increasing number of unprotected devices connected to the Internet, the constant emergence of new security threats, and the insufficiency of improved security technologies in compensating for the under-investment problem [7], motivates the study of mechanisms for improving network security. Several methods for increasing users' investments, and thus the reliability of the interconnected system, have been proposed in the literature. These mechanisms fall into two main categories, based on whether they *incentivize* or *dictate* user cooperation. Mechanisms that dictate user investment in security, e.g. regulations, audits, and third party inspections [8], leverage the power of an authority such as the government or an Internet service provider (ISP). These methods are only effective if the authority has enough power to accurately monitor users and establish a credible threat of punishment.

Among the mechanisms that incentivize user investment in security, *cyberinsurance* is one of the most commonly studied approaches [1], [9], [8]. Using insurance, users transfer part of the security risks to an insurer in return for paying a premium fee. Cyberinsurance is affected by the classic insurance problems of adverse selection (higher risk users seek more protection) and moral hazard (users lower their investment in self-protection after being insured). Therefore, the insurance company needs to somehow mitigate the information asymmetry and calculate the premium fees with these considerations in mind. An example of such solutions is when an insurer chooses to monitor investments and/or inspect users' devices to prevent the moral hazard problem, specifying the terms of the contract accordingly to ensure appropriate levels of investment in self-protection [1].

A method similar to insurance is proposed in [3], where a certifying authority classifies users based on whether or not they have made security investments, and ensures that certified users get adequate compensation in case of a security incident. Another theoretically attractive incentive mechanism that may result in optimal levels of investment is the *liability rule* [1], [2], where users are required to compensate others for the damages caused by their under-investment in security. However, these mechanisms are costly in that it is difficult to accurately determine the cause of a damage. Alternatively,

¹These terms are used interchangeably in this paper to denote a single unit in a connected system.

[2] proposes assigning a level of *due care*, in which following a security incident, a user is penalized only if its level of investment is lower than a pre-specified threshold. Finally, users can be incentivized to invest in security if they are assigned bonuses/penalties based on their security outcome (e.g. users get a reward if their security has not been breached), or get subsidized/fined based on their effort (e.g. users are given discounts if they buy security products) [5].

In this paper, we take a mechanism design approach to the security investment problem. Specifically, we present a game form, consisting of a message exchange process and an outcome function, through which users converge to an equilibrium where they make the socially optimal levels of investment in security. Our method is different from the previous solutions in several ways, highlighted as follows.

- 1) The proposed mechanism is applicable to the general model of interdependence proposed in [7]. This model allows continuous levels of effort (as opposed to a binary decision of whether or not to invest in security [1], [3], [6]).
- 2) It does not assume perfect protection once investment is made (unlike epidemic models [1], [8]). Another similar assumption is to decompose the risks of a user into direct and indirect (i.e. spreading from another infected user) risks, and assume perfect protection against direct risks only [8]. Nevertheless, none of these models can be descriptive of an IDS game, as no security technology can provide perfect protection against all threats.
- 3) It models the heterogeneity in users' preferences and their importance to the system by allowing for a more general utility function (in contrast to [1], [2], [4], [5], [9], [6]).
- 4) This mechanism not only improves the levels of investment (as also done in [7]), but in fact results in socially optimal investments in security.

The rest of this paper is organized as follows. In Section II, we present a model for the IDS game. We introduce the concept of price of anarchy in Section III, and highlight the inefficiency of Nash equilibria in an unregulated IDS game through a simple example. We discuss the decentralized mechanism and its optimality in Section IV. Section V illustrates that such optimal mechanism may fail to be individually rational, typically a trivial requirement in many other settings. Section VI concludes the paper with directions for future work.

II. MODEL AND PRELIMINARIES

Consider a collection of N users; this collection will also be referred to as the system. Each user i can choose a level $x_i \geq 0$ of effort/investment in security, incurring a cost $c_i > 0$ per unit of investment. Let $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$ denote the vector of investments. A user i 's *security risk* function is denoted by $f_i(\mathbf{x})$. The security risk function models the expected losses of an individual in case of a security breach. These functions vary among users depending on both their

security interdependencies and their valuations of security. We make the following assumptions about the functions $f_i(\cdot)$:

Assumption 1: $f_i(\cdot) > 0$ is differentiable and decreasing in x_j , for all i and all j .

The assumption of $\partial f_i / \partial x_j < 0$ models the positive externalities of security investments.

Assumption 2: $f_i(\cdot)$ is strictly convex.

The assumption of convexity means that initial investment in security offers considerable protection to the users [8], [10]. However, even with high effort, it is difficult to reduce the cost to zero, as there is no strategy that could prevent all malicious activity [7], [10].

The utility function of a user i is defined as:

$$u_i(\mathbf{x}) = -f_i(\mathbf{x}) - c_i x_i - t_i . \quad (1)$$

In (1), $g_i(\mathbf{x}) := f_i(\mathbf{x}) + c_i x_i$ is referred to as the *cost function* of user i [7], and represents all the costs associated with security investments and breaches. The term t_i is the *monetary transfer* that can be imposed on/awarded to users throughout the mechanism, which may itself depend on the vector of investments \mathbf{x} (as detailed shortly). This term is commonly known as *numeraire commodity* in the literature of mechanism design [11], as opposed to the *commodity of interest*, which are the security investments in our context. To illustrate the purpose of including this term in a user's utility function, note that externalities are defined as the side-effects of users' actions on one another, the costs or benefits of which are not accounted for when users pick their actions. A numeraire commodity is often used in problems involving externalities to bring such side-effects into strategic individuals' decision making process, a tactic referred to as "internalizing the externalities".

We make the following assumptions about the users:

Assumption 3: All users i are strategic, and choose their investment x_i in order to maximize their own utility function (1).

Assumption 4: The cost c_i and the functional form of $f_i(\cdot)$ are user i 's private information.

The Interdependent Security (IDS) game induced among these N strategic players is defined as the strategic game $(\{1, \dots, N\}, \{x_i \geq 0\}, \{u_i(\cdot)\})$. The *socially optimal* vector of security investments in this N user system is the vector \mathbf{x}^* maximizing the social welfare, as determined by the solution to the following centralized problem:

$$\begin{aligned} & \max_{(\mathbf{x}, \mathbf{t})} && \sum_{i=1}^N u_i(\mathbf{x}) \\ & \text{s.t.} && \sum_{i=1}^N t_i = 0, \quad \mathbf{x} \succeq 0. \\ & \equiv \min_{\mathbf{x}} && \sum_{i=1}^N g_i(\mathbf{x}) \\ & \text{s.t.} && \mathbf{x} \succeq 0. \end{aligned} \quad (2)$$

In other words, socially optimal solutions minimize the social cost $G(\mathbf{x}) := \sum_{i=1}^N g_i(\mathbf{x})$. By Assumption 2, there is a unique socially optimal investment profile \mathbf{x}^* for Problem (2). Also, due to Assumptions 3 and 4, there is no individual/user in the system with enough information to determine \mathbf{x}^* .

Accordingly, our goal is to find a mechanism, run by a manager/regulator, such that the induced interdependent security game has as its equilibrium the solution to the centralized problem (2) (also referred to as “implementing” the solution to (2)).

To determine the effort that users exert in an IDS game, with or without regulation (i.e., $t_i = 0$, $\forall i$), we will consider the vector of investments \mathbf{x} in a Nash equilibrium (NE) of the game $(\{1, \dots, N\}, \{x_i \geq 0\}, \{u_i(\cdot)\})$. Theoretically, Nash equilibria describe users’ actions in a game of complete information. However, due to Assumption 4, the model studied herein is one of incomplete information. The Nash equilibrium in this game can be interpreted as the convergence point of an iterative process, in which each user adjusts its action at each round based on its observations of other users’ actions, until unilateral deviations are no longer profitable [12], [7].²

A pure strategy Nash equilibrium of the IDS game is a vector of investments $\bar{\mathbf{x}}$, for which,

$$u_i(\bar{x}_i, \bar{\mathbf{x}}_{-i}) \geq u_i(x_i, \bar{\mathbf{x}}_{-i}), \quad \forall x_i \geq 0, \forall i. \quad (3)$$

We first ensure that the game studied indeed has a Nash equilibrium in the following result. The proof can be found in the Appendix.

Proposition 1: There always exists a pure strategy Nash equilibrium in an unregulated (i.e. $t_i = 0$, $\forall i$) IDS game modeled in this section.

III. PRICE OF ANARCHY IN AN UNREGULATED IDS GAME

Existence notwithstanding, the Nash equilibria of an unregulated IDS game are often inefficient. A common metric for quantifying the inefficiency of such equilibria is the *Price of Anarchy* (PoA), defined as the largest possible ratio between the worst possible social cost at a Nash equilibrium $\bar{\mathbf{x}}$ and at the social optimum \mathbf{x}^* . Formally, PoA ρ is defined as:

$$\begin{aligned} \rho &= \max_{\bar{\mathbf{x}}} \rho(\bar{\mathbf{x}}), \\ \rho(\bar{\mathbf{x}}) &:= \frac{G(\bar{\mathbf{x}})}{G(\mathbf{x}^*)} = \frac{\sum_{i=1}^N g_i(\bar{\mathbf{x}})}{\sum_{i=1}^N g_i(\mathbf{x}^*)}. \end{aligned} \quad (4)$$

In [7], the authors characterize the price of anarchy in an unregulated IDS game, i.e., the game in which no external mechanism is implemented. The NE of this game is defined in the same way as in (3), with $u_i(\cdot)$ replaced by $-g_i(\cdot)$. This means that without regulation, users selfishly pick effort levels that minimize their own cost. As a result, $\rho > 1$ for several plausible risk function models ([7, Lemma 1, Propositions 2, 3], reflecting under-investment in security. Below we present such an example, different from the aforementioned results presented in [7], and chosen for its simplicity.

²Alternatively, one may relax Assumption 4 and study a game of complete information, as is done in the majority of the current literature on IDS games.

Consider N interconnected users, and a *total effort* model for users’ risk function, such that

$$f_i(\mathbf{x}) = f\left(\sum_{j=1}^N x_j\right), \quad \forall i.$$

Furthermore, without loss of generality, assume $c_1 < c_2 < \dots < c_N$. At the Nash equilibrium of this game, each user will choose a level of investment $x_i \geq 0$ to minimize its own cost. Therefore, at the Nash equilibrium $\bar{\mathbf{x}}$ we must have:

$$\begin{aligned} \bar{x}_i = 0 & \quad \text{if} \quad \frac{\partial f(\bar{x}_i, \bar{\mathbf{x}}_{-i})}{\partial x_i} + c_i > 0, \\ \bar{x}_i > 0 & \quad \text{if} \quad \frac{\partial f(\bar{x}_i, \bar{\mathbf{x}}_{-i})}{\partial x_i} + c_i = 0. \end{aligned}$$

We conclude that only the user with the lowest cost will be exerting a non-zero effort at the Nash equilibrium $\bar{\mathbf{x}}$. Thus:

$$\partial f(\bar{x}_1, \mathbf{0}) / \partial x_1 = -c_1, \quad \text{and} \quad \bar{x}_j = 0, \quad \forall j > 1.$$

At the socially optimal equilibrium \mathbf{x}^* on the other hand, the levels of investment are determined by:

$$\begin{aligned} x_i^* = 0 & \quad \text{if} \quad N \frac{\partial f(x_i^*, \mathbf{x}_{-i}^*)}{\partial x_i} + c_i > 0, \\ x_i^* > 0 & \quad \text{if} \quad N \frac{\partial f(x_i^*, \mathbf{x}_{-i}^*)}{\partial x_i} + c_i = 0. \end{aligned}$$

Again the user with the lowest cost will be exerting all the effort at the equilibrium \mathbf{x}^* , however at a higher level, determined by:

$$\partial f(x_1^*, \mathbf{0}) / \partial x_1 = -c_1/N, \quad \text{and} \quad x_j^* = 0, \quad \forall j > 1.$$

The price of anarchy will therefore be given by:

$$\rho = \frac{N f(\bar{x}_1, \mathbf{0}) + c_1 \bar{x}_1}{N f(x_1^*, \mathbf{0}) + c_1 x_1^*}.$$

By the strict convexity of $f(\cdot)$, we have:

$$f(\bar{x}_1, \mathbf{0}) - f(x_1^*, \mathbf{0}) > \frac{\partial f(x_1^*, \mathbf{0})}{\partial x_1} (\bar{x}_1 - x_1^*).$$

Hence, $\rho > 1$. Figure 1 illustrates the levels of investment in both the socially optimal and the Nash equilibrium of this game. Based on fig. 1, it is easy to observe the under-investment in security in the Nash equilibrium of an unregulated game.

In the next section, we present a mechanism under which all Nash equilibria of the induced IDS game coincide with the socially optimal solution, i.e., we will have $\rho = 1$, closing the price of anarchy gap.

IV. A POSITIVE EXTERNALITY SECURITY INVESTMENT MECHANISM (PESIM)

In this section, we present a mechanism that implements the socially optimal solution to (2) in an informationally decentralized setting. This mechanism is adapted from [12], [13].

A decentralized mechanism is specified by a game form (\mathcal{M}, h) .

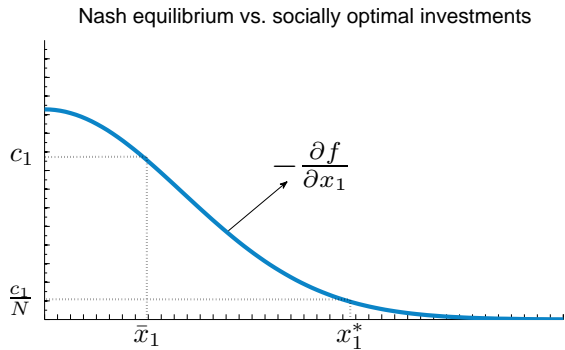


Fig. 1. Under-investment in security in an unregulated IDS game.

- The message space $\mathcal{M} := \prod_{i=1}^N \mathcal{M}_i$ specifies the set of permissible messages \mathcal{M}_i for each user i .
- The outcome function $h : \mathcal{M} \rightarrow \mathcal{A}$ determines the outcome of the game based on the users' messages. Here, \mathcal{A} is the space of all security investment profiles and tax profiles, i.e., (\mathbf{x}, \mathbf{t}) .

The game form, together with the utility functions, define a game, represented by $(\mathcal{M}, h(\cdot), \{u_i(\cdot)\})$. This will also be referred to as the regulated IDS game.

We say the message profile \mathbf{m}^* is a Nash equilibrium of this game, if

$$u_i(h(m_i^*, \mathbf{m}^*_{-i})) \geq u_i(h(m_i, \mathbf{m}^*_{-i})), \quad \forall m_i, \forall i. \quad (5)$$

The components of the proposed decentralized PESIM mechanism are specified as follows.

The Message Space: Each user i reports a message $m_i := (\pi_i, \mathbf{x}_i)$ to the regulator, with $\pi_i \in \mathbb{R}_+^N$ and $\mathbf{x}_i \in \mathbb{R}^N$. The component \mathbf{x}_i is user i 's proposal regarding the public good, i.e., the security investment profile, while π_i is user i 's suggestion regarding the private good, i.e., the price profile³.

The Outcome Function: The outcome function h takes the message profile \mathbf{m} as input and determines the security investment profile $\hat{\mathbf{x}}$ and the tax profile $\hat{\mathbf{t}}$ as follows:

$$\hat{\mathbf{x}}(\mathbf{m}) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \quad (6)$$

$$\begin{aligned} \hat{\mathbf{t}}_i(\mathbf{m}) &= (\pi_{i+1} - \pi_{i+2})^T \hat{\mathbf{x}}(\mathbf{m}) \\ &+ (\mathbf{x}_i - \mathbf{x}_{i+1})^T \text{diag}(\pi_i)(\mathbf{x}_i - \mathbf{x}_{i+1}) \\ &- (\mathbf{x}_{i+1} - \mathbf{x}_{i+2})^T \text{diag}(\pi_{i+1})(\mathbf{x}_{i+1} - \mathbf{x}_{i+2}), \forall i. \end{aligned} \quad (7)$$

In (7), for simplicity $N+1$ and $N+2$ are treated as 1 and 2, respectively. That is, $N+1$ denotes the modulo $(N \bmod 1)$, and so on.

This outcome function is interpreted as follows: first, (6) states that the contribution \hat{x}_i of each user i to the public good vector of investments $\hat{\mathbf{x}}$ is determined by the average of all users' proposals. The taxation term (7) is then used to

make sure that all investment profile proposals \mathbf{x}_i are the same at equilibrium, and are equal to the socially optimal security investments.

The tax term for user i itself consists of three different terms. The first term $(\pi_{i+1} - \pi_{i+2})^T \hat{\mathbf{x}}(\mathbf{m})$ is independent of user i 's proposal for prices, and depends only on the investment profile⁴. The second term determines the penalties for the discrepancy between user i 's proposal \mathbf{x}_i and user $(i+1)$'s proposal. This term will ensure eventual agreement between investment proposals put forward by different users. The third term does not depend on user i 's message, and is used only as a balancing term. In fact, at equilibrium, both the second and third terms will be equal to zero. Nevertheless, their inclusion is necessary to ensure convergence to the optimal security investment profile, and also for budget balance (i.e., the sum of all taxes equal zero) on and off the equilibrium. Note that having budget balance off equilibrium is an important property of the proposed mechanism, in order to prevent complications in an iterative message exchange process that leads to the desired Nash equilibrium.

We would also like to highlight the close relation between the tax term proposed in (7) and the positive externalities of users' actions. As illustrated later, at an equilibrium of the PESIM mechanism, the second and third terms in (7) disappear, so that the tax \hat{t}_i for user i reduces to $\hat{t}_i = \mathbf{l}_i^*{}^T \hat{\mathbf{x}}$, where $\mathbf{l}_i^* := \pi_{i+1}^* - \pi_{i+2}^*$ is known as the Lindhal price for user i . Furthermore, when users' monetary taxes are assessed according to Lindhal prices, the socially optimal investments \mathbf{x}^* will be individually optimal as well, i.e.,⁵

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \geq 0} g_i(\mathbf{x}) + \mathbf{l}_i^*{}^T \mathbf{x}. \quad (8)$$

As a result, it is easy to show that for all i , and all j for which $\hat{x}_j \neq 0$,

$$\frac{\partial g_i(\hat{\mathbf{x}})}{\partial x_j} < 0 \Rightarrow \mathbf{l}_{i,j}^* > 0 \Rightarrow \mathbf{l}_{i,j}^* \hat{x}_j^* > 0. \quad (9)$$

The interpretation of this observation is that by implementing the PESIM mechanism, user i will be paying a monetary tax to user j , which is proportional to the positive externality of j 's investment on user i 's costs (9).

It should be pointed out that for the time being, we have assumed users' participation in the mechanism is ensured, either through policy mandate (e.g., the government may require users to participate in the mechanism as a prerequisite for conducting business with it), or secondary financial incentive (e.g., product discount for joining the collection of users interested in the mechanism), such that the incentive for participation is separate from the mechanism itself. In Section V, we present a counter-example to illustrate why the individual rationality constraint, i.e., the condition that a user is better off by participating than staying out, may fail to hold, and discuss some implications of this observation.

³Note the use of the term *price profile* for the vectors π_i . As illustrated later, these terms are closely related to Lindhal prices, and will in turn be used to determine a *tax profile* \mathbf{t} .

⁴ $\pi_{i+1} - \pi_{i+2}$ is interpreted as the Lindhal price for the public good [13].

⁵See proof of Theorem 1 presented later in this section for the derivation of this result.

We close this section by presenting the theorems that establish the optimality of the proposed game form. Note that to prove this optimality, we first need to show that a profile $(\hat{\mathbf{x}}(\mathbf{m}^*), \hat{\mathbf{t}}(\mathbf{m}^*))$, derived at the NE \mathbf{m}^* of the induced game, is an optimal solution to the centralized problem (2), and therefore socially optimal. Furthermore, as the procedure for convergence to NE is not specified, we need to verify that the optimality property holds for all Nash equilibrium of the message exchange process. This guarantees that the outcome will converge to the socially optimal solution regardless of the realized NE. These two requirements are established in Theorem 1 below.

Theorem 1: Let $(\hat{\mathbf{x}}(\mathbf{m}^*), \hat{\mathbf{t}}(\mathbf{m}^*))$ be the investment and tax profiles obtained at the Nash equilibrium \mathbf{m}^* of the game $(\mathcal{M}, h(\cdot), \{u_i(\cdot)\})$. Then, $(\hat{\mathbf{x}}, \hat{\mathbf{t}})$ is an optimal solution to the centralized problem (2). Furthermore, if $\bar{\mathbf{m}}$ is any other Nash equilibrium of the proposed game, then $\hat{\mathbf{x}}(\bar{\mathbf{m}}) = \hat{\mathbf{x}}(\mathbf{m}^*)$.

Proof: Let \mathbf{m}^* be a Nash equilibrium of the message exchange process, resulting in an allocation $(\hat{\mathbf{x}}, \hat{\mathbf{t}})$. Assume user i updates its message from $m_i^* = (\pi_i^*, \mathbf{x}_i^*)$ to $m_i = (\pi_i, \mathbf{x}_i^*)$, that is, it only updates the price vector proposal. Therefore, according to (6), $\hat{\mathbf{x}}$ will remain fixed, while based on (7), the second term in \hat{t}_i will change. Since \mathbf{m}^* is an NE, unilateral deviations are not profitable. Mathematically,

$$\begin{aligned} & (\mathbf{x}_i^* - \mathbf{x}_{i+1}^*)^T \text{diag}(\pi_i^*)(\mathbf{x}_i^* - \mathbf{x}_{i+1}^*) \\ & \leq (\mathbf{x}_i^* - \mathbf{x}_{i+1}^*)^T \text{diag}(\pi_i)(\mathbf{x}_i^* - \mathbf{x}_{i+1}^*), \quad \forall \pi_i \geq 0. \end{aligned} \quad (10)$$

Hence, from (10) we conclude that for all i :

$$\mathbf{x}_i^* = \mathbf{x}_{i+1}^* \quad \text{or} \quad \pi_i^* = \mathbf{0}. \quad (11)$$

Using (11) together with (7) we conclude that at equilibrium, the second and third terms of a user's tax vanish. Denoting $\mathbf{l}_i^* := \pi_{i+1}^* - \pi_{i+2}^*$, we get:

$$\hat{\mathbf{t}}_i(\mathbf{m}^*) = \mathbf{l}_i^{*T} \hat{\mathbf{x}}(\mathbf{m}^*). \quad (12)$$

Now consider the utility function of the users at the Nash equilibrium \mathbf{m}^* . Since unilateral deviations are not profitable, a user's utility (1) should be maximized at the NE, i.e., for any choice of \mathbf{x}_i and $\pi_i \geq 0$:

$$\begin{aligned} & g_i(\hat{\mathbf{x}}(\mathbf{m}^*)) + \mathbf{l}_i^{*T} \hat{\mathbf{x}}(\mathbf{m}^*) \\ & \leq g_i\left(\frac{\mathbf{x}_i + \sum_{j \neq i} \mathbf{x}_j^*}{N}\right) + \mathbf{l}_i^{*T} \frac{\mathbf{x}_i + \sum_{j \neq i} \mathbf{x}_j^*}{N} \\ & \quad + (\mathbf{x}_i - \mathbf{x}_{i+1}^*)^T \text{diag}(\pi_i)(\mathbf{x}_i - \mathbf{x}_{i+1}^*) \end{aligned} \quad (13)$$

If we choose $\pi_i = \mathbf{0}$ and let $\mathbf{x}_i = N \cdot \mathbf{x} - \sum_{j \neq i} \mathbf{x}_j^*$, where \mathbf{x} is any vector of security investments, we get:

$$g_i(\hat{\mathbf{x}}(\mathbf{m}^*)) + \mathbf{l}_i^{*T} \hat{\mathbf{x}}(\mathbf{m}^*) \leq g_i(\mathbf{x}) + \mathbf{l}_i^{*T} \mathbf{x}, \quad \forall \mathbf{x}. \quad (14)$$

To show that the Nash equilibrium \mathbf{m}^* results in a socially optimal allocation, we sum up (14) over all i , and use the fact that $\sum_i \mathbf{l}_i^* = \mathbf{0}$ to get:

$$\sum_{i=1}^N g_i(\hat{\mathbf{x}}(\mathbf{m}^*)) \leq \sum_{i=1}^N g_i(\mathbf{x}), \quad \forall \mathbf{x}. \quad (15)$$

Therefore, $\hat{\mathbf{x}}(\mathbf{m}^*)$ is the optimal investment profile minimizing the social cost in problem (2). Furthermore, any tax profile \mathbf{t} satisfying the budget balance condition can be chosen as the tax profile in the optimal solution. Since the tax terms (12) are balanced, we conclude that $(\hat{\mathbf{x}}(\mathbf{m}^*), \hat{\mathbf{t}}(\mathbf{m}^*))$ solves (2) and is therefore socially optimal. Finally, since our choice of the NE \mathbf{m}^* has been arbitrary, the same proof holds for any other NE, and thus all NE of the mechanism result in the optimal solution to problem (2). ■

Finally, we establish the converse of this statement in Theorem 2, i.e., given an optimal investment profile, there exists an NE of the proposed game which implements this solution.

Theorem 2: Let \mathbf{x}^* be the optimal investment profile in the solution to the centralized problem (2). Then, there exists at least one Nash equilibrium \mathbf{m}^* of the game $(\mathcal{M}, h(\cdot), \{u_i(\cdot)\})$ such that $\hat{\mathbf{x}}(\mathbf{m}^*) = \mathbf{x}^*$.

The proof of this theorem is given in the appendix.

V. ON INDIVIDUAL RATIONALITY

Thus far, we have assumed user participation in the message exchange process is ensured using external incentive mechanisms. Alternatively, one could try to guarantee voluntary participation of strategic users by establishing that the so-called *individual rationality* condition is satisfied, i.e., users gain when participating in the mechanism as opposed to staying out.

Whether a mechanism is individually rational depends on the structure of the game form, as well as the actions available to users when opting out. A common assumption in the majority of public good and resource allocation problems, including the prior work on the decentralized mechanism presented in Section IV ([13], [14], [12]), is that users will get a zero share (of the public good or allotted resources) when staying out. Following this assumption, [13], [14], [12] establish the individual rationality of the presented mechanism. However, a similar line of reasoning is not applicable to the current problem.

The different nature of individual rationality in an IDS game can be intuitively explained as follows. By implementing a socially optimal equilibrium, (some) users will be required to increase their level of investment in security. In turn, the mechanism should either guarantee that these users enjoy a higher level of protection due to higher equilibrium investments from other participants, and/or are adequately compensated for their contribution by a monetary reward (negative taxation). On the other hand, by staying out, a user can still enjoy the positive externalities of other users' investments (although these may be lower when the mechanism has partial coverage), choose its optimal action accordingly, and possibly avoid taxation. Thus to establish individual rationality in such an IDS game is not nearly as trivial as in previous studies.

Indeed, the following counter-example shows that the benefits of staying out can overthrow that of participation, making a user better off when acting as a "loner".

Specifically, a loner is a user who refuses to participate in the mechanism, and later best-responds to the socially optimal strategy of the remaining $N - 1$ users who did participate. Arguably, these $N - 1$ users could also revise their strategy (investments) in response to this loner's best-response, leading to a sequential game. In this example we will compare the loner's utility in the socially optimal solution when participating in the mechanism, versus the utility at the outcome of the sequential game described above.

Consider a collection of N users. Without loss of generality, assume $c_1 < c_2 < \dots < c_N$. Assume user 1 is contemplating whether to participate or remain a free agent. We further assume all users have the same risk function $f_i(\mathbf{x}) = \exp(-\sum_{i=1}^N x_i)$ (an instance of the total effort model [2]).

It is easy to show that at the socially optimal solution \mathbf{x}^* to the N -player game, the user with the smallest cost would exert all the effort (see e.g. Section III, or [2]), such that:

$$\exp(-x_1^*) = c_1/N, \quad x_j^* = 0, \quad \forall j > 1.$$

By (12) in the proof of Theorem 1, the tax for user 1 is given by:

$$t_1^* = \mathbf{1}_1^{*T} \mathbf{x}^* = l_{11}^* x_1^*.$$

Re-writing (14) in the proof of Theorem 1 as

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \geq 0} g_1(\mathbf{x}) + \mathbf{1}_1^{*T} \mathbf{x},$$

and applying the KKT conditions, we conclude that:

$$\begin{aligned} l_{11}^* + \frac{\partial g_1}{\partial x_1}(\mathbf{x}^*) &= l_{11}^* - \exp(-x_1^*) + c_1 = 0 \\ \Rightarrow l_{11}^* &= -(1 - \frac{1}{N})c_1 \Rightarrow t_1^* = -(1 - \frac{1}{N})c_1 x_1^*. \end{aligned}$$

As expected, user 1 is getting a reward in this mechanism.

Now assume user 1 opts out of the decentralized mechanism. The remaining $N - 1$ users choose their strategies assuming user 1 exerts an effort of x_1 . Then, by the nature of the total effort game, the user with the smallest cost among these $N - 1$ players will exert all the effort (if any) such that:

$$\exp(-x_1 - \hat{x}_2) = c_2/(N - 1), \quad \hat{x}_j = 0, \quad \forall j > 2.$$

On the other hand, if user 1 is best responding to a choice of x_2 , it chooses an effort according to:

$$\exp(-\hat{x}_1 - x_2) = c_1.$$

Combining the last two equations, at an equilibrium $\hat{\mathbf{x}}$ of the sequential game we have:

$$\hat{x}_1 = \arg \min_{x_1 \geq 0} \exp(-x_1 - \max\{-\ln \frac{c_2}{N-1} - x_1, 0\}) + c_1 x_1.$$

From the above, we conclude that if $-\ln \frac{c_2}{N-1}$ is large enough, that is, if without user 1's participation, user 2 will exert a sufficiently high effort, user 1 will choose to free-ride. Otherwise, it may again exert all the effort, in which case $\exp(-\hat{x}_1) = c_1$.

Let us focus on this latter case. It is interesting to note that the overall level of security in the sequential game is lower than the coordinated socially optimal equilibrium.

We compare user 1's utility under the two scenarios.

$$\begin{aligned} u_1^{IN}(\mathbf{x}^*) &= -\exp(-x_1^*) - c_1 x_1^* + (1 - \frac{1}{N})c_1 x_1^*. \\ u_1^{OUT}(\hat{\mathbf{x}}) &= -\exp(-\hat{x}_1) - c_1 \hat{x}_1. \end{aligned}$$

Therefore,

$$\begin{aligned} u_1^{IN} - u_1^{OUT} &= -(\exp(-x_1^*) - \exp(-\hat{x}_1)) \\ &\quad - c_1(x_1^* - \hat{x}_1) + (1 - \frac{1}{N})c_1 x_1^* \\ &= -(\frac{c_1}{N} - c_1) - c_1(-\ln \frac{c_1}{N} + \ln c_1) \\ &\quad + (1 - \frac{1}{N})c_1(-\ln \frac{c_1}{N}) \\ &= \frac{c_1}{N} \left((N-1)(1 - \ln c_1) - \ln N \right). \end{aligned} \quad (16)$$

Based on (16), with any cost $c_1 \geq \exp(1)$, user 1's utility will decrease when participating, indicating that in this case the decentralized mechanism fails to satisfy individual rationality.

In light of the above observation, we conclude that although the proposed mechanism is incentive compatible and implements the socially optimal levels of investment in a Nash equilibrium, it fails to satisfy individual rationality in general. It remains an interesting question whether there are other mechanisms which would satisfy all requirements simultaneously, or alternatively whether this is a more fundamental challenge in designing mechanisms for resource allocation with positive externalities. The answer should shed light on questions such as whether security policies should be mandated (or alternatively incentivized), rather than being left to users' free will.

VI. CONCLUSION

In this paper, we have presented a decentralized mechanism, through which we can find and implement the socially optimal levels of investment in security in an interdependent security game. This mechanism is especially attractive as it is applicable to a wide range of user preferences, operates without the need for collecting information about these preferences, and does not need to centrally dictate the socially optimal outcome. We further consider the issue of individual rationality, often a trivial condition to satisfy in many resource allocation problems. We provide a counter example under the proposed mechanism, and argue that with positive externality, the incentive to stay out and free-ride on others' investment can make individual rationality much harder to satisfy in designing a mechanism.

The study of IDS games in the current framework can be further continued in several directions. First, the procedure and conditions under which the message exchange process converges to a Nash equilibrium remains an open problem, and is an interesting direction of future study. Alternatively, one could switch focus to Bayesian Nash equilibrium as the

solution concept for games of incomplete information, to better capture the uncertainty of users about their environment, including other users' valuations of security and the resources available to them. It is also interesting to study how the information obtained from alternative resources, e.g. IP blacklists, can help users attain a better understanding of their security risks and consequently make more effective investment decisions.

APPENDIX

In this appendix, we present the proofs to Proposition 1 and Theorem 2. The proof for Theorem 2 is technically similar to that presented in [12], [13], and the proof of Proposition 1 follows from [7, Proposition 1].

Proof of Proposition 1

We first show that the strategy space $x_i \in [0, \infty)$ of a user i can be effectively reduced to a convex and compact set.

Let $\text{BR}_i(\mathbf{x}_{-i})$ represent user i 's best response to the strategies $\mathbf{x}_{-i} \succeq 0$ of all the other users. Define $\hat{x}_i = \frac{f_i(\mathbf{0}) + \epsilon}{c_i}$, for some $\epsilon > 0$. By assumption 2, the functions $f_i(\cdot)$ are convex, and thus:

$$\begin{aligned} f_i(0, \mathbf{x}_{-i}) - f_i(\hat{x}_i, \mathbf{x}_{-i}) &\geq -\hat{x}_i \frac{\partial f_i(\hat{x}_i, \mathbf{x}_{-i})}{\partial x_i} \\ &= -\frac{f_i(\mathbf{0}) + \epsilon}{c_i} \frac{\partial f_i(\hat{x}_i, \mathbf{x}_{-i})}{\partial x_i} \end{aligned} \quad (17)$$

By assumption 1, $f_i(\hat{x}_i, \mathbf{x}_{-i}) \geq 0$, and $f_i(0, \mathbf{x}_{-i}) \leq f_i(\mathbf{0})$. Therefore, (17) reduces to:

$$f_i(\mathbf{0}) \geq -\frac{f_i(\mathbf{0}) + \epsilon}{c_i} \frac{\partial f_i(\hat{x}_i, \mathbf{x}_{-i})}{\partial x_i}. \quad (18)$$

Equation (18) in turn implies that $\frac{\partial f_i(\hat{x}_i, \mathbf{x}_{-i})}{\partial x_i} + c_i > 0$. Therefore, since user i 's cost is increasing at \hat{x}_i , a best response to minimize the cost should be such that $\text{BR}_i(\mathbf{x}_{-i}) \in [0, \hat{x}_i]$. Let $x_{max} := \max_i \hat{x}_i$. We conclude that for all i , the strategy sets can be effectively reduced to $x_i \in [0, x_{max}]$.

Since the strategy sets are non-empty, compact, and convex, and as the utility functions (1) are continuous and concave in x_i , the unregulated IDS game will always have at least one Nash equilibrium ([11, Proposition 8.D.3]). ■

Proof of Theorem 2

Consider the optimal security investment profile \mathbf{x}^* in the solution to the centralized problem (2). Our goal is to show that there indeed exists a Nash equilibrium \mathbf{m}^* of the mechanism for which $\hat{\mathbf{x}}(\mathbf{m}^*) = \mathbf{x}^*$.

We start by showing that given the investment profile \mathbf{x}^* , it is possible to find a vector of personalized (Lindhal) prices \mathbf{l}_i^* , for each i , such that,

$$\arg \min_{\mathbf{x} \succeq 0} g_i(\mathbf{x}) + \mathbf{l}_i^{*T} \mathbf{x} = \mathbf{x}^*. \quad (19)$$

First, we know that since \mathbf{x}^* is the solution to problem (2), it should satisfy the following KKT conditions, where $\boldsymbol{\lambda}_i \in$

\mathbb{R}_+^N , $\forall i$:

$$\begin{aligned} \sum_{i=1}^N (\nabla g_i(\mathbf{x}^*) - \boldsymbol{\lambda}_i^T) &= \mathbf{0}, \\ \boldsymbol{\lambda}_i^T \mathbf{x}^* &= 0 \quad \forall i. \end{aligned} \quad (20)$$

Choose $\mathbf{l}_i^* = -\nabla g_i(\mathbf{x}^*) + \boldsymbol{\lambda}_i^T$. Then,

$$\mathbf{l}_i^* + \nabla g_i(\mathbf{x}^*) - \boldsymbol{\lambda}_i^T = \mathbf{0}. \quad (21)$$

Equations (20) and (21) together are the KKT conditions for the convex optimization problem:

$$\min_{\mathbf{x} \succeq 0} g_i(\mathbf{x}) + \mathbf{l}_i^{*T} \mathbf{x}. \quad (22)$$

The KKT conditions are necessary and sufficient for finding the optimal solution to the convex optimization problem (22), and thus we have found the personalized prices satisfying (19).

We now proceed to finding a Nash equilibrium \mathbf{m}^* implementing the socially optimal solution \mathbf{x}^* . Consider the message profiles $\mathbf{m}_i^* = (\boldsymbol{\pi}_i^*, \mathbf{x}_i^*)$, for which $\mathbf{x}_i^* = \mathbf{x}^*$, and the price vector proposals $\boldsymbol{\pi}_i^*$ are found from the recursive equations:

$$\boldsymbol{\pi}_{i+1}^* - \boldsymbol{\pi}_{i+2}^* = \mathbf{l}_i^*, \quad \forall i. \quad (23)$$

Here, \mathbf{l}_i^* are the personalized prices defined at the beginning of the proof. The set of equations (23) always has a non-negative set of solutions $\boldsymbol{\pi}_i^* \succeq 0$, $\forall i$. This is because starting with a large enough $\boldsymbol{\pi}_1^*$, the remaining $\boldsymbol{\pi}_i^*$ can be determined using:⁶

$$\boldsymbol{\pi}_i^* = \boldsymbol{\pi}_{i-1}^* - \mathbf{l}_{i-1}^*, \quad \forall i \geq 2. \quad (24)$$

Now, first note that by (22), for all choices of $\mathbf{x} \succeq 0$, and all users i , we have:

$$g_i(\mathbf{x}^*) + \mathbf{l}_i^{*T} \mathbf{x}^* \leq g_i(\mathbf{x}) + \mathbf{l}_i^{*T} \mathbf{x}. \quad (25)$$

Particularly, if we pick $\mathbf{x} = \frac{\mathbf{x}_i + \sum_{j \neq i} \mathbf{x}_j^*}{N}$,

$$\begin{aligned} g_i(\mathbf{x}^*) + \mathbf{l}_i^{*T} \mathbf{x}^* \\ \leq g_i\left(\frac{\mathbf{x}_i + \sum_{j \neq i} \mathbf{x}_j^*}{N}\right) + \mathbf{l}_i^{*T} \frac{\mathbf{x}_i + \sum_{j \neq i} \mathbf{x}_j^*}{N}. \end{aligned} \quad (26)$$

Also, since by construction $\mathbf{x}_i^* = \mathbf{x}_{i+1}^*$, $\forall i$, the inequality is preserved for any choice of $\boldsymbol{\pi}_i \succeq 0$, when the two additional tax terms are added in as follows:

$$\begin{aligned} g_i(\mathbf{x}^*) + \mathbf{l}_i^{*T} \mathbf{x}^* + (\mathbf{x}_i^* - \mathbf{x}_{i+1}^*)^T \text{diag}(\boldsymbol{\pi}_i^*)(\mathbf{x}_i^* - \mathbf{x}_{i+1}^*) \\ - (\mathbf{x}_{i+1}^* - \mathbf{x}_{i+2}^*)^T \text{diag}(\boldsymbol{\pi}_{i+1}^*)(\mathbf{x}_{i+1}^* - \mathbf{x}_{i+2}^*) \\ \leq g_i\left(\frac{\mathbf{x}_i + \sum_{j \neq i} \mathbf{x}_j^*}{N}\right) + \mathbf{l}_i^{*T} \frac{\mathbf{x}_i + \sum_{j \neq i} \mathbf{x}_j^*}{N} \\ + (\mathbf{x}_i - \mathbf{x}_{i+1}^*)^T \text{diag}(\boldsymbol{\pi}_i)(\mathbf{x}_i - \mathbf{x}_{i+1}^*) \\ - (\mathbf{x}_{i+1}^* - \mathbf{x}_{i+2}^*)^T \text{diag}(\boldsymbol{\pi}_{i+1}^*)(\mathbf{x}_{i+1}^* - \mathbf{x}_{i+2}^*). \end{aligned} \quad (27)$$

Equation (27) can be more concisely written as:

$$\begin{aligned} u_i(h(\mathbf{m}_i^*, \mathbf{m}_{-i}^*)) &\geq u_i(h(\mathbf{m}_i, \mathbf{m}_{-i}^*)) , \\ \forall \mathbf{m}_i &= (\boldsymbol{\pi}_i, \mathbf{x}_i), \quad \forall i. \end{aligned} \quad (28)$$

⁶In (24), \mathbf{l}_0^* is interpreted as \mathbf{l}_N^* .

We conclude that the messages $\mathbf{m}_i^* = (\pi_i^*, \mathbf{x}^*)$ constitute an NE of the proposed mechanism. In other words, the message exchange process will indeed have an NE which implements the socially optimal solution of problem (2). ■

REFERENCES

- [1] H. Kunreuther and G. Heal, "Interdependent security," *Journal of Risk and Uncertainty*, vol. 26, no. 2-3, pp. 231–249, 2003.
- [2] H. Varian, "System reliability and free riding," *Economics of information security*, pp. 1–15, 2004.
- [3] M. Parameswaran, X. Zhao, A. B. Whinston, and F. Fang, "Reengineering the internet for better security," *Computer*, vol. 40, no. 1, pp. 40–44, 2007.
- [4] J. Grossklags, N. Christin, and J. Chuang, "Secure or insure?: a game-theoretic analysis of information security games," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 209–218.
- [5] J. Grossklags, S. Radosavac, A. A. Cárdenas, and J. Chuang, "Nudge: Intermediaries role in interdependent network security," in *Trust and Trustworthy Computing*. Springer, 2010, pp. 323–336.
- [6] M. Lelarge, "Economics of malware: Epidemic risks model, network externalities and incentives," in *47th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2009, pp. 1353–1360.
- [7] L. Jiang, V. Anantharam, and J. Walrand, "How bad are selfish investments in network security?" *IEEE/ACM Transactions on Networking*, vol. 19, no. 2, pp. 549–560, 2011.
- [8] A. Laszka, M. Felegyhazi, and L. Buttyán, "A survey of interdependent security games," *CRYSYS*, vol. 2, 2012.
- [9] R. Pal and L. Golubchik, "Analyzing self-defense investments in internet security under cyber-insurance coverage," in *IEEE 30th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2010, pp. 339–347.
- [10] (2012, October) Strategies to mitigate targeted cyber intrusions. [Online]. Available: <http://www.dsd.gov.au/infosec/top-mitigations/top35mitigationstrategies-list.htm>
- [11] A. Mas-Colell, M. D. Whinston, J. R. Green *et al.*, *Microeconomic theory*. Oxford university press New York, 1995, vol. 1.
- [12] S. Sharma and D. Teneketzis, "A game-theoretic approach to decentralized optimal power allocation for cellular networks," *Telecommunication Systems*, vol. 47, no. 1-2, pp. 65–80, 2011.
- [13] L. Hurwicz, "Outcome functions yielding walrasian and lindahl allocations at nash equilibrium points," *The Review of Economic Studies*, vol. 46, no. 2, pp. 217–225, 1979.
- [14] S. Sharma and D. Teneketzis, "A game-theoretic approach to decentralized optimal power allocation for cellular networks," in *Proceedings of the 3rd International Conference on Performance Evaluation Methodologies and Tools*.