TASHAROK: Using Mechanism Design for Enhancing Security Resource Allocation in Interdependent Systems

Mustafa Abdallah[1], Daniel Woods[2], Parinaz Naghizadeh[3], Issa Khalil[4],
Timothy Cason[1], Shreyas Sundaram[1], and Saurabh Bagchi[1]
[1]*Purdue University, West Lafayette, Indiana, USA* [2]*University of Innsbruck, Innsbruck, Austria*
[3]*Ohio State University, Columbus, Ohio, USA* [4]*Qatar Computing Research Institute, Doha, Qatar*
{abdalla0, cason, sundara2, sbagchi}@purdue.edu, daniel.j.woods@uibk.ac.at, naghizadeh.1@osu.edu, ikhalil@hbku.edu.qa

2022 IEEE Symposium on Security and Privacy (SP) | 978-1-6654-1316-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/SP46214.2022.9833591

***Abstract*—We consider interdependent systems managed by multiple defenders that are under the threat of stepping-stone attacks. We model such systems via game-theoretic models and incorporate the effect of behavioral probability weighting that is used to model biases in human decision-making, as descended from the field of behavioral economics. We then incorporate into our framework called** TASHAROK**, two types of tax-based mechanisms for such interdependent security games where the central regulator incentivizes defenders to invest well in securing their assets so as to achieve the socially optimal outcome. We first show that due to the nature of our interdependent security game, no reliable tax-based mechanism can incentivize the socially optimal investment profile while maintaining a weakly balanced budget. We then show the effect of behavioral probability weighting bias on the amount of taxes paid by defenders, and prove that higher biases make defenders pay more taxes under the two mechanisms. We then explore voluntary participation in tax-based mechanisms. To evaluate our mechanisms, we use four representative real-world interdependent systems where we compare the game-theoretic optimal investments to the socially optimal investments under the two mechanisms. We show that the mechanisms yield higher decrease in the social cost for behavioral decision-makers compared to rational decision-makers.**

***Index Terms*—Behavioral decision-making; Security games; Mechanism design; Interdependent systems; Attack graphs.**

I. Introduction

Today's interdependent systems face sophisticated attacks from external adversaries where the attacker aims to breach specific critical assets within the system [1], [2]. Such attacks pose a serious danger to large-scale critical infrastructure (e.g., the massive supply chain attack on SolarWinds in 2020 [3] and the recent Colonial Pipeline ransomware attack in May 2021 [4]). Such attacks have motivated several attempts to improve the cyber security of these systems [1], [5], [6], [7], [8]. Several challenges exist for these improvements. System operators often have limited security budgets that they need to allocate wisely within the systems they manage to reduce security risks. Moreover, coordination in large-scale systems that are composed of many interdependent subsystems managed by different operators is challenging as each operator has her local priority of securing her own subsystem.

Prior work has considered such security decision-making problems in both decision-theoretic [9], [10], [11], [12] and game-theoretic settings [13], [14], [15], [16]. The crux of the problem is that the security risk (usually captured by the probability of successful attack on system's critical assets) faced by a defender depends on her own security investments as well as those of other defenders securing subsystems that are interdependent with her subsystem. However, most existing work has relied on *classical models* of decision-making, where all defenders and attackers are assumed to be fully rational decision-makers [13], [16], [17].

In contrast, behavioral economics has shown that humans consistently deviate from these classical models of decision-making. Most notably, research in *behavioral economics* has shown that humans perceive gains, losses and probabilities in a skewed, nonlinear manner [18]. In particular, humans typically overweight low probabilities and underweight high probabilities, where this weighting function has an inverse S-shape, as shown later in Figure 2. Many empirical studies (e.g., [19], [18], [20]) have provided evidence for this class of behavioral models. These effects are relevant for evaluating the security of interdependent systems in which decisions on implementing security controls are *not* made purely by automated algorithms, but rather through human decision-making, albeit with help from threat assessment tools [12], [21]. The prevalence of human factors in security decision-making has been recognized through popular press articles [22], [23] and in academic studies [24], [25], [26]. Recent research has also shown that cybersecurity professionals' risk perceptions are as susceptible to systematic biases as those of the general population [27], [28] and related behavior of system-administrators securing systems [29].

Recent work has begun to model and predict the effect of behavioral decision-making on security investments [25], [30], [31], [32], [33]. However, none of this research sheds light on the mitigation of such cognitive biases and these works have only studied specific interdependent games. In contrast, we consider general defense allocation techniques that can be applied to any system whose failure scenarios are modeled by an attack graph. We consider tax-based mechanisms to guide behavioral decision-makers towards enhancing their security investments and incentivize them to achieve socially

DOI 10.1109/SP46214.2022.00106

optimal allocations that reduce the overall security risk. Fundamentally, our framework, TASHAROK[1], identifies the effects of behavioral bias on the design of mechanisms for improving security decisions in interdependent systems.

Throughout our paper, we consider two classes of defenders.
Behavioral defenders: These defenders make security investment decisions subject to the nonlinear probability weighting cognitive bias, found in prospect-theoretic models. They misperceive the probabilities of a successful attack on each edge of the attack graph of the system network.
Non-behavioral (rational) defenders: These defenders make security investment decisions based on the classical models of fully rational decision-making. Thus, they correctly perceive the probability on each edge within the attack graph.
Problem setup and mechanism design:
In this paper, we model a security setup of interdependent systems with multiple defenders. Each defender is responsible for defending a subnetwork of the whole network. In such interdependent systems, stepping-stone attacks are often used by external attackers to exploit vulnerabilities within the system in order to compromise critical targets. These stepping-stone attacks are captured via *attack graphs* [34].

We first show the difference between the Pure-Strategy Nash Equilibrium (PNE)[2] investments (by both rational and behavioral defenders) and the socially optimal investments via multiple motivating examples. We then design two tax-based mechanisms that enhance security investment decision-making for our interdependent security games. Such mechanisms use monetary payments/rewards to incentivize socially optimal (SO) security behavior, i.e., those minimizing the sum of the costs of all defenders due to a security attack. The two tax-based mechanisms are the 'Externality' mechanism [35] and the Vickrey-Clarke-Groves ('VCG') mechanism [36]. These mechanisms enhance the implemented security policy by incentivizing defenders to allocate their limited security resources to minimize the system's social cost.

We then show a fundamental result that there exists no reliable tax-based mechanism which can incentivize the socially optimal investment profile while maintaining a weakly balanced budget (i.e., the central regulator does not pay out-of-pocket money) for all instances of interdependent security games. We show the difference between our result and prior results in the security economics literature [36], [37] in Section VII. Our result shows that designing mechanisms in interdependent security games is more challenging compared to monolithic systems. We also show the effect of behavioral biases on the two mechanisms' outcomes in our interdependent security games framework.

We then evaluate our findings using four synthesized attack graphs that represent realistic interdependent systems and attack paths through them. These systems are DER.1 [21] (modeled by NESCOR), SCADA industrial control system modeled using NIST guidelines for ICS [16], E-commerce [17], and VOIP [17]. We do a benchmark comparison with four prior solutions for optimal security controls with attack graphs [25], [12], [38], [9]. In conducting our analysis, we address several domain-specific challenges in the context of security for interdependent systems. These include augmenting the attack graph with certain parameters such as edges' sensitivity to security investments (Equation 2), estimation of baseline attack probabilities (Table II), modifying mechanism formulations for our interdependent security games (Section V), and incorporating behavioral biases in our formulations (Section II).
Key insights:
Abstracting from the details, we provide three hitherto unknown insights into the security of interdependent systems.

1) A social planner (e.g., government agency) can achieve much lower security loss than each defender acting on her own. The difference increases when security defenders have more cognitive biases (Figure 9). The global planning is beneficial even if the planner is behavioral (Example 1). However, if the degree of interdependency is slight, then there is no need to go to the complexity of setting up central regulation — each defender acting independently (selfishly) achieves close to the optimal security (Figure 10(a)).
2) Our work supports recent proposals for companies to buy cyber insurance as part of their risk management strategy. In such process, the company would pay a tax (determined by the regulator depending on the system architecture (Figure 12)) and then transfer the financial risks related to network and computer incidents to that regulator.
3) Behavioral decision-making leads to suboptimal resource allocation and thus tax-based mechanisms can be more helpful in a system with behavioral defenders compared to non-behavioral (rational) decision-makers (Figure 11). In such mechanisms, we prove that behavioral biases make defenders pay *more* taxes compared to rational defenders.

In summary, this paper makes the following contributions:

1) We propose a *security investment guiding* technique for defenders of interdependent systems whose assets have mutual interdependencies. We show the effect of an important behavioral bias of human decision-making and selfishness of PNE decision-making on system security.
2) We consider two mechanism designs for interdependent security games modeled by attack graphs to guide decision-makers toward the socially optimal solution. In contrast to excludable public good games, we show that a weak budget balance condition is not guaranteed for all instances of interdependent security games.
3) We explore the voluntary participation in tax-based mechanisms and show that behavioral defenders participate under higher tax payments, compared to rational defenders.
4) We illustrate the benefits of our mechanisms through four real-world interdependent systems and analyze the different system parameters and the effect of behavioral decision-making on the mechanisms' outcomes and the overall security of these interdependent systems.

[1]TASHAROK is a word in Arabic that denotes several persons collaborating while carrying equal responsibilities for the sake of successful trade.

[2]A profile of security investments by defenders is said to be a PNE if no defender can decrease her cost by unilaterally changing her investment.

II. Background and Problem Setup

We now present a background on interdependent security games, establishing a theoretical basis that can be used to model multi-defender interdependent systems. These defenders can be different divisions within a large company, or different sectors of a country's economy. We formally define an interdependent system as "the system that has multiple defenders where each defender is responsible for defending a subnetwork of the whole network. In that system, there are dependencies among assets of different subnetworks which are captured via directed acyclic graph (DAG)". Figure 1 shows a simple example of our setup, which represents a system consisting of 3 interdependent defenders. An external attacker aims to exploit vulnerabilities within the network in order to compromise critical targets. We now formalize the attacker and defenders' goals and actions. The formulation in Sections II-A and II-B is well accepted in the security literature of attack graphs [12], [9], [39], [40]. We start discussing the notion of behavioral defenders from Section II-C where we tread ground not often trodden in the security literature, with some exceptions [25], [41], [42].

A. Threat Model

We consider security games consisting of one attacker and multiple defenders interacting through a directed acyclic attack graph $G = (V, \mathcal{E})$. The nodes V of the attack graph represent the assets in the system, while the edges $\mathcal{E}$ capture the attack progression between the assets. In particular, an edge from v_i to v_j, $(v_i, v_j) \in \mathcal{E}$, indicates that if asset v_i is compromised by the attacker, it can be used as a stepping-stone to launch an attack on asset v_j (e.g., if an attacker gains the password required to access a power plant's control software (v_i), it can use it to attempt to alter the operation of a generator (v_j)). We denote the baseline probability that the attacker can successfully compromise v_j given that it has compromised v_i, by the edge weight $p^0_{i,j} \in [0, 1]$. By "baseline probability" we mean the probability of successful compromise without any security investment in protecting the assets.[3] Suppose that the set of all defenders is given by $D = \{D_1, \ldots, D_k, \ldots, D_{|D|}\}$. The attacker initiates attacks on the network from a source node v_s (or multiple possible source nodes), and aims to reach a target node v_m, i.e., a critical node for any defender $D_k \in D$.

B. Defense Model

Each defender $D_k \in D$ is in control of a subset of assets $V_k \subseteq V$. Among all assets in the network, a subset $V_m \subseteq V$ are *critical* assets, the compromise of which entails a financial loss for the corresponding defender. Specifically, if asset $v_m \in V_m$ is compromised by the attacker, any defender D_k with $v_m \in V_k$ suffers a financial loss $L_m \in \mathbb{R}_{>0}$. Note that the critical assets of different defenders can be overlapping if they share common critical assets (e.g., the SCADA system in Figure 6). We emphasize that different critical assets can have heterogeneous loss valuations (Section VI).

[3]We emphasize that $p^0_{i,j}$ can also represent the pre-existing (inherent) security investments on the edge (v_i, v_j) (e.g., old software patched).

To protect the critical assets from being reached through stepping-stone attacks, the defenders can choose to invest their resources in strengthening the security of the edges in the network. Specifically, let $x^k_{i,j}$ denote the non-negative real investment of a defender D_k on edge $(v_i, v_j) \in \mathcal{E}_k$ (it suffices for an edge to belong to $\mathcal{E}_k$ if it belongs to at least one attack path from the source node v_s to one critical asset $v_m \in V_k$), and let $x_{i,j} = \sum_{D_k \in D} x^k_{i,j}$ be the total investment on that edge by all eligible defenders. Then, the probability of successfully compromising v_j starting from v_i is given by $p_{i,j}(x_{i,j})$. In addition, let $s_{i,j} \in [1, \infty)$ denote the sensitivity of edge (v_i, v_j) to the total investment $x_{i,j}$. For larger sensitivity values, the probability of successful attack on the edge decreases faster with each additional unit of security investment on that edge; in other words, edges that are easier to defend will have larger sensitivity.

Let P_m be the set of all attack paths from v_s to v_m. The defender assumes the worst-case scenario, i.e., the attackerexploits the most vulnerable path to each target. Note that previous works considered such an adversary model that chooses the most vulnerable path to target assets (e.g., [16], [13]). Mathematically, this can be captured via the following total loss function for D_k:

$$\hat{C}_k(\mathbf{x}) = \sum_{v_m \in V_k} L_m \Big(\max_{P \in P_m} \prod_{(v_i, v_j) \in P} p_{i,j}(x_{i,j}) \Big) . \quad (1)$$

In the above cost function in (1), we assume that the defense cost is negligible with respect to the huge financial cost under successful attack. We let the probability of successfully compromising v_j starting from v_i be given by,

$$p_{i,j}(x_{i,j}) = p^0_{i,j} \exp\Big(- s_{i,j}\ x_{i,j} \Big). \quad (2)$$

That is, the probability of successful attack on an edge (v_i, v_j) decreases exponentially with the sum of the investments on that edge by all defenders. This probability function falls within a class commonly considered in security economics [43], [10], [16], [32]. Note that (2) is a log-convex function.

C. Behavioral Probability Weighting

As mentioned in the introduction, the behavioral economics literature has shown that humans consistently misperceive probabilities by overweighting low probabilities, and underweighting high probabilities [18], [44]. More specifically, humans perceive a "true" probability p as probability $w(p)$, where $w(\cdot)$ is known as a *probability weighting function*. A commonly studied form for this weighting function was formulated by Prelec in [44], shown in Figure 2, given by

$$w(p) = \exp\Big[- (-\log(p))^\alpha \Big], \quad p \in [0, 1], \quad (3)$$

where $\alpha \in (0, 1]$ controls the extent of misperception. When $\alpha = 1$, we have $w(p) = p$ for all $p \in [0, 1]$, which corresponds to correct perception of probabilities, i.e., a non-behavioral (rational) defender (agent).

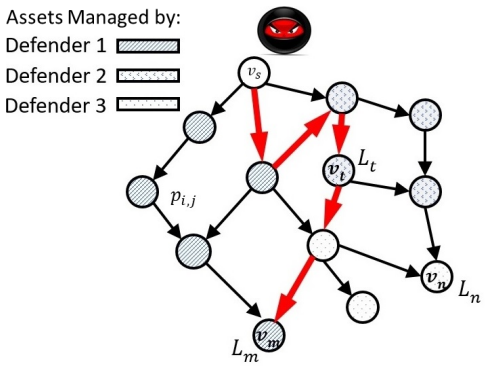


Fig. 1: An overview of the interdependent security framework. The interdependencies between assets are represented by edges. An attacker tries to compromise critical assets using stepping-stone attacks starting from v_s. The bold (red) edges show one such attack path.

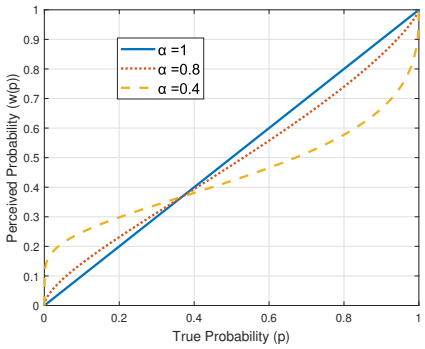


Fig. 2: The Prelec probability weighting function. The parameter $\alpha \in (0, 1]$ controls the extent of overweighting and underweighting, with $\alpha = 1$ indicating non-behavioral or rational decision-making. The smaller the value of α the more behavioral the action is (see dotted curves in the figure).

D. Perceived Cost of a Behavioral Defender

We now incorporate this probability weighting function into the interdependent security game defined in Section II-B. In a *Behavioral Security Game*, each defender misperceives attack success probability on each edge according to the probability weighting function in (3). She then chooses her investments $x_k := \{x^k_{i,j}\}_{(v_i,v_j)\in\mathcal{E}_k}$ to minimize her *perceived* loss

$$C_k(x_k, \mathbf{x}_{-k}) = \sum_{v_m \in V_k} L_m \Big(\max_{P \in P_m} \prod_{(v_i,v_j)\in P} w\left(p_{i,j}(x_{i,j})\right) \Big) , \tag{4}$$

subject to her total security investment budget B_k, i.e., $\sum_{(v_i,v_j)\in\mathcal{E}_k} x^k_{i,j} \leq B_k$[4], and non-negativity of the investments, i.e., $x^k_{i,j} \geq 0$. We now show that the total loss (4) is convex.

Lemma 1. *Let the probability of successful attack function on each edge $p_{i,j}(x_{i,j})$ be twice-differentiable and log-convex. Then, the total loss function in* (4) *is convex in investment $x_{i,j}$.*

The proof of Lemma 1 follows from the second derivative of the total loss function in (4) with respect to $x_{i,j}$ and the properties of the probability weighting function in (3).

E. Socially Optimal Investments

It is also common in the literature to measure the sub-optimality of Nash equilibria (attained by interdependent security games between multiple selfish defenders) by comparing them to socially optimal (SO) investments. Formally, the socially optimal investment levels $\mathbf{x}^*$ are those that maximize the social welfare (i.e., these investments minimize the sum of all defenders' costs), which is given by

$$\mathbf{x}^* = \underset{\substack{\mathbf{x}\succeq\mathbf{0};\\ \mathbf{1}^T\mathbf{x}\leq\sum_{k=1}^{|D|} B_k}}{\operatorname{argmin}} \sum_{k=1}^{|D|} C_k(\mathbf{x}), \tag{5}$$

where $|D|$ is the number of defenders.

[4]Our findings will also follow if each defender invests any amount subject to a maximum budget. The stakeholder (defender) can use any amount from such a maximum budget limit for enhancing the security of her subnetwork.

A comparison of the Nash equilibria and the socially optimal solution often reveals sub-optimal investment in security by defenders at PNE where each defender only cares about her own critical assets. In the literature, there are several works that have proposed mechanisms for decreasing this inefficiency gap, by incentivizing improved security investments [37], [45]. However, these works studied specific games where each defender has a single asset in which she allocates her resources [37] or considered that all defenders have a common asset [45]. Moreover, all of these works considered only classical models of rational decision-making introduced earlier. On the contrary, we consider an attack graph based system where each defender has the ownership of a subset of nodes. Further, the interdependency between defenders is captured via overlapping paths for reaching different defenders' assets, and we model the behavioral probability weighting bias as well. These two distinctions make our setup more challenging compared to prior work and more representative of the reality of interdependent system security with humans acting as security decision-makers.

III. Mechanism Design Setup

The focus of the present paper is designing and evaluating regulatory mechanisms, specifically monetary taxation, to incentivize socially optimal security behavior for defenders in interdependent security games. Our goal is to find a mechanism, run by a central regulator (e.g., a government agency), such that the induced interdependent security game has as its equilibrium the solution to the centralized problem (5) (also referred to as "implementing" the socially optimal solution). Such mechanisms incentivize optimal behavior by assessing a tax t_k to each participating defender D_k; this tax may be positive, negative, or zero, indicating payments, rewards, or no payment, respectively. Similar to prior work [37], [46], we assume that defenders' costs are quasi-linear; i.e., linear in the tax term t_k. Therefore, the total (security) cost for a defender D_k when she is assigned a tax t_k is

$$C_k(\mathbf{x}, t_k) := C_k(\mathbf{x}) + t_k, \tag{6}$$

where the tax amount t_k can in general be a function of the total security investment $\mathbf{x}$ or the overall state of system's security (as will be explained later in Section V) where each mechanism corresponds to one form of t_k.

Remark 1. Following the previous works [37], [46], [47], we assume that the money used for the taxes paid by each defender comes from a separate pool from the pool from which the security enhancement budget of each defender is drawn. However, we believe that considering them to be from the same pool is an interesting direction for the future extensions.

Proposition 1. *There always exists a Pure-Strategy Nash Equilibrium in an unregulated (i.e. $t_i = 0$, $\forall i$) Behavioral Security Game as modeled in this section.*

The proof of the above result follows by noting from Lemma 1 that the cost function of each defender is convex in the security investment level $\mathbf{x}$ (equivalently the payoff is concave function in $\mathbf{x}$), thus this game is an instance of concave games which always have a PNE [48].

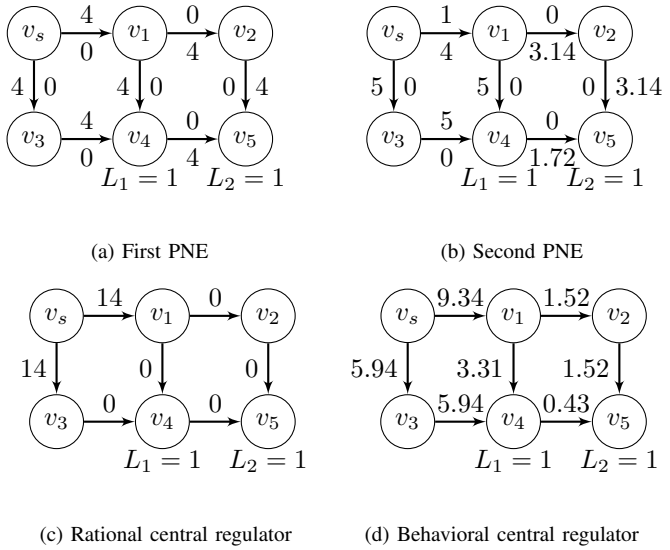
Mechanism Properties: In addition to implementing the socially optimal solution, incentive mechanisms are often designed so as to satisfy one main property. When using taxation, the mechanism designer prefers to maintain *weak budget balance* (WBB) [35], [46]; i.e., $\sum_{i=1}^{N} t_i \geq 0$. In other words, the regulator does not pay out to the defenders. In contrast, $\sum_{i=1}^{N} t_i < 0$ implies a budget deficit, i.e., the mechanism would require spending external resources by the designer. At first, we consider two mechanism designs where participation by defenders is mandatory (Sections V-A and V-B) and then we consider the mechanism where participation is voluntary (Section V-C). The mandatory participation maps to the realistic case that a government agency can make participation in cyber-insurance a prerequisite for companies to receive security funding or business opportunities [26]; see for example the recent California proposal for mandatory cyber insurance [49].

IV. Motivational Examples

Having provided the game notations and the general tax-based mechanism, we now provide a couple of examples to show the difference between the social optimal solution (given by (5)) and the PNE solution (where each defender best responds to the aggregate optimal investments of other defenders) to reach the PNE of Behavioral Security Games.

Example 1. Consider the attack graph of Figure 3. There are two defenders, D_1 and D_2, where defender D_1 aims to protect node v_4, and defender D_2 wishes to protect node v_5. Suppose that D_1 has a budget $B_1 = 16$ and D_2 has $B_2 = 12$, and let the probability of successful attack on each edge (v_i, v_j) be given by $p_{i,j}(x_{i,j}) = e^{-x_{i,j}}$ (assuming $p^0_{i,j} = 1$). Moreover, both defenders have behavioral bias with $\alpha_1 = \alpha_2 = 0.5$. Figures 3a and 3b illustrate two distinct PNE for this game.

We obtained these multiple Nash equilibria by varying the starting investment decision of defender D_1 and then following best response dynamics until the investments converged to an equilibrium. It is interesting to note that these two Nash equilibria lead to different costs for the defenders.



(a) First PNE (b) Second PNE

(c) Rational central regulator (d) Behavioral central regulator

Fig. 3: An instance of a Behavioral Security Game with multiple PNE and its corresponding social optimal solution. The costs for each defender are lower with the central regulator than with PNE. Defenders D_1 and D_2 are behavioral decision-makers with $\alpha_1 = \alpha_2 = 0.5$. In (a) and (b), the numbers above/left and below/right of the edges represent investments by D_1 and D_2, respectively. In (c) and (d) these numbers represent investments by rational and behavioral (with $\alpha = 0.5$) central regulator, respectively.

Difference between PNE and social optimal: First, for the Nash equilibrium of Figure 3a, defender D_1's perceived expected cost, given by (4), is equal to $\exp(-4)$, while her true expected cost, given by (1), is equal to $\exp(-8)$. Defender D_2 has a perceived expected cost of $\exp(-6)$, and a true expected cost of $\exp(-12)$. In contrast, for the Nash equilibrium in Figure 3b, defender D_1 has a perceived expected cost of $\exp(-4.5)$ and a true expected cost of $\exp(-10)$. Defender D_2 has a perceived expected cost of $\exp(-5.78)$ and a true expected cost of $\exp(-11.28)$. As a result, the equilibrium in Figure 3a is preferred by defender D_2, while the equilibrium in Figure 3b has a lower expected cost (both perceived and real) for defender D_1.

Second, we calculate the optimal investments by a social planner for such network. We assume that this social planner would have the same total budget (i.e., the sum of the two budgets of defenders D_1 and D_2) and calculate the optimal investment of that social planner (given by (5)). Figure 3c shows that the rational social planner would distribute her budget equally (only) on the edges (v_s, v_1) and (v_s, v_3) while Figure 3d shows that the behavioral social planner (with $\alpha = 0.5$) would distribute investments on all edges. We emphasize that the true expected cost of defender D_1 is $\exp(-14.0)$ and the true expected cost of defender D_2 is $\exp(-14.0)$ under rational central planning. On the other hand, the true expected cost of D_1 is $\exp(-11.88)$ and the true expected cost of defender D_2 is $\exp(-12.31)$ under behavioral central planning. In other words, rational central planning is better for both defenders and for the system as a whole.

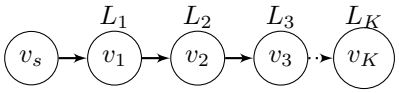


Fig. 4: An attack graph where the social optimal investment is better than the PNE's investments for all behavioral defenders.

Key takeaways: For both scenarios (rational social planner and behavioral social planner), the true costs are better (lower) for both defenders than in both of the attained PNEs. Moreover, the system's social cost is lower under such socially optimal solutions. This example sheds light on the inefficiency of the PNEs compared to the social optimal solution. In this context, the notion of *Price of Anarchy (PoA)* is often used to quantify the inefficiency of Nash equilibrium compared to the socially optimal outcome [50]. The Price of Anarchy is defined as the ratio of the highest total system cost at a PNE to the total system cost at the social optimum. In Example 1, the PoA under rational and behavioral social planning is 205.41 and 30.11, respectively, indicating a 205X and 30X reduction in expected security loss with central planning. The higher the PoA is, the greater is the motivation for centralized design of a mechanism that incentivizes the defenders to enhance their investments and achieve social optimal.

Example 2. Consider the attack graph in Figure 4, where the probability of successful attack on each edge (v_i, v_j) is given by (2) with $p^0_{i,j} = 1$. This graph contains $|D| = K$ defenders, and each defender D_k is responsible for defending target node v_k. Assume the total security budget B is divided equally between the K defenders (i.e., each defender has a security budget of $\frac{B}{K}$). Let all nodes $v_1, v_2, \cdots, v_K$ have same loss which is L. Then, the socially optimal solution would put all the budget B on the first edge (v_s, v_1), so that all nodes have probability of successful attack given by $\exp(-B)$.

We now characterize the cost under the PNE for behavioral defenders. This PNE is given by the investment profile where each defender D_k puts their entire budget $\frac{B}{K}$ on the edge coming into their node v_k. To show this, first consider defender D_1. Since investments on edges other than (v_s, v_1) do not affect the probability of successful attack at node v_1, it is optimal for defender D_1 to put all her investment on (v_s, v_1). Now, given D_1's investment on (v_s, v_1), defender D_2 should optimally spread her budget of $\frac{B}{K}$ over the two edges (v_s, v_1) and (v_1, v_2) in order to minimize her cost (4). Thus, D_2's optimization problem, given D_1's investment, is

$$\underset{x^2_{s,1}+x^2_{1,2}=\frac{B}{K}}{\text{minimize}} \quad e^{-(\frac{B}{K}+x^2_{s,1})^{\alpha_2}-\left(x^2_{1,2}\right)^{\alpha_2}}. \tag{7}$$

The unique optimal solution of (7) (for all $\alpha_2 \in (0, 1)$) would be to put all $\frac{B}{K}$ into the edge (v_1, v_2), i.e., $x^2_{1,2} = \frac{B}{K}$ and zero on the edge (v_s, v_1), i.e., $x^2_{s,1} = 0$.

Continuing this analysis, we see that if defenders $D_1, D_2, \ldots, D_{k-1}$ have each invested $\frac{B}{K}$ on the edges incoming into their nodes, it is optimal for defender D_k to also invest their entire budget $\frac{B}{K}$ on the incoming edge to v_k. Thus, investing $\frac{B}{K}$ on each edge is a PNE. Therefore, the true cost of defender D_1 under this PNE is given by $K\exp(-\frac{B}{K})$, which is much larger than this of the social optimal solution. Thus, the PoA in this game instance grows exponentially in the sum of budgets B.

In total, the two examples show the importance of attaining social optimal solution for both per-defender total real loss and the social cost (sum of defenders' real total losses).

V. Mechanism Types and Properties

We now provide two incentive mechanisms in our interdependent security games, and identify features of the interdependent systems that affect the properties attainable through these mechanisms. Specifically, we explain and study the performance of the two mechanisms (the Externality mechanism and the VCG mechanism) within our class of interdependent security games.

A. The Externality Mechanism

We now introduce the Externality mechanism inspired by the work of Hurwicz [51]. A main design goal of this mechanism is to guarantee a complete redistribution of taxes; i.e., strong budget balance. This mechanism has been adapted in [35], where it is shown to achieve social optimality, guarantee participation, and maintain a balanced budget, in allocation of power in cellular networks. However, the recent work [37] has shown that this is not the case in security games where each defender has a single asset in which she allocates her resources. However, that work only considered classical decision-making models (where all defenders are assumed to be fully rational decision-makers), and did not consider interdependency (attack graph models).

Let us denote the total tax paid by defender D_k at the equilibrium as t^*_k, which depends on the investment vector $\mathbf{x}$, i.e., $t^*_k = \mathbf{l^*_k x}$. We denote $\mathbf{l^*_k} := \{l^{kn^*}_{ij}\}_{(v_i,v_j)\in\mathcal{E}_n, D_n\in D}$ where $l^{kn^*}_{ij} = -L_k \frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*)$ is the positive externality of defender D_k due to defender D_n's investment on the edge (v_i, v_j).

To have the designed mechanism achieve the social optimal, the socially optimal investments $\mathbf{x}^*$ will be individually optimal as well; in other words, we have

$$\mathbf{x}^* \in \underset{\substack{\mathbf{x}\succeq\mathbf{0};\\ \mathbf{1}^T\mathbf{x}\leq\sum_{k=1}^{|D|}B_k}}{\text{argmin}} \quad C_k(\mathbf{x}) + \mathbf{l^*_k x}. \tag{8}$$

As a result, the Karush-Kuhn-Tucker (KKT) conditions on (8) yield that the tax term of defender D_k under the Externality mechanism in our interdependent security games is given by:

$$t^*_k(\mathbf{x}^*) = \sum_{n=1}^{|D|} \sum_{(v_i,v_j)\in\mathcal{E}_n} t^{kn^*}_{ij}. \tag{9}$$

In other words, the total tax paid by defender D_k is a summation of the taxes over all edges, where the tax on each edge depends on the sum of the externalities of all defenders on that edge. Specifically, the investment by defender D_n on the edge (v_i, v_j) is denoted by $x^{n^*}_{i,j}$.

Thus, the tax term that D_k pays due to the externality of defender D_n's investment on the edge (v_i, v_j) is given by

$$t^{kn^*}_{ij} = -L_k x^{n^*}_{ij} \frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*). \tag{10}$$

Interpretation of the Externality Mechanism: The interpretation of the above tax terms is that by implementing this externality mechanism, each defender D_k will be financing part of defender $D_n \neq D_k$'s reimbursement. According to (9) and (10), this amount is proportional to the positive externality of D_n's investment on D_k's cost.

Individual rationality and budget deficit: It can be shown that despite attaining the socially optimal solution, these taxes may fail to satisfy the weak budget balance constraint in our behavioral interdependent security games. We characterize this finding via the following result.

Proposition 2. *There exists an interdependent security game instance where the Externality mechanism cannot implement the social optimal while guaranteeing weak budget balance.*

Proof. See Appendix A. □

Interpretation: Proposition 2 shows a budget deficit case for the Externality mechanism in which the central regulator has to spend out-of-pocket money to incentivize the defenders to achieve the socially optimal solution in the context of our interdependent security games (modeled by attack graphs). Thus, we show for the first time that the prior result of Externality mechanism [35], [51], [52], social optimality *and* balanced budget, is not guaranteed in interdependent systems.

Now, we turn our attention to the effect of defender's behavioral bias on amount of taxes paid by the defender.

Theorem 1. *Consider a set of defenders D and an underlying attack graph G. Suppose that the joint investment profile by all defenders except D_k, denoted by $\mathbf{x}_{-k}$, is fixed. Suppose that $p_{i,j}(x_{i,j}) \in (0, \frac{1}{e}]$. Then the tax paid by defender D_k under Externality mechanism, denoted by $t_k^*(\mathbf{x}^*)$ in* (9)*, is a decreasing function in α_k. In other words, the behavioral defender pays more taxes compared to a rational defender.*

Proof. See Appendix B. □

Behavioral level and the amount of taxes: Theorem 1 shows that under appropriate conditions, the behavioral defender would pay more taxes compared to a rational defender under the Externality mechanism. The reason for such an increase in taxes is that the perception of the behavioral defender of the externality from other defenders' investments (via the drop in her perceived cost from such investments) induce the defender to pay more taxes for such a (perceived) increased safety level. We emphasize that the central regulator does not enforce rational decision-making on defenders but serves as a coordinator that facilitates the mechanism-based game between the defenders and incentivizes the optimal behavior of each defender by assessing a tax t_k (via creating the tax scheme upfront).

B. The VCG Mechanism

The second mechanism that we consider here is the VCG mechanism [36], [53], also commonly known as the Pivotal Mechanism. This is a family of mechanisms in which the central planner incentivizes users (defenders) to reveal their true preferences in dominant strategies through the appropriate design of taxes for users with quasi-linear utilities (or costs). This leads to achieving the socially optimal solution. In this mechanism, each defender D_k receives a monetary transfer equal to the amount he contributes to the rest of the society. This ingenious, but simple, idea leads to aligning the incentives of all players with the social cost.

VCG Mechanism Explanation: Let $\mathbf{x}^*_{-\mathbf{k}}$ denote the equilibrium (by all defenders except D_k) under exit of user D_k (i.e., assuming D_k is not spending anything on defense), which is given by

$$\mathbf{x}^*_{-\mathbf{k}} = \underset{\substack{\mathbf{x} \succeq \mathbf{0}; \\ \mathbf{1}^T \mathbf{x} \leq \sum_{j \neq k} B_j}}{\operatorname{argmin}} \sum_{j \neq k} C_j(\mathbf{x}). \tag{11}$$

Let $\overline{\mathbf{x}}$ represents a PNE investment vector by all defenders (including defender D_k). Thus, the taxes paid by D_k in the VCG mechanism under $\overline{\mathbf{x}}$ for our interdependent security games are given by

$$t_k^* = \sum_{j \neq k} C_j(\overline{\mathbf{x}}) - \sum_{j \neq k} C_j(\mathbf{x}^*_{-\mathbf{k}}). \tag{12}$$

Interpretation of the VCG mechanism: Intuitively, each defender receives a monetary transfer which is equivalent to her "contribution" to the rest of the society. For instance, if the defender D_k's investments makes the system worse, i.e., the social cost (without counting defender D_k) under the social optimal (including defender D_k's investments) is higher than the social cost without including her in the system, then the tax amount t_k^* would be positive. In other words, the mechanism penalizes the defender D_k for worsening the system. On the other hand, if defender D_k's investments makes the system better (i.e., with less social cost), t_k would be negative (i.e., D_k would receive such amount as a reward).

We now characterize the weak budget balance constraint and different amount of taxes paid by defenders under the VCG mechanism in our interdependent security games, respectively.

Proposition 3. *There exists an interdependent security game instance in which the VCG tax-based incentive mechanism cannot implement the socially optimal solution while guaranteeing weak budget balance.*

Proof. See Appendix C. □

Intuition: This result shows a budget deficit case for the VCG mechanism in which the central regulator has to spend out-of-pocket money to incentivize the defenders to achieve the social optimal solution. It was shown that the VCG mechanism achieves social optimality, and achieves weak budget balance in many private and public good games (see [36], [54], [55] for more details and related background). However, we show for the first time that this is not satisfied in interdependent security games. Fundamentally this is because, in interdependent security games a defender can free ride (i.e., under-invest in security and depend on investments from other defenders). Thus, such defender needs to be incentivized to achieve the socially optimal solution.

Effect of behavioral level on amount of taxes: We now show that higher behavioral bias (i.e., smaller α) leads to the payment of more taxes (by defenders) under the VCG mechanism. The reason for such increase in the taxes paid is that if any defender $D_k \in \mathcal{D}$ becomes more behavioral, her investments become more suboptimal and consequently increase (worsen) the system's social cost compared to the case in which D_k is not a member of the society. Thus, the VCG mechanism imposes more taxes on D_k in such scenario. We validate this finding in our evaluation (Section VI).

C. Voluntary Participation Mechanism Design

We next explore voluntary participation in interdependent security games modeled by attack graphs. To participate in the mechanism, a defender $D_k \in \mathcal{D}$ should have a preference for being part of the mechanism over opting out. In other words, the overall cost of defender D_k under the mechanism, which is the defender's cost under the attained joint investment profile by the mechanism plus the taxes paid by the defender to the central regulator (planner), must be lower than or equal to defender D_k's cost under PNE (for all defenders). Formally, a defender $D_k \in \mathcal{D}$ participates in the mechanism if

$$C_k(\mathbf{x}^*) + t_k \leq C_k(\bar{\mathbf{x}}),$$

where $C_k(\mathbf{x}^*)$ is defender D_k's cost under the socially optimal outcome (induced by the mechanism) and $C_k(\bar{\mathbf{x}})$ is the corresponding PNE (state of anarchy) with no defender $D_k \in \mathcal{D}$ being a part of the mechanism.

We first define a class of directed acyclic attack graphs (DAG) defined as a "Layered DAG" [56] which is a special case of a DAG where nodes are partitioned into l layers and the DAG has certain properties.[5]

Definition 1. *Let v_i^j be the j-th node in layer i and $H_i = \{v_i^j | \forall j\}$ be the set of all nodes in layer i. In a layered DAG, $\mathcal{E}$ only contains edges that connect nodes in H_i to nodes in H_{i+1}, $\forall 1 \leq i \leq l-1$.*

Amount of Taxes and Voluntary Participation: We now present result on voluntary participation in our tax-based framework for the introduced class of layered attack graphs.

Proposition 4. *Suppose that G denotes a layered DAG that has K behavioral defenders (with $\alpha_k \in (0,1)$), where each layer k has a single node v_k and under ownership of a defender $D_k \in D$. Suppose that the probability of successful attack on each edge (v_i, v_j) is given by (2) with $p_{i,j}^0 = 1$. Suppose that each defender has security budget $\frac{B}{K}$ and that L_i is the financial loss of asset v_i. Then, we have*

1) *If $t_i \leq L_i \left[\exp(-\frac{iB}{K}) - \exp(-B)\right]$, then defender D_i would participate in the mechanism*
2) *The maximum amount of tax t_i^{max} that a defender can be charged and participate in the mechanism is decreasing in the defender index $i, \forall i = 1, \cdots, K$.*[6]

Proof. See Appendix D. □

Intuition: The above result shows two main insights about taxation and participation in the mechanism. First, in the layered DAG, each defender would prefer participation in the mechanism if the amount of taxes she pays is less than or equal to the difference between the socially optimal solution and the state of anarchy (PNE). Otherwise, the defender would prefer to not participate in the mechanism since she can have a lower cost without participation. Second, the social planner can impose more taxes on the defenders that are nearer from the attacker's source node v_s in the attack graph compared to those who are far from the source node. The reason is that the latter can free-ride on the security investments of the former and will prefer PNE over the mechanism if they are charged high amount of taxes. For instance, in the DAG considered in Proposition 4, the maximum amount of tax to be imposed on the last defender D_K to participate in the mechanism is zero.

Remark 2. We also observe similar results of amount of taxes and participation in all of our four case studies (in Section VI) which have different attack graph structures.

VI. Evaluation

Our evaluation aims to answer the following questions:

- What is the gain of using mechanism design for incentivizing behavioral defenders toward the socially optimal solution?
- How does the level of behavioral bias affect the mechanism design outcomes?
- What is the maximum tax payment under which the defender prefers to participate in a tax-based mechanism over the state of anarchy (PNE)?

A. Dataset Description

We use four synthesized attack graphs that represent real-world interdependent systems to evaluate our setups. Specifically, we consider four popular interdependent systems from the literature which are: DER.1 [21], SCADA [16], E-commerce [17], and VoIP [17]. In all of these systems, nodes represent the progression of attack steps (e.g., unauthorized control of a physical generator in DER.1, taking privilege of control unit software in SCADA). Note that for each of our applications, it could be either air-gapped (here the attack would be from an insider attacker) or externally accessible (here the attack would be from an external adversary).

Now, we give a brief explanation of these systems and their associated failure scenarios. We generate the attack graphs of these systems using the CyberSage tool [21] which maps system's failure scenarios into an attack graph given the workflow of that system, security goals, and attacker's model.

[5]The layered DAG structure represents stepping-stone nature of attacks on the critical assets within the system that we consider here where attacker uses one asset in one layer to progressively attack other assets in deeper layers.

[6]Defender D_i's asset is closer to attacker's source node than defender D_{i+1}'s asset and thus defender D_i securing her asset benefits all D_j, with $j > i$.

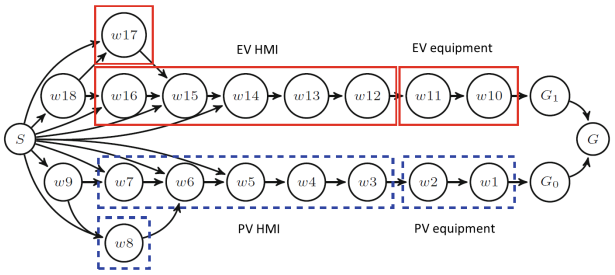


Fig. 5: Attack Graph of DER System

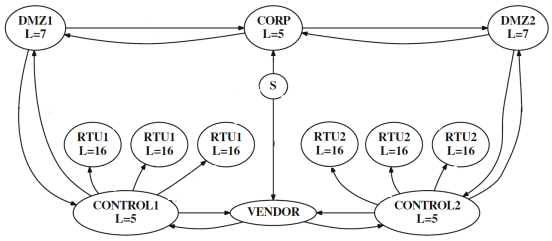


Fig. 6: Attack Graph of SCADA System

DER.1 System Description: The US National Electric Sector Cybersecurity Organization Resource (NESCOR) Technical Working Group has introduced a framework for evaluating the risks of cyber attacks on the electric grid. A distributed energy resource (DER) is described as a cyber-physical system consisting of entities such as generators, storage devices, and electric vehicles, that are part of the smart energy distribution system. The DER.1 failure scenario has been identified as the riskiest failure scenario affecting distributed energy resources according to the NESCOR ranking [21]. As shown in Figure 5, there are two critical equipment assets: a PhotoVoltaic (PV) generator and an electric vehicle (EV) charging station. Each equipment is accompanied by a Human Machine Interface (HMI), the only gateway through which the equipment can be controlled. The DER.1 failure scenario is triggered when the attacker gets access to the HMI. Once the attacker gets access to the system, she changes the DER settings and gets physical access to the DER equipment so that they continue to provide power even during a power system fault. Through this manipulation, the attacker can cause serious physical damage to the system.

SCADA System Description: The SCADA system is composed of two control subsystems, where each incorporates a number of cyber components, such as control subnetworks and remote terminal units (RTUs), and physical components, such as, valves controlled by the RTUs. We followed the NIST guidelines for industrial control systems for such architecture [57], where each subsystem is separated from external networks through a demilitarized zone (DMZ). The system implements firewalls both between the DMZ and the external networks, as well as between the DMZ and its control subnetwork. Therefore, an attacker must bypass two different levels of security to gain access to these control subnetworks. These two subsystems are interdependent via the shared corporate network, as well as due to having a common vendor for their control equipment. The resulting attack graph of the described system is shown in Figure 6. The "Corp" and the "Vendor" nodes connect the two subnetworks belonging to the two different defenders and can be used as jump points to spread an attack from one control subsystem to the other. This system has six critical assets (3 RTUs, Control Unit, CORP, and DMZ). The compromise of a control network "CONTROL i" will lead to loss of control of all 3 connected RTUs.

E-commerce System Description: The E-commerce system overview is shown in Figure 7. The web server sits in a DMZ separated by a firewall from the other two servers, which are connected to a network not accessible from the Internet. All connections from the Internet and through servers are controlled by the firewall. Rules state that the web and application servers can communicate, and the web server can be reached from the Internet. Here, the attacker is assumed to be external and thus her starting point is the Internet and uses stepping-stone attacks with the goal of having access to the MySQL database, represented by node 19 in the attack graph. For this system, we follow the attack graph generated by [17] (Figure 7 (on right), shaded nodes are detectors, not attack steps), based on popular vulnerabilities databases [58].

VoIP System Description: As shown in Figure 8, the VoIP system is composed of three zones; a DMZ for the servers accessible from the cloud, an internal network for local resources (e.g., computers, mail server and DNS server), and an internal network that is consisted of only VoIP components. This architecture follows the NIST security guidelines for deploying a secure VoIP system [59]. In this context, the VoIP network consists of a Proxy, voicemail server, and software-based and hardware-based phones. The firewall has rules to control the traffic between the three zones. Note that the DNS and mail servers in the DMZ are the only accessible nodes to the Internet. The PBX server can route calls to the Internet or to a public-switched telephone network (PSTN). The ultimate attack goal is to eavesdrop on VoIP communication. Figure 8 shows the resultant attack graph.

Having explained the failure scenarios of our four interdependent systems. Next, we present our experimental setup which includes simulation parameters and the procedure.

B. Experimental Setup

The simulations are based on our proposed game-theoretic models in Section II and mechanism-based models in Section V with the following parameters. Each system has two defenders. For DER, E-commerce, and VoIP, we have the financial losses $L_i = L = \$2M, \forall i$. The losses of the critical assets within SCADA (in Million dollars) are shown in Figure 6. We used the probability of successful attack function in (2) in our simulations. To estimate the baseline probability of successful attack on each edge (i.e., without any security investment), we first create a table of CVE-IDs (from real vulnerabilities reported in the CVE database for 2000-2020). We then followed [34] to convert the main attack's metrics

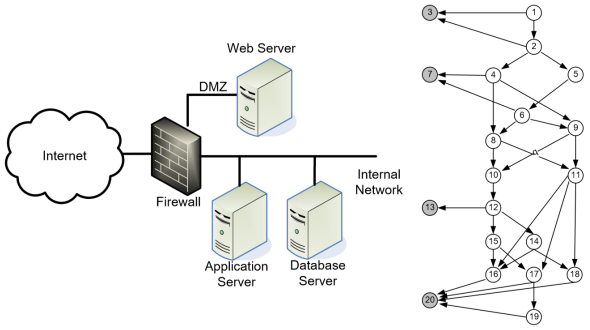


Fig. 7: A high level network overview of E-commerce (on left) adapted from [17]. The resultant attack graph (on right).

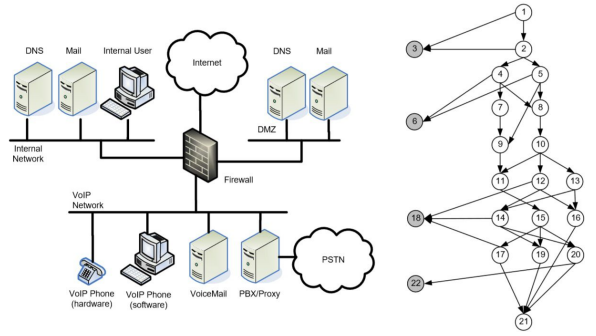
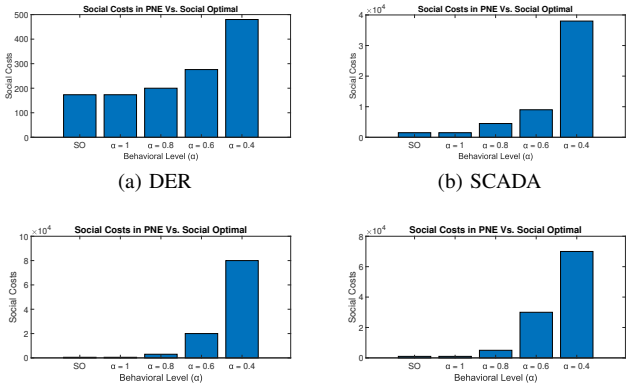


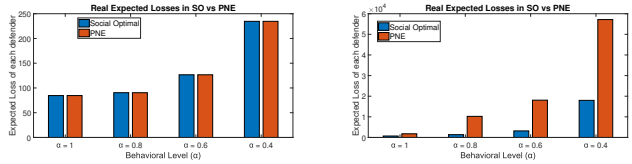
Fig. 8: A high level network overview of VoIP (on left) adapted from [17] and its resultant attack graph (on right).



(a) DER (b) SCADA

(c) E-commerce (d) VoIP

Fig. 9: A comparison of social costs under the socially optimal allocation (induced by mechanism) versus the PNE. We observe that the social cost under the socially optimal allocation is much lower than the social cost under PNE with behavioral defenders.



(a) DER (b) SCADA

Fig. 10: A comparison of expected loss of each defender under the social optimal (SO) versus the PNE under different behavioral levels. We observe that the expected loss under SO is lower than (same in DER) that under PNE irrespective of behavioral level.

(i.e., attack vector, attack complexity) to a baseline probability of successful attack. Table II (in Appendix E) illustrates this process for DER.1 and SCADA systems. We sweep the behavioral bias α such that $\alpha \in [0.4, 1]$; this is consistent with the range of behavioral parameters from prior experimental studies [19], [60]. We consider a symmetric security budget across the defenders (unless otherwise stated). For Nash Equilibrium, we run the best response dynamics until the game reaches the Nash Equilibrium while the social optimal is found using (5). TASHAROK refers to the setup with any of the two proposed mechanisms since both mechanisms lead to the social optimal, albeit with different tax collections.

C. Evaluation Results

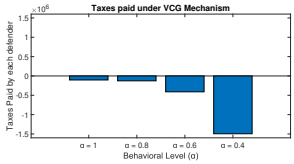
Next, we show our findings from different experiments for the four interdependent systems. Mainly, we compare the security investments (by both classes of decision-makers), the social costs under different investments, the per-defender expected loss, the amount of taxes (payments) under the two mechanisms (from Section V), the effect of behavioral decision-making, and the trends in voluntary participation.

Security Investments: We observe that the socially optimal allocation leads to distributing investments only on min-cut edges[7]. On the other hand, behavioral defenders distribute their investments across the network. This finding motivates the importance of incentivizing behavioral defenders to achieve social optimal investments since this would lead to reducing the per-defender real cost and the social cost as shown next.

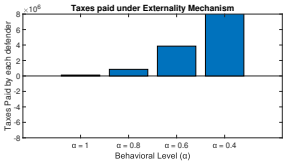
Social Costs: Figure 9a-9d demonstrate the reduction in social cost (which is the sum of the real costs of all defenders) following the implementation of the mechanism for the four systems. We observe that the mechanism design is more helpful for moderate and highly behavioral defenders since the behavioral investments under PNE is much worse than the social optimal solution. Numerically, as a result of risk reduction following the implementation of the mechanism, we see that the gain for society (represented by the ratio of the social cost under PNE to social cost under the mechanism) is 3X for DER, 180X for SCADA, 450X for E-commerce, and 390X for VoIP when the defenders are highly behavioral (i.e., $\alpha_1 = \alpha_2 = 0.4$). This result shows that the social cost under the socially optimal allocation is much lower than that under PNE, and the gap is higher for highly behavioral defenders and for systems with higher degree of interdependency.

Defender's Real Expected Loss: Figure 10a and 10b illustrate the real expected losses of all defenders under both the PNE and the socially optimal outcome (incentivized by the mechanism). Here the social planner is made to be behavioral along with the defenders, at the same level (same value of α). From the result, we see that implementing the proposed mechanisms would incentivize risk reduction for each defender for SCADA system while keeping the risk the same for the DER system. This happens due to the loose interdependency in the DER system. With such loose interdependency, the social optimality is achieved simply by the defenders individually spending their security resources efficiently.

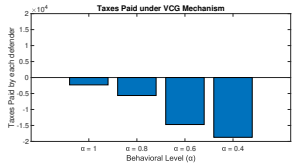
[7]The min-cut edges are the edges in the minimal set that can be removed to disconnect the graph. Here the same concept is applied to our attack graph.



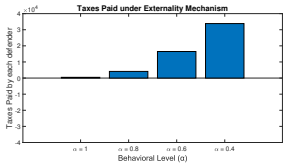
(a) VCG Mechanism (SCADA)



(b) Externality Mechanism (SCADA)



(c) VCG Mechanism (E-commerce)



(d) Externality Mech. (E-commerce)

Fig. 11: The amount of taxes paid by each defender under the studied mechanisms. For the VCG Mechanism, the player receives payment (i.e., pay negative taxes). On the other hand, under the Externality mechanism each defender pays positive taxes.

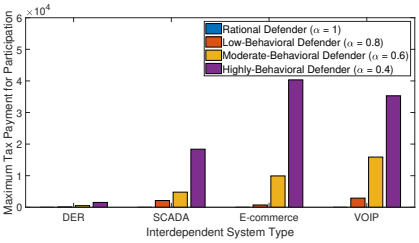
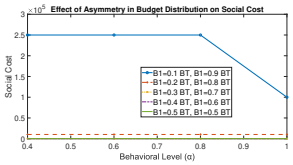


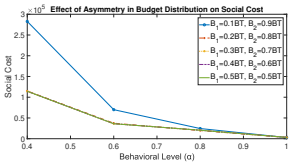
Fig. 12: The maximum amount of tax payment under which each defender participates in the mechanism for the four studied interdependent systems. The highly behavioral defender is willing to participate under higher tax payment.

Tax Payment Amounts: Here, we compare tax payments under different scenarios for both mechanisms that we study here. First, for the DER.1 system, it has the nature that each subnetwork is mainly affected by the corresponding defender. Therefore, under VCG mechanism, both defenders can reach the social optimal without paying taxes (i.e., budget balance for the central regulator). Hence, we omit this figure. However, for the SCADA system since the two subnetworks are mainly interdependent (i.e., if the attacker access both subnetworks via the Corp and the Vendor nodes, as explained earlier), the budget balance condition is not satisfied for the VCG mechanism. Figure 11a shows such insight where each of the two defenders is paid by the central regulator in the VCG mechanism since each defender makes the SCADA system more secure by her investments. We note also that although behavioral defenders invest suboptimally, they also benefit other defenders in the network (reduce the social cost) and thus need also to be paid by the VCG mechanism regulator. On the other hand, Figure 11b shows that the budget balance condition is satisfied with the Externality mechanism since each defender pays for the positive externalities on her cost due to other defender's investments. Figure 11c-11d show similar findings for E-commerce system due to interdependency among servers via firewalls and internet. We omit similar tax figures for VoIP.

Amount of Taxes and Voluntary Participation: Human bias is an important factor when trying to understand how



(a) DER



(b) SCADA

Fig. 13: The effect of asymmetry in security budget distribution across defenders on the social cost for different behavioral levels. Such effect is more pronounced under high budget asymmetries.

stakeholders would react to the security tax. Thus, we consider next the voluntary participation of the defenders under any quasi-linear tax-based mechanism. This requires calculation of the maximum tax payment under which each defender would participate in the mechanism. Figure 12 shows such maximum tax amount for our four interdependent systems under different behavioral levels. The highly behavioral defender is willing to participate in the mechanism even under higher tax payments since her suboptimal investments are far from the socially optimal level. For her, paying higher taxes and allocating resources according to the social optimum would yield lower total real loss compared to opting out and achieving PNE.

Asymmetry in Security Budget Distribution Across Defenders: We now study the effect of asymmetry in security budget distribution across the defenders and its effect on social cost. Figure 13a and 13b show the results for the DER and SCADA systems, respectively. For both systems, we observe that the social cost is higher with very high budget asymmetry and moderate to high behavioral level (i.e., one defender has 10% of the total security budget and the second defender has the remaining 90%). This observation can be explained by two facts. First, with a suboptimal behavioral allocation, the defender that has much less security budget wastes her constrained budget on non-critical edges. Second, the much richer behavioral player also allocates her resources suboptimally. This leads to this magnified increase in the social cost under extreme budget asymmetry. We observe similar findings for E-commerce and VoIP systems (Figures omitted).

Number of Defenders: We create a network with multiple defenders that contains replicas of these two subnetworks, and assume that new installed equipment corresponds to a new defender's subnetwork. We consider a symmetric distribution of the security budget over all defenders, with each defender having same security budget. Figure 16 shows that as the number of defenders increases, the difference between total losses between non-behavioral and behavioral games increases in a super-linear manner. For instance, when the number of defenders is 4, a change from non-behavioral to behavioral defenders ($\alpha = 0.6$) increases the loss by 8.65%, while the same change in α in the larger network with 16 defenders results in a substantial increase of 26.17%. This is due to the interdependencies between the subnetworks. For instance, if there are two defenders, each will incur a loss in two cases: when either her target asset is successfully compromised or the other defender's target asset is successfully compromised (as it can lead to the compromise of their common goal G).

On the other hand, if there are 16 defenders, for each defender, there are 16 possible paths through which she suffers a loss. This also explains why the total loss in the system increases as the number of defenders increases—the individual budget of each defender stays the same but the number of ways in which her asset can be compromised increases linearly.

Asymmetry of Cognitive Bias across Defenders: We study the effect of the asymmetry in cognitive bias across different defenders, in contrast to the previous experiments where both defenders had the same behavioral bias. Figure 17a-17b (in Appendix F) shows the effect of such heterogeneity on the social cost. We observe that two defenders with moderate behavioral levels ($\alpha_1 = 0.6, \alpha_2 = 0.6$) would invest better and consequently reduce the social cost more compared to two heterogeneous defenders (with $\alpha_1 = 0.8, \alpha_2 = 0.4$), although the mean behavioral level is the same in the two cases.

Sensitivity of Edges to Investments: Finally, we show the effect of edges' sensitivity to investments on each defender's real expected loss for different behavioral levels. We present this experiment in Appendix F. The high-level takeaway is that a defender whose edge is more sensitive to investments (i.e., the probability of successful attack goes down faster) gets more tax payments under the mechanisms.

D. Baseline Systems

We compare TASHAROK with four baseline systems under rational defenders: the seminal work of [12] for security investment with attack graphs on attack graph generation and investment decision analysis[8], [38] for placing security resources using defense in depth technique which traverses all edges that can be used to compromise each critical asset and distribute resources equally on them, the recent work [25] that explored behavioral decision-making in a non-cooperative setup (PNE characterization), and the recent work [9] that showed that attackers follow shorter paths to exploit target assets in the generated attack graphs. Table I shows such comparison by calculating the social cost under each work's defense allocation, indicating the superiority of TASHAROK for almost all our interdependent systems (note similar results between our proposed approach and most baselines for DER.1 in Table I due to the weak interdependency in this system). Since three of the four baselines (except [25]) do not design for behavioral defenders, we do not consider such defenders in this experiment. The result bears out the fact that the defense investments given by [25] and [12] are identical under rational decision-making.

TABLE I: Comparison of TASHAROK ***and baseline systems in terms of the social cost under each system's defense allocation (lower is better). For*** TASHAROK***, we consider a rational social planner.*** TASHAROK ***gives the best defense allocation among the techniques (resp. the lowest social cost).***

System Type	S&P02 [12]	Milcomm06 [38]	AsiaCCS21 [25]	CCS21 [9]	TASHAROK
DER.1	**173.390**	600.451	**173.390**	**173.390**	**173.390**
SCADA	513.230	4.023×10^4	513.230	5.902×10^3	**222.210**
E-commerce	47.014	8.115×10^4	47.014	2.493×10^4	**45.001**
VoIP	184.120	1.525×10^5	184.120	1.4859×10^4	**110.21**

[8]More recent approaches (e.g., [61], [62], [63]) follow same strategy of [12].

VII. Related Work

Game-theoretic modeling of security: Game theory has been used to describe the interactions between attackers and defenders and their effects on system security. A commonly used model in this context is that of two-player games, where a single attacker attempts to compromise a system controlled by a single defender [64], [65]. Game theoretic models have also been used in [14], [66], [67] to study the interaction between one defender and (multiple) attackers attempting Distributed Denial of Service attacks. Our work differs from both of these lines of literature by considering the interdependencies between multiple defenders in the network. Game theoretic models have also been used to study critical infrastructure security [13], [68], censorship-resilient proxy distribution [15], and protecting networks from cascade attacks [69], [70]. The major difference of our work is that we analyze behavioral models of decision-making while existing work has focused on classical game-theoretic models of rational decision-making. Moreover, previous research does not consider mechanism designs to improve security as we do here.

Human behavior in security and privacy: Notable departures from classical economic models within the privacy literature are [41], [71], which identify the effects of behavioral decision-making on personal privacy choices. The importance of considering similar models in the study of system security has been recognized in the literature [60], [72]. Prior work considers models from behavioral economics in the context of security applications, but based only on psychological studies [42], [73] and human subject experiments [74], [75], [30] for the end user. Our work differs by exploring a rigorous mathematical model of decision-makers' behavior. We also model the interaction between multiple defenders (in contrast to only one defender in these studies), consider interdependent assets (in contrast to studies that examine binary decisions for isolated assets), and study the mechanism design for enhancing security decision-making of behavioral defenders (in contrast to these studies that did not consider any mitigation). A few studies provide theoretical treatment of behavioral decision-making in specific classes of interdependent security games [33], [32], [76]. These works, however do not consider any mitigation for behavioral decision-making.

Mechanism design in security: The motivation for considering mechanism design models in the security literature comes from two main characteristics of security games with multiple defenders. First, the security investments of each defender can help other defenders, similar to public good provision with positive externalities. Second, defenders can therefore free ride and depend on security investments by other defenders. This leads to an inefficiently low overall security level of the system [35], [36]. This motivates the study of mechanisms for improving network security, and ideally, incentivizing user cooperation and driving the system to a socially optimal state of enhanced overall security, e.g., [37], [46]. However, to the best of our knowledge, no previous work in mechanism design has investigated

behavioral decision-making effects and considered attack graphs that can model any large-scale interdependent system.

Our presented impossibility result differs from those in the existing literature, which builds on the seminal work of Green and Laffont [77]. We differ in terms of the selected equilibrium solution concept, the set of properties the mechanism is required to satisfy, the space of cost functions, or the nature of the system type. For instance, the Myerson and Satterthwaite result [36] considers a Bayesian Nash solution while we have a Pure Nash implementation. On the other hand, Maskin's work on implementation theory [78] considers Nash equilibrium for the complete information setup. However it requires that all NE be socially optimal which cannot be guaranteed in interdependent security games (see Section IV). Finally, the line of work [37], [46] has considered quasi-linear costs (where the tax is added to the original cost function) and Nash equilibrium solutions of the mechanisms, which has two main differences from our present work. First, they consider utilities with classical decision-making models, without the cognitive biases that we consider here. Second, they do not consider interdependent systems with attack graphs.

VIII. Discussion

(1) Existence of bias in security decision-makers: Numerous academic studies of even the most highly-trained specialists have shown that experts are also susceptible to systematic failures of human cognition (e.g., [79], [80]). Specifically, the work [80] has conducted a survey of experiments that considered behavior of students against experts in a wide variety of professions. This survey reports only one out of thirteen considered studies found that professionals make decisions more closely in line with standard economic theory. Moreover, recent research has shown that cybersecurity professionals' probability perceptions are as susceptible to systematic biases as those of the general population [27], [28]. Finally, even if security experts exhibit weaker biases, this can still result in sub-optimal security investments and their effects may be magnified due to the magnitude of losses associated with compromised real-world assets that these experts control.
(2) Guiding security decision-makers: We acknowledge that the security state of a system cannot be fully summarized by only one measure. However, we believe that our framework provides an important estimate of the probability of successful attack (resp. expected financial loss). We compose that estimate from something that is easier to grasp — the loss due to each asset in the system being compromised. We believe that our work opens up a new dimension of *intervention* in securing interdependent systems. Our framework allows for a quantification of the improvements in security that can be obtained by incentivizing security professionals to invest better via our proposed mechanisms. Furthermore, this framework can guide operators of large-scale interdependent systems (akin to social planners), by allowing the operator to investigate subsystems within the system where sub-optimal security investments might have been made by subordinates operating those subsystems and by calculating the taxes charged to each subordinate to participate in the mechanism and enhance overall system security (social cost). As shown in Section VI, this would depend on the nature of the network and the interdependency among different defenders. We emphasize that our focus in this paper is to explore the benefit of such mechanism for human decision-makers with cognitive biases securing interdependent systems and explore its outcomes.
(3) Mechanism design to solve behavioral bias in different security problems: Our proposed adaptation of the Externality and VCG mechanisms to interdependent security games (Section V) can be further used for different security problems. Examples include defending isolated assets with heterogeneous valuations, e.g., for enhancing security decisions to defend different airports [81] or preventing DAG-based ransomware attacks [82]. Recent work has shown the effect of cognitive biases on security resource allocations in such settings using decision- and game-theoretic analysis [76], [33]. However, these studies do not consider any mitigation for such biases. Thus, using mechanism design to improve such biases would be an avenue for future work.

IX. Conclusion

We studied interdependent systems that are managed by multiple stakeholders and are prone to cyber attacks that progress in a stepping-stone manner. We modeled such attack scenarios using a game-theoretic framework and captured the attack progression and interdependency via attack graphs, in our framework called TASHAROK. We then analyzed two tax-based mechanism types for our interdependent security setups where the central regulator incentivizes defenders to achieve socially optimal allocations. We then showed that a mechanism designer cannot guarantee social optimality without paying money to incentivize defenders in all instances of our interdependent security games. We also showed the effect of behavioral bias on the two mechanisms' outcomes where higher bias leads to paying more taxes. We then explored the relation between the tax amount and the voluntary participation of defenders in the mechanism and showed that behavioral defenders choose to participate in the mechanism even under higher tax payments, compared to rational defenders. We evaluated TASHAROK via four real-world interdependent systems and showed the effect of mechanisms on social cost and the effect of behavioral decision-making on the mechanisms' outcomes. We compared the security cost achieved by security allocations of TASHAROK compared to those of four baseline solutions from the attack graph literature. We found that even with rational defenders TASHAROK either equals or outperforms the baselines.

We believe that our study can help central regulators and interdependent systems' defenders attain improved understanding of their security risks and consequently make more effective investment decisions to mitigate such risks, including additional risk from decisions under cognitive biases. Future avenues of research include characterizing the achievable security allocation, as well as the associated mechanisms, and exploring attackers with cognitive bias.

ACKNOWLEDGMENT

We thank the anonymous reviewers for their valuable comments to improve the quality of this paper. This material is based in part upon work supported by the National Science Foundation under Grant Number CNS-1718637, Wabash Heartland Innovation Network (WHIN) project from Lilly Endowment Inc. NSF CCF-1919197, and Army Research Lab under Contract number W911NF-2020-221. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security – a survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802–1831, 2017.

[2] H. Chen, M. Pendleton, L. Njilla, and S. Xu, "A survey on ethereum systems security: Vulnerabilities, attacks, and defenses," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–43, 2020.

[3] I. Week, "The 10 biggest cyber security attacks of 2020," https://searchsecurity.techtarget.com/news/252494362/10-of-the-biggest-cyber-attacks, Jan 2021, [Online; accessed 1-October-2021].

[4] J. Robertson and W. Turton, "Colonial Pipeline ransomware attack," https://www.bloomberg.com/news/articles/2021-05-09/colonial-hackers-stole-data-thursday-ahead-of-pipeline-shutdown, May 2021, [Online; accessed 30-October-2021].

[5] H. Choi, S. Kate, Y. Aafer, X. Zhang, and D. Xu, "Cyber-physical inconsistency vulnerability identification for safety checks in robotic vehicles," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 2020, pp. 263–278.

[6] J. Ding, Y. Atif, S. F. Andler, B. Lindström, and M. Jeusfeld, "Cps-based threat modeling for critical infrastructure protection," *ACM SIGMETRICS Performance Evaluation Review*, vol. 45, no. 2, pp. 129–132, 2017.

[7] A. B. Sharma, F. Ivančić, A. Niculescu-Mizil, H. Chen, and G. Jiang, "Modeling and analytics for cyber-physical systems in the age of big data," *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 4, pp. 74–77, 2014.

[8] NREL, "Cybersecurity Threat Evaluation on renewable energy systems," https://www.nrel.gov/news/program/2021/nrel-joins-industry-in-leading-cybersecurity-threat-evaluation-for-us-wind-fleet.html, April 2021, [Online; accessed 1-February-2022].

[9] A. Nadeem, S. Verwer, S. Moskal, and S. J. Yang, "Enabling visual analytics via alert-driven attack graphs," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 2420–2422. [Online]. Available: https://doi.org/10.1145/3460120.3485361

[10] Y. Baryshnikov, "IT Security Investment and Gordon-Loeb's 1/e rule." in *Workshop on the Economics of Information Security (WEIS)*, 2012.

[11] M. Mirkin, Y. Ji, J. Pang, A. Klages-Mundt, I. Eyal, and A. Juels, "Bdos: Blockchain denial-of-service," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 601–619. [Online]. Available: https://doi.org/10.1145/3372297.3417247

[12] O. Sheyner, J. Haines, S. Jha, R. Lippmann, and J. M. Wing, "Automated generation and analysis of attack graphs," in *Proceedings 2002 IEEE Symposium on Security and Privacy*. IEEE, 2002, pp. 273–284.

[13] A. Laszka, M. Felegyhazi, and L. Buttyan, "A survey of interdependent information security games," *ACM Computing Surveys (CSUR)*, vol. 47, no. 2, p. 23, 2015.

[14] G. Yan, R. Lee, A. Kent, and D. Wolpert, "Towards a bayesian network game framework for evaluating ddos attacks and defense," in *Proceedings of the 2012 ACM conference on Computer and communications security (CCS)*, 2012, pp. 553–566.

[15] M. Nasr, S. Farhang, A. Houmansadr, and J. Grossklags, "Enemy at the gateways: Censorship-resilient proxy distribution using game theory." in *NDSS*, 2019.

[16] A. R. Hota, A. Clements, S. Sundaram, and S. Bagchi, "Optimal and game-theoretic deployment of security investments in interdependent assets," in *International Conference on Decision and Game Theory for Security*, 2016, pp. 101–113.

[17] G. Modelo-Howard, S. Bagchi, and G. Lebanon, "Determining placement of intrusion detectors for a distributed application through bayesian network modeling," in *International Workshop on Recent Advances in Intrusion Detection*. Springer, 2008, pp. 271–290.

[18] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the econometric society*, vol. 47, pp. 263–291, 1979.

[19] R. Gonzalez and G. Wu, "On the shape of the probability weighting function," *Cognitive psychology*, vol. 38, no. 1, pp. 129–166, 1999.

[20] D. R. Cavagnaro, M. A. Pitt, R. Gonzalez, and J. I. Myung, "Discriminating among probability weighting functions using adaptive design optimization," *Journal of risk and uncertainty*, vol. 47, no. 3, pp. 255–289, 2013.

[21] S. Jauhar, B. Chen, W. G. Temple, X. Dong, Z. Kalbarczyk, W. H. Sanders, and D. M. Nicol, "Model-based cybersecurity assessment with nescor smart grid failure scenarios," in *Dependable Computing (PRDC), 2015 IEEE 21st Pacific Rim International Symposium on*. IEEE, 2015, pp. 319–324.

[22] I. Week, "IT Leadership: 3 Tips for Making Better Investments in Security," https://www.informationweek.com/strategic-cio/3-tips-for-making-better-investments-in-security/a/d-id/13298802, September 2017, [Online; accessed 30-October-2021].

[23] F. T. Council, "CISO Should Stand For Chief Influence Security Officer," https://www.forbes.com/sites/forbestechcouncil/2018/09/24/ciso-should-stand-for-chief-influence-security-officer/, September 2018, [Online; accessed 20-September-2021].

[24] D. Dor and Y. Elovici, "A model of the information security investment decision-making process," *Computers & security*, vol. 63, pp. 1–13, 2016.

[25] M. Abdallah, D. Woods, P. Naghizadeh, I. Khalil, T. Cason, S. Sundaram, and S. Bagchi, "Morshed: Guiding behavioral decision-makers towards better security investment in interdependent systems," in *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*, 2021, pp. 378–392.

[26] S. Dambra, L. Bilge, and D. Balzarotti, "Sok: Cyber insurance–technical challenges and a system security roadmap," in *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2020, pp. 1367–1383.

[27] K. Mersinas, B. Hartig, K. M. Martin, and A. Seltzer, "Experimental elicitation of risk behaviour amongst information security professionals." in *14th Workshop on the Economics of Information Security (WEIS)*, 2015.

[28] ——, "Are information security professionals expected value maximizers?: An experiment and survey-based test," *Journal of Cybersecurity*, vol. 2, no. 1, pp. 57–70, 12 2016. [Online]. Available: https://doi.org/10.1093/cybsec/tyw009

[29] C. Tiefenau, M. Häring, K. Krombholz, and E. Von Zezschwitz, "Security, availability, and multiple information sources: Exploring update behavior of system administrators," in *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, 2020, pp. 239–258.

[30] P. Morgner, C. Mai, N. Koschate-Fischer, F. Freiling, and Z. Benenson, "Security update labels: establishing economic incentives for security patching of iot consumer products," in *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2020, pp. 429–446.

[31] M. Abdallah, T. Cason, S. Bagchi, and S. Sundaram, "The effect of behavioral probability weighting in a simultaneous multi-target attacker-defender game," in *2021 European Control Conference (ECC)*. IEEE, 2021, pp. 933–938.

[32] M. Abdallah, P. Naghizadeh, A. R. Hota, T. Cason, S. Bagchi, and S. Sundaram, "Behavioral and game-theoretic security investments in interdependent systems modeled by attack graphs," *IEEE Transactions on Control of Network Systems*, 2020.

[33] A. Sanjab, W. Saad, and T. Başar, "Prospect theory for enhanced cyber-physical security of drone delivery systems: A network interdiction game," in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[34] J. Homer, S. Zhang, X. Ou, D. Schmidt, Y. Du, S. R. Rajagopalan, and A. Singhal, "Aggregating vulnerability metrics in enterprise networks using attack graphs," *Journal of Computer Security*, vol. 21, no. 4, pp. 561–597, 2013.

[35] S. Sharma and D. Teneketzis, "A game-theoretic approach to decentralized optimal power allocation for cellular networks," *Telecommunication systems*, vol. 47, no. 1, pp. 65–80, 2011.

[36] D. C. Parkes, *Iterative combinatorial auctions: Achieving economic and computational efficiency*. University of Pennsylvania, PA, 2001.

[37] P. Naghizadeh and M. Liu, "Exit equilibrium: Towards understanding voluntary participation in security games," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[38] R. Lippmann, K. Ingols, C. Scott, K. Piwowarski, K. Kratkiewicz, M. Artz, and R. Cunningham, "Validating and restoring defense in depth using attack graphs," in *IEEE Military Communications Conference*. IEEE, 2006, pp. 1–10.

[39] O. Stan, R. Bitton, M. Ezrets, M. Dadon, M. Inokuchi, O. Yoshinobu, Y. Tomohiko, Y. Elovici, and A. Shabtai, "Extending attack graphs to represent cyber-attacks in communication protocols and modern it networks," *IEEE Transactions on Dependable and Secure Computing*, 2020.

[40] O. Stan, R. Bitton, M. Ezrets, M. Dadon, M. Inokuchi, Y. Ohta, T. Yagyu, Y. Elovici, and A. Shabtai, "Heuristic approach for countermeasure selection using attack graphs," in *2021 IEEE 34th Computer Security Foundations Symposium (CSF)*. IEEE, 2021, pp. 1–16.

[41] A. Acquisti, "Nudging privacy: The behavioral economics of personal information," *IEEE security & privacy*, vol. 7, no. 6, 2009.

[42] R. Anderson, "Security economics: a personal perspective," in *Proceedings of the 28th Annual Computer Security Applications Conference*. ACM, 2012, pp. 139–144.

[43] L. A. Gordon and M. P. Loeb, "The economics of information security investment," *ACM Transactions on Information and System Security (TISSEC)*, vol. 5, no. 4, pp. 438–457, 2002.

[44] D. Prelec, "The probability weighting function," *Econometrica*, vol. 66, no. 3, pp. 497–527, 1998.

[45] H. Varian, "System reliability and free riding," in *Economics of information security*. Springer, 2004, pp. 1–15.

[46] P. Naghizadeh and M. Liu, "Opting out of incentive mechanisms: A study of security as a non-excludable public good," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2790–2803, 2016.

[47] M. M. Khalili, X. Zhang, and M. Liu, "Contract design for purchasing private data using a biased differentially private algorithm," in *Proceedings of the 14th Workshop on the Economics of Networks, Systems and Computation*, 2019, pp. 1–6.

[48] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica: Journal of the Econometric Society*, pp. 520–534, 1965.

[49] M. Rasch, "California Proposal for Mandatory Cyber Insurance," https://securityboulevard.com/2020/03/california-proposal-for-mandatory-cyber-insurance//, 2020, [Online; accessed 21-October-2021].

[50] T. Roughgarden, "The price of anarchy is independent of the network topology," *Journal of Computer and System Sciences*, vol. 67, no. 2, pp. 341–364, 2003.

[51] 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.

[52] L. Mathevet, "Supermodular mechanism design," *Theoretical Economics*, vol. 5, no. 3, pp. 403–443, 2010.

[53] W. Conen and T. Sandholm, "Partial-revelation vcg mechanism for combinatorial auctions," in *AAAI/IAAI*, 2002, pp. 367–372.

[54] T. Groves and M. Loeb, "Incentives and public inputs," *Journal of Public economics*, vol. 4, no. 3, pp. 211–226, 1975.

[55] A. Wolitzky, "Mechanism design with maxmin agents: Theory and an application to bilateral trade," *Theoretical Economics*, vol. 11, no. 3, pp. 971–1004, 2016.

[56] S. Milani, W. Shen, K. S. Chan, S. Venkatesan, N. O. Leslie, C. Kamhoua, and F. Fang, "Harnessing the power of deception in attack graph-based security games," in *International Conference on Decision and Game Theory for Security*. Springer, 2020, pp. 147–167.

[57] K. Stouffer, "Guide to industrial control systems (ics) security," *NIST special publication*, vol. 800, no. 82, pp. 16–16, 2011.

[58] "gtraq Vulnerability Database," https://www.securityfocus.com/, September 2021, [Online; accessed 18-September-2021].

[59] D. R. Kuhn, T. J. Walsh, and S. Fries, "Security considerations for voice over ip systems," *NIST special publication*, vol. 800, 2005.

[60] D. Woods, M. Abdallah, S. Bagchi, S. Sundaram, and T. Cason, "Network defense and behavioral biases: An experimental study," *Experimental Economics*, pp. 1–33, 2021.

[61] M. Zhang, L. Wang, S. Jajodia, A. Singhal, and M. Albanese, "Network diversity: a security metric for evaluating the resilience of networks against zero-day attacks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 1071–1086, 2016.

[62] X. Sun, J. Dai, P. Liu, A. Singhal, and J. Yen, "Using bayesian networks for probabilistic identification of zero-day attack paths," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 10, pp. 2506–2521, 2018.

[63] H. S. Lallie, K. Debattista, and J. Bal, "A review of attack graph and attack tree visual syntax in cyber security," *Computer Science Review*, vol. 35, p. 100219, 2020.

[64] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu, "A survey of game theory as applied to network security," in *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*. IEEE, 2010, pp. 1–10.

[65] H. S. Bedi, S. Roy, and S. Shiva, "Game theory-based defense mechanisms against DDoS attacks on TCP/TCP-friendly flows," in *IEEE Symposium on Computational Intelligence in Cyber Security (CICS)*. IEEE, 2011, pp. 129–136.

[66] Z. Xu and J. Zhuang, "A study on a sequential one-defender-n-attacker game," *Risk Analysis*, vol. 39, no. 6, pp. 1414–1432, 2019.

[67] M. Mirkin, Y. Ji, J. Pang, A. Klages-Mundt, I. Eyal, and A. Juels, "Bdos: Blockchain denial-of-service," in *Proceedings of the 2020 ACM SIGSAC conference on Computer and Communications Security*, 2020, pp. 601–619.

[68] L. Perelman and S. Amin, "A network interdiction model for analyzing the vulnerability of water distribution systems," in *Proceedings of the 3rd international conference on High confidence networked systems*. ACM, 2014, pp. 135–144.

[69] R. J. La, "Interdependent security with strategic agents and cascades of infection," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1378–1391, 2015.

[70] R. La, "Influence of clustering on cascading failures in interdependent systems," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 351–363, 2018.

[71] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509–514, 2015.

[72] L. F. Cranor, "A framework for reasoning about the human in the loop." *Proc. 1st Conference on Usability, Psychology, and Security, Usenix Assoc.*, 2008.

[73] R. Anderson and T. Moore, "Information security economics–and beyond," in *Annual International Cryptology Conference*. Springer, 2007, pp. 68–91.

[74] E. M. Redmiles, M. L. Mazurek, and J. P. Dickerson, "Dancing pigs or externalities?: Measuring the rationality of security decisions," in *Proceedings of the 2018 ACM Conference on Economics and Computation*. ACM, 2018, pp. 215–232.

[75] T. Wu, R. Zhang, W. Ma, S. Wen, X. Xia, C. Paris, S. Nepal, and Y. Xiang, "What risk? i don't understand. an empirical study on users' understanding of the terms used in security texts," in *Proceedings of 2020 ACM Asia Conference on Computer and Communications Security*, 2020.

[76] M. Abdallah, P. Naghizadeh, T. Cason, S. Bagchi, and S. Sundaram, "Protecting assets with heterogeneous valuations under behavioral probability weighting," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 5374–5379.

[77] J. Green and J.-J. Laffont, *Incentives in public decision-making*. Elsevier North-Holland, 1979.

[78] E. Maskin, "The theory of implementation in nash equilibrium: A survey," *Cambridge, Mass.: Dept. of Economics, Massachusetts Institute of Technology*, 1983.

[79] L. Haynes, B. Goldacre, D. Torgerson *et al.*, "Test, learn, adapt: developing public policy with randomised controlled trials," *Cabinet Office-Behavioural Insights Team*, 2012.

[80] G. R. Fréchette and A. Schotter, *Handbook of experimental economic methodology*. Oxford University Press, USA, 2015.

[81] G. Kuper, F. Massacci, W. Shim, and J. Williams, "Who should pay for interdependent risk? policy implications for security interdependence among airports," *Risk Analysis*, vol. 40, no. 5, pp. 1001–1019, 2020.

[82] A. Zimba, Z. Wang, and H. Chen, "Reasoning crypto ransomware infection vectors with bayesian networks," in *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*. IEEE, 2017, pp. 149–151.

APPENDIX

A. Proof of Proposition 2

Proof. The proof of Proposition 2 follows by the following counter example in which we give an instance of interdependent security game that has a budget deficit.

Example 3. Consider the attack graph in Figure 14. This graph contains 2 rational defenders ($\alpha_1 = \alpha_2 = 1$), and each defender D_k is responsible for defending target node v_k. Let the defender D_1's node have loss equal to L_1, and the defender D_2's node have loss L_2. From (8), the cost of defender D_1 is given by $C_1(\mathbf{x}) = L_1 e^{-x^1_{s,1}-x^2_{s,1}} + \begin{bmatrix} l^{11^*}_{s,1} & l^{12^*}_{s,1} & l^{11^*}_{1,2} & l^{12^*}_{1,2} \end{bmatrix} \begin{bmatrix} x^1_{s,1} \\ x^2_{s,1} \\ x^1_{1,2} \\ x^2_{1,2} \end{bmatrix}$. Thus, the Lagrangian of the defender D_1 is given by

$$\mathcal{L}(\mathbf{x},\mu) = L_1 e^{-x^1_{s,1}-x^2_{s,1}} + \begin{bmatrix} l^{11^*}_{s,1} & l^{12^*}_{s,1} & l^{11^*}_{1,2} & l^{12^*}_{1,2} \end{bmatrix} \begin{bmatrix} x^1_{s,1} \\ x^2_{s,1} \\ x^1_{1,2} \\ x^2_{1,2} \end{bmatrix} + \mu[x^1_{s,1} + x^1_{1,2} - B_1]. \tag{13}$$

Applying KKT conditions [59] to (13) yields

$$-L_1 e^{-x^1_{s,1}-x^2_{s,1}} + l^{11^*}_{s,1} + \mu = 0 \tag{14}$$
$$-L_1 e^{-x^1_{s,1}-x^2_{s,1}} + l^{12^*}_{s,1} = 0 \tag{15}$$
$$l^{11^*}_{1,2} + \mu = 0 \tag{16}$$
$$l^{12^*}_{1,2} = 0. \tag{17}$$

Thus, by solving (15)-(18) we have

$$\mathbf{l^*_1} = \begin{bmatrix} L_1 e^{-x^1_{s,1}-x^2_{s,1}} - \mu & L_1 e^{-x^1_{s,1}-x^2_{s,1}} & -\mu & 0 \end{bmatrix}$$
$$t^*_1 = \mathbf{l^*_1} \begin{bmatrix} x^1_{s,1} & x^2_{s,1} & x^1_{1,2} & x^2_{1,2} \end{bmatrix}^T.$$

Similarly, calculating the Lagrangian of defender D_2 and doing a similar analysis to that of defender D_1 and letting $\beta_2 = e^{-x^1_{s,1}-x^2_{s,1}-x^1_{1,2}-x^2_{1,2}}$, the tax terms of defender D_2 are

$$\mathbf{l^*_2} = L_2 \begin{bmatrix} \beta_2 & \beta_2 - \frac{\mu'}{L_2} & \beta_2 & \beta_2 - \frac{\mu'}{L_2} \end{bmatrix},$$
$$t^*_2 = \mathbf{l^*_2} \begin{bmatrix} x^1_{s,1} & x^2_{s,1} & x^1_{1,2} & x^2_{1,2} \end{bmatrix}^T.$$

Now, we calculate summation of taxes for the two defenders. Note that under social optimal $\mathbf{x}^*$, we have $x^1_{s,1} = B_1, x^1_{1,2} = 0$, $x^2_{s,1} = B_2$, and $x^2_{1,2} = 0$. Thus, the taxes terms are

$$t^*_1 = 2L_1 B_1 e^{-B_1-B_2} - \mu B_1, \;\; t^*_2 = 2L_2 B_2 e^{-B_1-B_2} - \mu' B_2.$$

For simplicity, suppose that $L_1 = L_2 = L$ and $B_1 = B_2 = B$, we thus have $t^*_1 = 2LBe^{-2B} - \mu B$ and $t^*_2 = 2LBe^{-2B} - \mu' B$. Therefore, summing the taxes of the two defenders yield $\sum_{i=1}^2 t^*_i = 4LBe^{-2B} - B(\mu + \mu')$. Note that if $4Le^{-2B} < \mu + \mu'$ (which can happen under large budget B and small loss L, e.g., $L = 4$ and $B = 50$ yields $4Le^{-2B} = 2.97 \times 10^{-43}$), we would have $\sum_{i=1}^2 t^*_i < 0$. □

B. Proof of Theorem 1

Proof. We prove this result by showing that the amount of taxes paid by defender D_k, given by $t^*_k(\mathbf{x}^*)$ is a decreasing function in the defender D_k's behavioral level α_k. Recall from (9) the tax term of defender D_k under the Externality mechanism in our interdependent security games is given by:

$$t^*_k(\mathbf{x}^*) = \sum_{n=1}^{|D|} \sum_{(v_i,v_j)\in\mathcal{E}_n} t^{kn^*}_{ij} \overset{(10)}{=} -L_k \sum_{n=1}^{|D|} \sum_{(v_i,v_j)\in\mathcal{E}_n} x^{n^*}_{ij} \frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*). \tag{18}$$

Note that the marginal derivative of defender D_k's cost w.r.t. the investment of defender D_n on the edge (v_i, v_j) follows from differentiating (4) and is given by

$$\frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*) = \sum_{v_m\in V_k} L_m \exp\left(-\sum_{(v_i,v_j)\in\overline{P}} [-\log(p_{i,j}(x_{i,j}))]^{\alpha_k}\right) \times \alpha_k[-\log(p_{i,j}(x_{i,j}))]^{\alpha_k-1} \times \frac{p'_{i,j}(x_{i,j})}{p_{i,j}(x_{i,j})},$$

where $\overline{P} = \underset{P\in P_m}{\operatorname{argmax}} \prod_{(v_i,v_j)\in P} w\left(p_{i,j}(x_{i,j})\right)$. Now, differentiating $\frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*)$ w.r.t. α_k yields

$$\frac{d}{d\alpha_k}\frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*) = \sum_{v_m\in V_k} L_m \exp\left(-\sum_{(v_i,v_j)\in\overline{P}} [-\log(p_{i,j}(x_{i,j}))]^{\alpha_k}\right) \times [-\log(p_{i,j}(x_{i,j}))]^{\alpha_k-1} \times \frac{p'_{i,j}(x_{i,j})}{p_{i,j}(x_{i,j})} \times \left(-\sum_{(v_i,v_j)\in\overline{P}} \log(-\log(p_{i,j}(x_{i,j}))) \times [-\log(p_{i,j}(x_{i,j}))]^{\alpha_k} - 1 + \left(\frac{\alpha_k(\alpha_k-1)}{\log(p_{i,j}(x_{i,j}))} \times \frac{p'_{i,j}(x_{i,j})}{p_{i,j}(x_{i,j})}\right)\right).$$

Since $0 < p_{i,j}(x_{i,j}) \le \frac{1}{e}$, we have $1 \le -\log(p_{i,j}(x_{i,j})) < \infty$ and $0 \le \log(-\log(p_{i,j}(x_{i,j}))) < \infty$. Thus, the firt term is negative. Moreover, since $p_{i,j}(x_{i,j})$ is decreasing in the defense investment $x_{i,j}$, we have $p'_{i,j}(x_{i,j}) < 0$, and since $\alpha_k \in (0,1]$ and $-\infty < \log(p_{i,j}(x_{i,j})) \le -1$ (from above), the third term is non-positive. Therefore, the whole term

$$\left(-\sum_{(v_i,v_j)\in\overline{P}} \log(-\log(p_{i,j}(x_{i,j}))) \times [-\log(p_{i,j}(x_{i,j}))]^{\alpha_k}\right) - 1 + \left(\frac{\alpha_k(\alpha_k-1)}{\log(p_{i,j}(x_{i,j}))} \times \frac{p'_{i,j}(x_{i,j})}{p_{i,j}(x_{i,j})}\right)$$

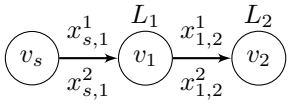


Fig. 14: An attack graph where the Externality mechanism has individual rationality (achieves social optimal solution) but does not have weakly budget balance.

is negative. Finally, from the above analysis and noting that $\exp(x) > 0 \forall x \in (0, \infty)$, the whole term $\sum_{v_m \in V_k} L_m \exp\left(-\sum_{(v_i,v_j)\in \overline{P}} [-\log(p_{i,j}(x_{i,j}))]^{\alpha_k}\right) \times [-\log(p_{i,j}(x_{i,j}))]^{\alpha_k - 1} \times \frac{p'_{i,j}(x_{i,j})}{p_{i,j}(x_{i,j})}$ is negative. Therefore, we have $\frac{d}{d\alpha_k} \frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*) > 0$.

Now, differentiating (18) w.r.t. α_k with noting that the joint investment profile $\mathbf{x}_{-k}$ is fixed yields

$$\frac{d}{d\alpha_k} t^*_k(\mathbf{x}^*) = -L_k \sum_{n=1}^{|D|} \sum_{(v_i,v_j)\in \mathcal{E}_n} x^{n^*}_{ij} \frac{d}{d\alpha_k} \frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*),$$

which is negative since $\frac{d}{d\alpha_k} \frac{\partial C_k}{\partial x^n_{i,j}}(\mathbf{x}^*) > 0$ and since $\exists D_n$ s.t. $x^{n^*}_{i,j} > 0$ for at least one edge (v_i, v_j). □

C. Proof of Proposition 3

Proof. We prove this impossibility by the following counter example with one family of instance game as shown below.
Example 4. Consider the instance of interdependent security game of k rational defenders on the attack graph shown in Figure 15. We now show the details of taxes calculation.

First the PNE solution is given by $\overline{\mathbf{x}} = \left[\frac{\sum_{i=1}^k B_i}{2} \quad \frac{\sum_{i=1}^k B_i}{2} \quad 0 \quad \cdots \quad 0\right]$. In other words, the total budget, which is the sum of the budgets of all defenders, would be distributed equally between the two min-cut edges (v_s, v_1), and (v_s, v_2). For each defender D_i, the total social cost (not counting D_i) is given by

$$\sum_{j=1, j\neq i}^{k} C_j(\overline{\mathbf{x}}) = \left(\sum_{j\neq i} L_j\right) \times \left(e^{-\frac{\sum_{j=1}^k B_j}{2}}\right).$$

Now, if defender D_i was not a member of the society, the equilibrium without defender D_i, denoted by $\mathbf{x}^*_{-\mathbf{i}}$ is given by $\mathbf{x}^*_{-\mathbf{i}} = \left[\frac{\sum_{j=1, j\neq i}^k B_j}{2} \quad \frac{\sum_{j=1, j\neq i}^k B_j}{2} \quad 0 \quad \cdots \quad 0\right]$. Therefore, the amount of tax paid by defender D_i is given by

$$\begin{aligned} t^*_i &= \sum_{j\neq i} C_j(\overline{\mathbf{x}}) - \sum_{j\neq i} C_j(\mathbf{x}^*_{-\mathbf{i}}) \\ &= \left(\sum_{j\neq i} L_j\right) \times \left(e^{-\frac{\sum_{j=1, j\neq i}^k B_j}{2}}\right) \times \left(e^{-\frac{B_i}{2}} - 1\right), \end{aligned}$$

which is negative for each defender D_i with a positive security budget (with $B_i > 0$). Therefore, summing the taxes of all players yields that $t^*_i < 0$. □

This shows the budget deficit under VCG mechanism. Note that we assume that all defenders have finite budget (when securing real-world interdependent systems).

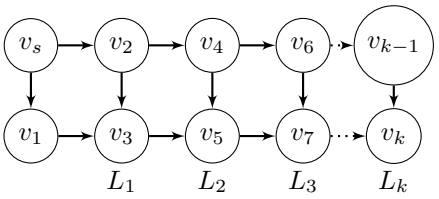


Fig. 15: An example for a graph structure (with k defenders) in which the VCG mechanism achieves the socially optimal allocation but has a budget deficit.

D. Proof of Proposition 4

Proof. From the Proposition statement, the socially optimal solution would put all the budget B on the first edge (v_s, v_1), so that all nodes have a probability of successful attack given by $\exp(-B)$. Now, we prove the first part (i) as follows.

The PNE for behavioral defenders is given by the investment profile where each defender D_k puts their entire budget $\frac{B}{K}$ on the edge coming into their node v_k (similar to analysis in Example 2). Therefore, the true cost of defender D_1 under this PNE is $L_1 \exp(-\frac{B}{K})$.

Now, to have defender D_1 participate in the mechanism we must have

$$\begin{aligned} C_1(\mathbf{x}^*) + t_1 \leq C_1(\overline{\mathbf{x}}) \\ \iff L_1 \exp(-B) + t_1 \leq L_1 \exp(-\frac{B}{K}) \\ \iff t_1 \leq L_1 \left[\exp(-\frac{B}{K}) - \exp(-B)\right]. \end{aligned}$$

For defenders $D_2, D_3, \cdots, D_{K-1}$, defender D_i would participate in the mechanism if

$$\begin{aligned} C_i(\mathbf{x}^*) + t_i \leq C_i(\overline{\mathbf{x}}) &\iff L_i \exp(-B) + t_i \leq L_i \exp(-\frac{iB}{K}) \\ &\iff t_i \leq L_i \left[\exp(-\frac{iB}{K}) - \exp(-B)\right]. \end{aligned}$$

This concludes the proof of the first part.

Now, we prove the second part (ii) From part (i), a defender $D_i \in D$ can participate while paying at most the max amount of tax $t^{max}_i = L_i \left[\exp(-\frac{iB}{K}) - \exp(-B)\right]$. Differentiating t^{max}_i w.r.t the defender index i yields

$$\frac{\partial t^{max}_i}{\partial i} = L_i \times \exp(-\frac{iB}{K}) \times \frac{-B}{K},$$

which is negative since the exponential function range is $(0, \infty)$, L_i is the non-negative financial loss when defender D_i's asset is compromised, and $\frac{B}{K}$ is the non-negative security budget of each defender $D_i \in D$. This concludes the proof of the second part. □

E. Estimation of Baseline Probabilities of Successful Attack

We show the estimation of baseline probability of successful attack in Table II. The first column represent the vulnerability CVE-ID (from real-world vulnerabilities reported in CVE database). The second column represent the corresponding edge(s) in the attack graph. The third column represents the

attack vector type (physical, local, or network). The fourth column is the score generated following the seminal work [34].

TABLE II: Baseline probability of successful attack for vulnerabilities in SCADA and DER.1 systems.

Vulnerability (CVE-ID)	Edge(s)	Attack Vector	Score
SCADA application			
Control Unit (CVE-2018-5313)	(Vendor,Control1),(Vendor,Control2)	Local	0.78
Remote authentication (CVE-2010-4732)	(S, Vendor)	Network	0.9
Remote cmd injection (CVE-2011-1566)	(Control,RTU1),(Control,RTU2)	Network	1.0
Authentication bypassing (CVE-2019-6519)	(Corp,DMZ1),(Corp,DMZ2)	Network	0.75
DER.1 application			
Physical access (CVE-2017-10125)	(w_9, w_7),(w_{18}, w_{16})	Physical	0.71
Network access (CVE-2019-2413)	(w_9, w_8),(w_{18}, w_{17})	Network	0.61
Software access (CVE-2018-2791)	(w_7, w_6),(w_8, w_6)	Network	0.82
Sending cmd (CVE-2018-1000093)	(w_6, w_5),(w_{15}, w_{14})	Network	0.88

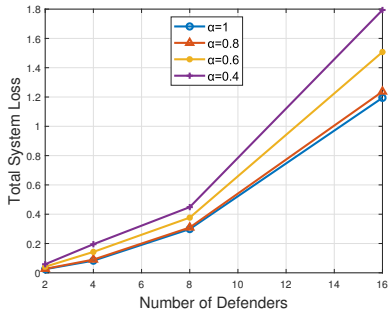


Fig. 16: Total loss (in Millions) as a function of the number of defenders. We observe that the loss increases super-linearly (i.e., the per-defender loss is increasing as system size grows). This is due to the increased risks resulting from interdependencies in the defenders' critical assets.

F. Evaluation-Extended

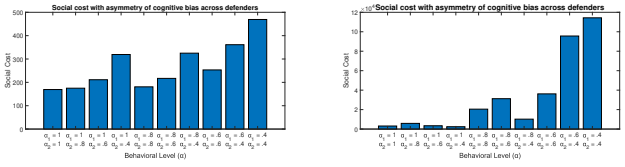
Here, we extend our evaluation presented in Section V. Specifically, we study the effect of the asymmetry in behavioral cognitive bias and asymmetry in edges sensitivities to security investments across different defenders.

Asymmetry in Cognitive Bias across Defenders: We study the effect of the asymmetry in behavioral level (cognitive bias) across different defenders and its effect on the social cost, for the four case studies we consider in our evaluation. Figure 17a and Figure 17b shows the result for such experiment for the DER and SCADA systems, respectively. We observe that the difference in the social cost is more pronounced with very high bias asymmetry, e.g., the social cost with $\alpha_1 = 1, \alpha_2 = 0.4$ is about 2X the social cost with $\alpha_1 = 1, \alpha_2 = 1$ for DER. A similar insight observed for the SCADA systems. Moreover, we notice that two defenders with moderate behavioral levels ($\alpha_1 = 0.6, \alpha_2 = 0.6$) would invest better and consequently better protect the overall system compared to two defenders with one less behavioral and one more behavioral (with $\alpha_1 = 0.8, \alpha_2 = 0.4$), although the mean behavioral level is the same in these two cases. This sheds the light on the effect of suboptimality of behavioral defender on the overall system.

Sensitivity of Edges to Investments: Finally, we consider the effects of different sensitivities of edges to security investments. Recall from (2) that edges with higher sensitivity are those for which the probability of successful attack decreases faster with each unit of security investment. In this experiment, for both DER and SCADA systems, we assume that the defender D_1 has lower edges' sensitivities to her investments compared to the defender D_2. Formally, we let $s^1_{i,j} = 0.5$ (for D_1) and $s^2_{i,j} = 1$ (for D_2). That can be mapped into realistic scenario where D_1's methods for investing on edges are less effective in reducing the probability of successful attack compared to D_2.

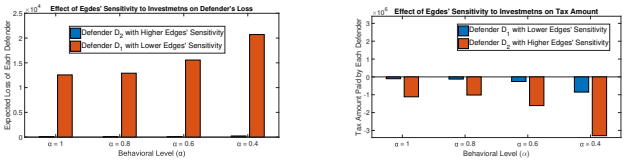
For DER system, we show the effect of edges' sensitivity to investments on each defender's real expected loss for different behavioral levels in Figure 18a. We observe that the defender with the higher edges' sensitivity (here, D_1) would have much lower expected loss compared to the defender with the lower edges' sensitivity (here, D_2) irrespective of the behavioral level of the defender. However, both defenders pay zero amount of taxes under all behavioral levels due to loose interdependency across the two defenders' subnetworks in DER (as explained earlier in Section VI).

For SCADA system, we show the effect of sensitivity of edges to investments on the amount of taxes paid by each defender under the VCG mechanism. Figure 18b shows that D_2 would receive more amount of payments compared to D_1 for all behavioral levels. The intuition here is that D_2 is more beneficial to the society as her investments reduce the social cost more compared to the investments of D_1 (since the sensitivity of edges to D_2's investments is twice the sensitivity of edges to D_1's investments). Therefore, under the VCG mechanism, D_2 would receive much more amount due to her contribution to the society. Moreover, we note that the effect of edges' sensitivity is more pronounced under higher behavioral bias (i.e., less α) and therefore the difference in the amount of taxes among the two defenders increases as defenders become more behavioral (since D_1 even wastes her budget on edges that has less sensitivity to those non-critical edges).



(a) DER (b) SCADA

Fig. 17: The effect of asymmetry in behavioral cognitive bias across defenders on the social cost.



(a) Effect of asymmetry in edges' sensitivity to investments across the two defenders on the loss of each defender on DER system.

(b) Effect of asymmetry in edges' sensitivity to investments across the two defenders on tax amount under VCG mechanism on SCADA system.

Fig. 18: The effect of asymmetry in edges' sensitivity to investments across the two defenders on the loss of each defender and the amount of taxes paid by the defender under the VCG mechanism.