games

MDPI

Article

Interdependent Defense Games with Applications to Internet Security at the Level of Autonomous Systems †

Hau Chan [1], Michael Ceyko [2,3] and Luis Ortiz [4,*]

1 Department of Computer Science, Trinity University, San Antonio, TX 78212, USA; hchan@trinity.edu
2 Department of Computer Science, Stony Brook University, Stony Brook, NY 11794, USA; mike@ceyko.net
3 Tumblr, 35 E 21 Street, Ground Floor, New York, NY 10010, USA
4 Department of Computer and Information Science, University of Michigan-Dearborn, Dearborn, MI 48128, USA
* Correspondence: leortiz@umich.edu; Tel.: +1-613-593-5239

† This paper is an extended version of our paper published in Hau Chan and Luis E. Ortiz. Computing Nash Equilibria in Interdependent Defense Games. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, Texas, USA, 25–30 January 2015; Bonet, B., Koenig, S., Eds.; AAAI Press, 2015; pp. 842–850; and Hau Chan, Michael Ceyko and Luis E. Ortiz. Interdependent Defense Games: Modeling Interdependent Security under Deliberate Attacks. In Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, USA, 14–18 August 2012; de Freitas, N., Murphy, K.P., Eds.; AUAI Press, 2012; pp. 152–162. The first and third authors contributed equally to this work. The second author contributed substantially, but less than the first and third author overall.

Academic Editor: Christos Dimitrakakis
Received: 1 October 2016; Accepted: 28 December 2016; Published: 16 February 2017

Abstract: We propose *interdependent defense* (*IDD*) *games*, a computational game-theoretic framework to study aspects of the interdependence of risk and security in multi-agent systems under deliberate external attacks. Our model builds upon *interdependent security* (*IDS*) *games*, a model by Heal and Kunreuther that considers the source of the risk to be the result of *a fixed randomized-strategy*. We adapt IDS games to model *the attacker's deliberate behavior*. We define the attacker's pure-strategy space and utility function and derive appropriate cost functions for the defenders. We provide a complete characterization of mixed-strategy Nash equilibria (MSNE), and design a simple *polynomial-time* algorithm for computing *all* of them for an important subclass of IDD games. We also show that an efficient algorithm to determine whether some attacker's strategy can be a part of an MSNE in an instance of IDD games is unlikely to exist. Yet, we provide a *dynamic programming* (*DP*) algorithm to compute an approximate MSNE when the graph/network structure of the game is a directed tree with a single source. We also show that the DP algorithm is a *fully polynomial-time approximation scheme*. In addition, we propose a generator of random instances of IDD games based on the real-world Internet-derived graph at the level of autonomous systems (≈27 K nodes and ≈100 K edges as measured in March 2010 by the DIMES project). We call such games Internet games. We introduce and empirically evaluate two heuristics from the literature on learning-in-games, *best-response gradient dynamics* (*BRGD*) and *smooth best-response dynamics* (*SBRD*), to compute an approximate MSNE in IDD games with arbitrary graph structures, such as randomly-generated instances of Internet games. In general, preliminary experiments applying our proposed heuristics are promising. Our experiments show that, while BRGD is a useful technique for the case of Internet games up to certain approximation level, SBRD is more efficient and provides better approximations than BRGD. Finally, we discuss several extensions, future work, and open problems.

Keywords: computational game theory; interdependent security; equilibrium computation; equilibrium characterization; fully polynomial-time approximation scheme; learning in games; autonomous-systems internet security application

1. Introduction

Attacks carried out by hackers and terrorists in recent decades have led to increased efforts by both government and the private sector to create and adopt mechanisms to prevent future attacks. This effort has yielded a more focused research attention to models, computational and otherwise, that facilitate and help to improve (homeland) security for both physical infrastructure and cyberspace. In particular, there has been quite a bit of recent research activity in the general area of game-theoretic models for terrorism settings (see, e.g., Bier and Azaiez [1] and Cárceles-Poveda and Tauman [2]).

Interdependent security (*IDS*) *games* are one of the earliest models resulting from a game-theoretic approach to model security in non-cooperative environments composed of free-will self-interested individual decision-makers. Originally introduced and studied by economists Kunreuther and Heal [3], IDS games model general abstract security problems. In those problems, an individual within a population considers whether to voluntarily invest in some protection mechanisms or security against a risk they may face. Individuals do so knowing that the cost-effectiveness of the decision depends on the investment decisions of others in the population. This is because of transfer risks; that is, the "bad event" may be transferable from a compromised individual to another.

In their work, Kunreuther and Heal [3] provided several examples based on their economics, finance, and risk management expertise. We refer the reader to their paper for more detailed descriptions. As a canonical example of the real-world relevance of IDS settings and the applicability of IDS games, Heal and Kunreuther [4] used this model to describe problems such as airline baggage security. In their setting, individual airlines may choose to invest in additional complementary equipment to screen passengers' bags and check for hazards such as bombs that could cause damage to their passengers, planes, buildings, or even reputations. However, mainly due to the large amount of traffic volume, it is impractical for an airline to go beyond applying security checks to bags incoming from passengers and include checks to baggage or cargo transferred from other airlines. On the other hand, if an airline invests in security, they can still experience a bad event if the bag was transferred from an airline that does not screen incoming bags, rendering their investment useless. [1] Thus, we can see how the cost-effectiveness of an investment can be highly dependent on others' investment decisions. Another recent application of the IDS model is on container shipping transportation [6], in which the objective is to study the effect that investment decisions about container screening on some ports may have on neighboring ports.

Some security-related problems in cyberspace are similar, but slightly different in nature to the airline scenario just described. Consider a network where all computers fully trust all other computers and freely exchange information. Each user has complete control over his own computer and can decide if he wants to protect the individual user's computer from hackers, by installing a firewall, for example. However, that individual user cannot directly control or impose others in the network to protect themselves too. Thus, in order for an individual to feel secure about storing his information on the network, that individual user not only has to think about the security of his own computer, but also the security of *other* computers on the internal network. This is because any other computer may

[1] Note that even if full screening were performed, the Christmas Day 2009 episode in Detroit [5] serves as a reminder that transfer risk still exists.

access that individual user's computer as well. If any computer were hacked, that individual user's information would potentially be exposed to the outside world.

Two potential outcomes immediately arise out of the cyber-security scenario. If one does not think enough people have invested in security, then one will not invest either, because any investment will contribute negligibly to the overall protection of one's data. Also, and this is the aspect that perhaps differentiates cyberspace from the airline security scenario, if nearly everyone has invested in security, one may no longer feel the need to protect oneself. This is because the network is already mostly secure and the amount of work required to protect oneself outweighs the minimal change in overall security. Thus, as many invest, fewer may want to invest.

In this work, we build on the literature in IDS games. In particular, we adapt the model to situations in which the abstract "bad event" results from the *deliberate* action of an attacker. The "internal agents" (e.g., airlines and computer network users or administrators), whom we also often refer to as "defenders" or "sites," have the voluntary choice to individually invest in security to defend themselves against a direct or indirect offensive attack, modulo, of course, the cost-effectiveness to do so. As a result, we formally define a new model class: *interdependent defense* (*IDD*) *games*.

In our model, both the attacker(s) [2] and the defenders, modeled as players in a game, make decisions based on cost-benefit analysis. The attacker wants to find the most cost-effective way to attack nodes in the network, and does so by maximizing explicit personal preferences. Similarly, each defender (node in a network) takes into account the costs, as well as potential losses, risks, and actions not only of the attacker, but also of other nodes in the network, when making security investment decisions.

A side benefit of explicitly modeling the attacker, as we do in our model, is that the probability of an attack results directly from the equilibrium analysis. Building IDS games can be hard because it requires *a priori* knowledge of the likelihood of an attack. Attacks of this kind are considered rare events and thus notoriously difficult to statistically estimate in general.

1.1. Related Work

Johnson et al. [7] and Fultz and Grossklags [8] independently developed non-cooperative game models similar to ours. Johnson et al. [7] extend IDS games by modeling uncertainty about the source of the risk (i.e., the attacker) using a Bayesian game over risk parameters. Fultz and Grossklags [8] propose and study a non-graphical game-theoretic model for the interactions between attackers and nodes in a network. In their model, each node in the network can decide whether to contribute (by investment) to the overall safety of the network or to individual safety. The attackers can attack any number of nodes, but with each attack there is an increased probability that the attacker might get caught and suffer penalties or fines. Hence, while their game has IDS characteristics, it is technically not within the standard IDS game framework introduced by Heal and Kunreuther.

Most of the previous related work explores the realm of information security and is application/network specific (see Roy et al. [9] for a survey on game theory application to network security). Syverson and Systems [10] suggest the use of game-theoretic models (non-cooperative or cooperative) to model the relationship between the attacker and the nodes in the network. Past literature has largely focused on two-person (an attacker and a defender) games where the nodes in the network are regarded as a single entity (or a central defender). For example, Lye and Wing [11] look at the interactions between an attacker and the (system) administrator using a two-player stochastic game. Recent work uses a Stackelberg game model in which the defender (or leader) commits to a mixed strategy to allocate resources to defend a set of nodes in the network, and the follower (or

---

[2] Throughout this article, we often used "attacker(s)" and "aggressor(s)" interchangeably as a way to remind the reader that our model handles a variety of interdependent security settings beyond airline or Internet infrastructure security.

attacker) optimally allocates resources to attack a set of "targets" in the network given the leader's commitment [12–16].

Very recent work by Smith et al. [17] and Lou and Vorobeychik [18] extends the traditional Stackelberg settings to multiple leaders and use different equilibrium concepts than MSNE. Laszka et al. [19] have used their model to study the interaction among an attacker and multiple defenders in spear-phishing attack settings. The main distinctions of their work to ours are that the defenders are not interconnected, the attacker has some fixed number of attacks (more than one), and the equilibrium concepts studied are different.

Other work strives to understand the motivation of the attackers. For example, Liu [20] focuses on understanding the attacker's intent, objectives, and strategies, and derive a (two-player) game-theoretic model based on those factors. As another example, Cremonini and Nizovtsev [21] use cost-benefit analysis (of attackers) to address the issue of the optimal amount of security (of the nodes in the network).

1.2. Brief Overview of the Article and the Significance of Our Contributions

We adapt the standard non-cooperative framework of IDS games, which we present and briefly discuss in Section 2, to settings in which the source of the risk is the result of a deliberate, strategic decision by an external attacker. In particular, we design and propose *interdependent defense (IDD) games*, a new class of games that, in contrast to standard IDS games, model the attacker *explicitly*, while maintaining a core component of IDS systems: the potential *transferability* of the risk resulting from an attack. We note that the explicit modeling of risk transfer is an aspect of our model that has not been a focus of previous game-theoretic attacker-defender models of security discussed earlier in Section 1.1.

We formally define and study IDD games in depth in Section 3. There, we also present some characterizations of their MSNE which have immediate computational and algorithmic implications.

In Section 4, we study several computational questions about IDD games. We first provide a *polynomial-time* algorithm to compute *all* MSNE for an important subclass of IDD games in Section 3.4. In that subclass, there is only one attack, the defender nodes are *fully transfer-vulnerable* (i.e., investing in security does nothing to reduce their external/transfer risk), and transfers are *one-hop*. [3] We describe this subclass in more detail in Section 3.4.

Before continuing, we would like to address two aspects of IDD games brought up in the last paragraph. Note that considering a single attacker is a typical assumption in security settings (see previous work discussed earlier in Section 1.1). It is also reasonable because we can view many attackers as a single attacker. Allowing at most one attack prevents immediate representational and computational intractability problems because, as we state at the beginning of Section 3.2, the number of the attacker's pure strategies grows *exponentially* with the number of attacks. Finally, because the attacker has no fixed target, it seems practically ineffective for the attacker to consider or go beyond plans of attacks involving multiple (>2) transfers: such plans are complex, time consuming, and costly. Having said that, there has been increased interest over the last few years in the network security community to explicitly model multiple attackers [22]. For instance, Merlevede and Holvoet [22] view the multi-attacker setting as a very important future direction. Here, we focus most of our technical results and experiments to the single-attacker setting. Yet, we present multi-attacker extensions of our proposed model in Appendix E, where we also discuss some ideas and very preliminary technical results for the multi-attacker setting.

In Section 4.2, we formally prove that our results for computing all MSNE in a subclass of IDS games in polynomial time is unlikely to extend to arbitrary IDD games. Given that, we move on to explore approximate MSNE in IDD games in Section 4.3 and provide a *fully polynomial-time approximation scheme (FPTAS)* for the case in which the graph over the sites is a directed tree-like

---

[3] We note that the original IDS games were also *fully transfer-vulnerable* and assumed one-hop transfers.

network. We note that the attacker is still connected to every site in the network. To place the significance of this result in context, we note that despite the apparent simplicity of the subgraph over the sites, there may be very important real-world applications in supply chains (e.g., see [23]) and the power grid [24], for example.

Our computational results in Section 4 are significant, and some initially surprising to us, within the context of the state of the art in computational and algorithmic game theory. Computing all MSNE in graphical IDS games is hard in general. For instance, the so-called *Nash-extension computational problem* in general IDS games is NP-complete [25]. To place our computational contributions in an even broader context, note that deciding whether an arbitrary graphical game has a *pure-strategy Nash equilibrium* (*PSNE*) is in general NP-complete [26]. In addition, many problems related to computing MSNE with particular properties, even in normal-form games, are NP-complete [27]. Indeed, computing an MSNE is PPAD-complete, even in two-player games [28], thus considered computationally intractable in general. Some alternative proofs of PPAD-completeness for two-player games are polynomial time reductions from graphical games, even if each player in the graphical game has at most 3 players [29]. We refer the reader to Papadimitriou [30] for more information on the complexity class PPAD and to Daskalakis et al. [31] for high-level information about the most recent computational results on the complexity of computing MSNE in normal-form games, and indirectly graphical games too. Also, computing *all* MSNE is rarely achieved and counting-related problems are often #P-complete. We refer the reader to, for example, Conitzer and Sandholm [32], and the references therein, for additional information. We do not know of any other *non-trivial* game for which there exists a *polynomial-time* algorithm to compute *all* MSNE except the one we provide in Section 4.1 here and the algorithm for uniform-transfer IDS games of Kearns and Ortiz [25]. Finally, while our hardness results build on those for graphical IDS games [25], there does not exist any analogous to our FPTAS for graphical IDS games. It seems that an FPTAS in the IDS setting may be possible by modifying the one we design here, and its proof, to that setting.

We provide experimental results in Section 5. In our experiments, we study the application of learning-in-games heuristics to compute approximate MSNE to both fixed and randomly-generated instances of IDD games. We focus on the class of games with at most one simultaneous attack and one-hop transfers. Our particular object of study is a very large Internet-derived graph at the level of *autonomous systems (AS)* ($\approx$27 K nodes and $\approx$100 K edges) obtained from DIMES [33,34] for March 2010, the last network graph available to us. Cybersecurity scenarios motivate this study. We refer the reader to Roy et al. [9] for some examples. Here, we propose a generative model of single-attack IDD games based on the aforementioned Internet graph. For simplicity, we refer to the models that a simulator that we built based on the generative model outputs as *Internet games* (*IGs*). In our experiments, we employ simple best-response heuristics from learning in games [35] to compute (approximate) MSNE in IGs. In particular, we perform a series of experiments to both show the large-scale feasibility and scalability of the model and approach. We also explore the behavior of the internal players and the attacker in the resulting equilibria, and the properties of the network-structure *induced* at an equilibrium.

In Section 6, we provide a discussion of future work, some open problems, and a summary of our contributions.

2. Interdependent Security Games, and a Generalization

Each *player* i in a *finite set* $[n] \equiv \{1, 2 \ldots, n\}$ of n players of an IDS game has a binary choice, *to invest* ($a_i = 1$) or *not to invest* ($a_i = 0$) in security mechanisms to protect themselves from a potential *bad event*. For each player i, the parameters C_i and L_i correspond to the *cost of investment and loss induced by the bad event*, respectively. We define the ratio of the two parameters, the player's "cost-to-loss" ratio, as $R_i \equiv C_i / L_i$. Bad events can occur through both *direct* and *indirect* means. The *direct risk*, or *internal risk*, parameter p_i, is the probability that player i will experience a bad event because of direct contamination. The standard IDS model assumes that investing will completely protect the player from direct contamination; hence, internal risk is only possible when $a_i = 0$. The *indirect-risk*

parameter q_{ji} is the probability that player j is directly "contaminated," does not experience the bad event, but transfers it to player i who ends up experiencing the bad event. This discussion leads us to the following definition of standard IDS games.

Definition 1. (Standard IDS Games) *An* IDS game *is defined by a tuple* $(n, \mathbf{C}, \mathbf{L}, \mathbf{p}, \mathbf{Q})$*, where* $\mathbf{C} \equiv (C_i)_{i\in[n]}$*,* $\mathbf{L} \equiv (L_i)_{i\in[n]}$*,* $\mathbf{p} \equiv (p_i)_{i\in[n]}$*,* $\mathbf{Q}$ *is a matrix representation of the* q_{ij}*'s, where* $(i, j) \in [n]^2$*. Implicit in Definition 1 is that* $q_{ii} = 0$ *for all* i*.*

Before we provide a formal definition of the *model semantics* in the upcoming paragraphs in this section, as a preview, we want to highlight that the standard IDS model assumes that the interactions between players are unaffected by investment in security. Said differently, each individual player's *transferred risk* is the same regardless of whether the player invests in security, or not.

We now formally define a *(directed) graphical-games* [36,37] version of IDS games, as first introduced by Kearns and Ortiz [25]. The reason we introduce graphical IDS games here is to emphasize that the matrix $\mathbf{Q}$ is often sparse and their *non-zero entries* lead to a *network structure* captured by an *induced graph* $G = ([n], E)$. That way the representation size of a graphical IDS game is not n^2, as in standard IDS games, but essentially the number of *directed* edges E of G, which could potentially be significantly smaller than n^2. In addition, we can exploit the sparse representation and the network structure to provide provably tractable computational solutions in some cases.

Definition 2. (Graphical IDS Games) *The parameters* q_{ij}*'s induce a directed graph* $G = ([n], E)$ *such that* $E \equiv \{(i, j)|q_{ij} > 0\}$*. Indeed, we assume that* $\mathbf{Q}$ *has a sparse-matrix representation as a list of non-zero* q_{ij} *values for each edge* $(i, j) \in E$*. Thus, the representation size of* $\mathbf{Q}$ *is* $O(|E|)$*. A* graphical IDS game *is defined by the tuple* $(n, G, \mathbf{C}, \mathbf{L}, \mathbf{p}, \mathbf{Q})$*.*

We now discuss the game-theoretic *semantics* of (graphical) IDS games. For each player $i \in [n]$, let $\mathrm{Pa}(i) \equiv \{j \in [n] \mid q_{ji} > 0\}$ be the set of players that are *parents* of player i in G (i.e., the set of players that player i is exposed to via transfers); and $\mathrm{PF}(i) \equiv \mathrm{Pa}(i) \cup \{i\}$ be the *parent family* of player i, which *includes* i. Denote by $k_i \equiv |\mathrm{PF}(i)|$ the *size of the parent family* of player i. Similarly, let $\mathrm{Ch}(i) \equiv \{j \in [n] \mid q_{ij} > 0\}$ be the set of players that are *children* of player i (i.e., the set of players to whom player i can present a risk via transfer) and $\mathrm{CF}(i) \equiv \mathrm{Ch}(i) \cup \{i\}$ the *(children) family* of player i, which *includes* i. The *probability that player* i *is safe from player* j, as a function of player j's decision, is

$$e_{ij}(a_j) \equiv a_j + (1 - a_j)(1 - q_{ji}) = (1 - q_{ji})^{1-a_j} \,. \tag{1}$$

Equation (1) results from noting that if j invests, then it is impossible for j to transfer the bad event, while if j does not invest, then j either experiences the bad event or transfers it to another player, but never both. [4] The player receiving the transfer still has the chance of not experiencing the bad event. However, without some form of screening of transfers, this chance is usually very low.

Denote by $\mathbf{a} \equiv (a_1, \ldots, a_n) \in \{0, 1\}^n$ the *joint action* of all n players. Also denote by $\mathbf{a}_{-i}$ the *joint action of all players except* i, and for any subset $I \subset [n]$ of players, denote by $\mathbf{a}_I$ the *sub-component of the joint action corresponding to those players in* I *only.* We define i's *overall safety* from *all* other players as

$$s_i(\mathbf{a}_{\mathrm{Pa}(i)}) \equiv \textstyle\prod_{j\in\mathrm{Pa}(i)} e_{ij}(a_j) \,, \tag{2}$$

and equivalently the *overall risk* from *some* other players as

$$r_i(\mathbf{a}_{\mathrm{Pa}(i)}) \equiv 1 - s_i(\mathbf{a}_{\mathrm{Pa}(i)}) \,. \tag{3}$$

[4] Or as Heal and Kunreuther [38] put it, "You Only Die Once."

Note that, as given in Equation (2) (and Equation (3)), each players' external safety (and risk) is a direct function of its parents only, *not* all other players. Note also that from the perspective of each players' preferences, as quantified by their cost functions presented next (Equation (4)), the source of each player's transfer risk is independent over the player's parents in the graph. This independence assumption comes from the original definition of standard IDS games, but there the graph is fully connected (i.e., $\mathrm{Pa}(i) = [n] - \{i\}$) [3]. From these definitions, we obtain player i's *overall cost function*: the cost of joint action $\mathbf{a} \in \{0,1\}^n$, corresponding to the (binary) investment decision of all players, is

$$M_i(a_i, \mathbf{a}_{\mathrm{Pa}(i)}) \equiv a_i[C_i + r_i(\mathbf{a}_{\mathrm{Pa}(i)})L_i] + (1 - a_i)[p_i + (1 - p_i)r_i(\mathbf{a}_{\mathrm{Pa}(i)})]L_i \,. \tag{4}$$

Whether players invest depends solely on what they can gain or lose by investing. If the overall cost of investing is less than the overall cost of not investing, the player will invest. Applying this logic to cost function M_i in Equation (4), player i will invest if

$$C_i + r_i(\mathbf{a}_{\mathrm{Pa}(i)})L_i < [p_i + (1 - p_i)r_i(\mathbf{a}_{\mathrm{Pa}(i)})]L_i \tag{5}$$

so that the investment cost and the losses due to a transferred event do not outweigh the losses from an internal or transferred bad event. Similarly, if the inequality in the last expression is reversed or is replaced by equality, player i will not invest or would be indifferent, respectively. Rearranging the expression for the best-response condition for strictly playing $a_i = 1$, given in the last equation (Equation (5)), and letting $\Delta_i \equiv R_i/p_i = \frac{C_i}{p_i L_i}$, the *cost-to-expected-loss ratio* of player i, we get the following *best-response correspondence* $\mathcal{BR}_i : \{0,1\}^{k_i - 1} \to 2^{\{0,1\}}$ for player i: [5] for all $\mathbf{a}_{\mathrm{Pa}(i)} \in \{0,1\}^{k_i - 1}$,

$$\mathcal{BR}_i(\mathbf{a}_{\mathrm{Pa}(i)}) \equiv \begin{cases} \{1\}, & \text{if } s_i(\mathbf{a}_{\mathrm{Pa}(i)}) > \Delta_i, \\ \{0\}, & \text{if } s_i(\mathbf{a}_{\mathrm{Pa}(i)}) < \Delta_i, \\ \{0,1\}, & \text{if } s_i(\mathbf{a}_{\mathrm{Pa}(i)}) = \Delta_i. \end{cases} \tag{6}$$

In other words, whether it is cost-effective for player i to invest or not depends on a simple threshold condition on the player's safety: Does the player feel safe enough from others?

Definition 3. *A joint-action* $\mathbf{a}^* \in \{0,1\}^n$ *is a* pure-strategy Nash equilibrium (PSNE) of a graphical IDS game $(n, G, \mathbf{C}, \mathbf{L}, \mathbf{p}, \mathbf{Q})$ *(see Definition 2) if* $a_i^* \in \mathcal{BR}_i(\mathbf{a}^*_{\mathrm{Pa}(i)})$ *for all players i (i.e.,* $\mathbf{a}^*$ *is a mutual best-response, as defined in Equation (6)).*

2.1. Generalized IDS Games

In the standard IDS game model, investment in security does not reduce transfer risks. However, in some IDS settings (e.g., vaccination and cyber-security), it is reasonable to expect that security investments would include mechanisms to reduce transfer risks. This motivates our first modification to the traditional IDS games: allowing the investment in protection to not only make us safe from direct attack but also partially reduce (or even eliminate) the transfer risk. Thus, we introduce a new real-valued parameter $\alpha_i \in [0,1]$ representing the probability that a transfer of a potentially bad event will go *unblocked* by i's security, assuming i has invested. Thus, we redefine player i's overall cost as [6]

$$M_i(a_i, \mathbf{a}_{\mathrm{Pa}(i)}) \equiv a_i[C_i + \alpha_i r_i(\mathbf{a}_{\mathrm{Pa}(i)})L_i] + (1 - a_i)[p_i + (1 - p_i)r_i(\mathbf{a}_{\mathrm{Pa}(i)})]L_i \,. \tag{7}$$

We call the generalization $\boldsymbol{\alpha}$*-IDS games*, where $\boldsymbol{\alpha} \equiv (\alpha_1, \alpha_2, \ldots, \alpha_n)$ corresponds to the vector composed of the parameter values of each player. This discussion leads to the following definition.

---

[5] By $2^{\{0,1\}}$ we mean the *power set* of $\{0,1\}$ which equals $\{\emptyset, \{0\}, \{1\}, \{0,1\}\}$.

[6] A similar extension was also proposed independently by Heal and Kunreuther [39].

Definition 4. *A* graphical $\boldsymbol{\alpha}$-IDS game, *or simply* $\boldsymbol{\alpha}$-IDS game, *is given by a tuple* $(n, G, \mathbf{C}, \mathbf{L}, \mathbf{p}, \mathbf{Q}, \boldsymbol{\alpha})$, *where each tuple-entry is as defined in the discussion above, and the semantics of the cost functions* M_i*'s of the players is as defined in Equation (7).*

From a game-theoretic perspective, the key aspect of the $\boldsymbol{\alpha}$ parameters is that they determine the characteristics of the best-response behavior of each player. That is, it allows us to model players that may behave in a way that is consistent with behavior that ranges from *strategic complementarity* (e.g., airline setting, where $\alpha_i = 1$), all the way to *strategic substitutability* (e.g., vaccination setting, where $\alpha_i = 0$), based on the relationship between α_i and $1 - p_i$.

The corresponding definition of PSNE for $\boldsymbol{\alpha}$-IDS games is analogous to that given in Definition 3. Hence, we do not formally re-state it here.

3. Interdependent Defense Games

Building from generalized IDS games, in this section we introduce *interdependent defense (IDD) games*. We begin by introducing an additional player, the *attacker*, who *deliberately* initiates bad events: now bad events are no longer "chance occurrences" without any strategic deliberation. [7] The attacker has a *target decision* for each player - a choice of attack ($b_i = 1$) or not attack ($b_i = 0$) player i. Hence, the attacker's pure strategy is denoted by the vector $\mathbf{b} \in \{0, 1\}^n$. (We discuss an extension of the model to multiple attackers in Appendix E.)

Changing from "random" *non-strategic* attacks, whose probability of occurrence is determined independent of the actions of the internal players, to *intentional* attacks, in which actions are deliberately carried out by an external actor, leads us to alter p_i and q_{ij}. This is because their original definitions imply extra meaning with respect to the new aggressor.

The game parameter p_i implicitly "encodes" b_i because $b_i = 0$ implies $p_i = 0$. Thus, we redefine

$$p_i \equiv p_i(b_i) \equiv b_i \widehat{p}_i \tag{8}$$

so that player i has *intrinsic risk* $\widehat{p}_i$, and only has *internal risk* if targeted (i.e, $b_i = 1$). The new parameter $\widehat{p}_i$ represents the (*conditional*) probability that an attack is successful at site player i *given* that site i was directly targeted and did not invest in protection.

Similarly, the game parameter q_{ij} "encodes" $b_i = 1$, because a prerequisite is that i is targeted before it can transfer the bad event to j. We redefine

$$q_{ij} \equiv q_{ij}(b_i) \equiv b_i \widehat{q}_{ij} \tag{9}$$

so that $\widehat{q}_{ij}$ is the intrinsic transfer probability from player i to player j, independent of b_i. The new parameter $\widehat{q}_{ij}$ represents the (*conditional*) probability that an attack is successful at player j *given* that it originated at player i, did not occur at i, but was transferred undetected to j.

Because the p_i's and q_{ij}'s depend on the attacker's action $\mathbf{b}$, so do the safety and risk functions. In particular, we now have

$$e_{ij}(a_j, b_j) \equiv a_j + (1 - a_j)(1 - b_j \widehat{q}_{ji}) = (1 - \widehat{q}_{ji})^{b_j(1-a_j)} \text{ , and} \tag{10}$$

$$s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) \equiv \prod_{j \in \mathrm{Pa}(i)} e_{ij}(a_j, b_j) \equiv 1 - r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) \text{ ,} \tag{11}$$

---

[7] By "strategic" here we mean that the action of an individual entity may depend on those of others in the population.

where $\widehat{p}_i$ and $\widehat{q}_{ji}$ are as defined in Equations (8) and (9), respectively. Hence, for each site player i, the *cost* function becomes

$$M_i(a_i, \mathbf{a}_{\mathrm{Pa}(i)}, b_i, \mathbf{b}_{\mathrm{Pa}(i)}) \equiv a_i[C_i + \alpha_i r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})L_i] + (1 - a_i)[b_i\widehat{p}_i + (1 - b_i\widehat{p}_i)r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})]L_i, \tag{12}$$

where $\widehat{p}_i$ and r_i are as defined in Equations (8) and (11), respectively. Let

$$\widehat{\Delta}_i \equiv R_i / \widehat{p}_i \equiv \frac{C_i}{\widehat{p}_i L_i} \tag{13}$$

and

$$\widehat{s}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) \equiv b_i\, s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) + \frac{1 - \alpha_i}{\widehat{p}_i} r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) \ , \tag{14}$$

where s_i is as defined in Equation (11). The *pure-strategy best-response correspondence of each site i* is

$$\mathcal{BR}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) \equiv \begin{cases} \{1\}, & \text{if } \widehat{s}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) > \widehat{\Delta}_i, \\ \{0\}, & \text{if } \widehat{s}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) < \widehat{\Delta}_i, \\ \{0,1\}, & \text{if } \widehat{s}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) = \widehat{\Delta}_i \end{cases} \ , \tag{15}$$

where $\widehat{\Delta}_i$ and $\widehat{s}_i$ are as defined in Equations (13) and (14), respectively.

We assume that the attacker wants to cause as much damage as possible. Here, we define the *utility/payoff* function U quantifying the objective of the attacker as

$$U(\mathbf{a}, \mathbf{b}) \equiv \textstyle\sum_{i=1}^{n} M_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) - a_i C_i - b_i C_i^0 \ , \tag{16}$$

where C_i^0 is *the attacker's own "cost" to target player i.* [8]

Of course, many other utility functions of varied complexity are also possible. Indeed, one can consider increasingly complex and sophisticated utility functions that may explicitly parse out the involved costs and induced losses in finer-grain and painstaking detail. For instance, we could decompose the cost to the attacker to target a specific site into different components such as, perhaps, planning and setup costs, carry-out costs, and the costs of getting caught or retaliated against, to name a few. We leave those more complex variants for future work.

Definition 5. *A* single-attacker graphical IDD game, *or simply* IDD game, *is given by the tuple* $(n, G, \mathbf{C}, \mathbf{L}, \mathbf{C}^0, \widehat{\mathbf{p}}, \widehat{\mathbf{Q}}, \boldsymbol{\alpha})$, *where the tuple's entries, as well as the model semantics, are as defined in the preceding discussion (Equations (8), (9), (12) and (16)), and the matrix* $\widehat{\mathbf{Q}}$ *is analogous to the matrix* $\mathbf{Q}$ *described in Definition 2.*

The attacker's *pure-strategy best-response correspondence* $\mathcal{BR}_0 : \{0,1\}^n \to 2^{\{0,1\}^n}$: [9]

$$\mathcal{BR}_0(\mathbf{a}) \equiv \arg\max_{\mathbf{b} \in \{0,1\}^n} U(\mathbf{a}, \mathbf{b}) \ , \tag{17}$$

where U is as defined in Equation (16).

---

[8] We should note that the terms "$-a_i C_i$" are actually strategically irrelevant, and could have been removed. That doing so is sound will become clear when we define the best-response correspondence of the attacker (Equation (17)). We decided to keep those terms to explicitly express the notion that the attacker does not care about the cost for investments that any player may incur.

[9] Note that $\mathcal{BR}_0(\mathbf{a}) = \arg\max_{\mathbf{b} \in \{0,1\}^n} \sum_{i=1}^{n} M_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) - b_i C_i^0$, because the term $-\sum_{i=1}^{n} a_i C_i$ is not a function of $\mathbf{b}$.

Definition 6. *A pure-strategy profile* $(\mathbf{a}^*, \mathbf{b}^*) \in \{0,1\}^{2n}$ *is a PSNE of an IDD game if, for each player* i, $a_i^* \in \mathcal{BR}_i(\mathbf{a}^*_{\mathrm{Pa}(i)}, \mathbf{b}^*_{\mathrm{PF}(i)})$, *and for the attacker,* $\mathbf{b}^* \in \mathcal{BR}_0(\mathbf{a}^*)$, *where* $\mathcal{BR}_i$ *and* $\mathcal{BR}_0$ *are as defined in Equations (15) and (17), respectively.*

3.1. Conditions on Model Parameters

We now introduce the following reasonable restrictions on the game *parameters*. We employ these conditions without loss of generality from a mathematical and computational standpoint, as we now discuss.

The first condition states that every site's investment cost is positive and (strictly) smaller than the *conditional* expected *direct* loss if the site were to be attacked directly ($b_i = 1$). That is, if a site knows that an attack is directed against it, the site will prefer to invest in security, unless the *external risk* is too high. This condition is reasonable because otherwise the player will never invest regardless of what other players do (i.e., "not investing" would be a *dominant strategy*).

Assumption 1. *For all sites* $i \in [n]$, $0 < C_i < \widehat{p}_i L_i$.

The second condition states that, for all sites i, the attacker's cost to attack i is positive and (strictly) smaller than the expected loss achieved (i.e., gains from the perspective of the attacker) if an attack initiated at site i is successful, either directly at i or at one of its children (after transfer). That is, if an attacker knows that an attack is rewarding (or able to obtain a positive utility), it will prefer to attack some nodes in the network. This assumption is also reasonable; otherwise the attacker will never attack regardless of what other players do (i.e., not attacking would be a dominant strategy, leading to an easy problem to solve).

Assumption 2. *For all sites* $i \in [n]$, $0 < C_i^0 < \widehat{p}_i L_i + \sum_{j \in \mathrm{Ch}(i)} \widehat{q}_{ij} \alpha_j L_j$.

In what follows, we study the problem of finding and computing MSNE in IDD games under Assumptions 1 and 2.

3.2. There Is No PSNE in Any IDD Game with at Most One (Simultaneous) Attack

Note that the attacker has in principle an *exponential* number of pure strategies. For IDD games this translates to a number of *simultaneous* attacks being 2^n. This presents several challenges.

One is the question of how one would compactly represent an attacker's strategy (or policy) over 2^n events, when the representation size of the model is $N = O(|E| + n)$, which is $O(n^2)$ in the worst case. Thus, one would need a representation of the attacker's policy that is *polynomial* in N.

Another related question is how realistic is for the attacker to have such a huge amount of power.

One way to deal with the compact representation, while at the same time realistically constraining the attacker's power, is to limit the number of simultaneous attack sites to some small finite number $K \ll n$. Even then, the number of pure strategies will grow *exponentially* in the number of potential attacks K, which still renders the attacker's pure-strategy space unrealistic, especially on a very large network with about 30 K nodes and 100 K edges, like the one we study in our experiments (Section 5). Worst-case, we need to consider up to 2^n number of pure strategies for K attacks as K goes to n. The simplest version of this constraint is to allow at most a single (simultaneous) attack (i.e., $K = 1$).

Assumption 3. *The set of pure strategies of the attacker is*

$$\mathcal{B} = \{\mathbf{b} \in \{0,1\}^n \mid \textstyle\sum_{i=1}^n b_i \leq 1\}.$$

We emphasize that Definition 7 does not explicitly preclude multiple attacks, just that they cannot occur simultaneously.

For convenience, we introduce the following definition.

Definition 7. *We say an IDD game is a game with a* single simultaneous attack*, or simply a* single-attack *game for brevity, if Assumption 3 holds (i.e., at most one simultaneous attack is possible).*

The following technical results on single-attack IDD games will also be convenient.

Lemma 1. *The following holds in any single-attack IDD game* $([n], G, \mathbf{C}, \mathbf{L}, \widehat{\mathbf{p}}, \widehat{\mathbf{Q}}, \boldsymbol{\alpha})$*: for all players* $i \in [n]$*, for all* $\mathbf{a}_{\mathrm{Pa}(i)}$ *and* $\mathbf{b}_{\mathrm{PF}(i)}$*,*

$$r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) = \sum_{j \in \mathrm{Pa}(i)} b_j(1 - a_j)\widehat{q}_{ji}, s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) = 1 - r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) \text{ and} \tag{18}$$

$$b_i s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) = b_i \, . \tag{19}$$

The proof is in Appendix B.1.

Proposition 1. *In any single-attack IDD game* $([n], G, \mathbf{C}, \mathbf{L}, \widehat{\mathbf{p}}, \widehat{\mathbf{Q}}, \boldsymbol{\alpha})$*, we have that for all players* $i \in [n]$*, for all* $\mathbf{a}$ *and* $\mathbf{b}$*,*

$$U(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} b_i U_i(\mathbf{a}) \, , \tag{20}$$

where for all $i = 1, \ldots, n$*,*

$$U_i(\mathbf{a}) \equiv (1 - a_i) \left(\widehat{p}_i L_i + \sum_{j \in \mathrm{Ch}(i)} \widehat{q}_{ij}(a_j \alpha_j + (1 - a_j)) L_j \right) - C_i^0 \, . \tag{21}$$

The proof is in Appendix B.2.

We note that single-attack IDD games are instances of *graphical multi-hypermatrix games* [40], in which the local hypergraph of each site $i \in [n]$ has vertices $\mathrm{PF}(i) \cup \{0\}$, where 0 denotes the attacker node, and hyperedges $\{\{i\}, \{0, i\}\} \cup \bigcup_{j \in \mathrm{Pa}(i)} \{\{0, i, j\}\}$, while the local hypergraph of the attacker has $n + 1$ vertices $[n] \cup \{0\}$ and hyperedges $\{\{0\}\} \cup \bigcup_{i \in [n]} \left(\{\{0, i\}\} \cup \bigcup_{j \in \mathrm{Ch}(i)} \{\{0, i, j\}\} \right)$. This is implicit in the expression of the attacker's payoff function (Equations (20) and (21) in Proposition 1) and the resulting expression for the sites' cost functions (Equation (12)) after substituting the corresponding expressions in Equation (18). Thus, as a standard graphical game, the game graph of a single-attack IDD game has the attacker connected to each of the n sites. Yet, looking at it from the perspective of the subgame over the sites only, given a fixed $\mathbf{y}$, the resulting subgame among the sites only is a *graphical polymatrix game* in parametric-form, which for 2-action games is strategically equivalent to an *influence game* [41]. This view of the subgame over the sites only will be useful for our hardness results (Section 4.2) and the discussion within the concluding section (Section 6).

It turns out that when one combines Assumptions 1 and 2 on the parameters with Assumption 3, no PSNE is possible, as we formally state in the next proposition (Proposition 2). This is typical of attacker-defender settings. This technical result eliminates PSNE as a universal solution concept for natural IDD games in which at most one simultaneous attack is possible. The main significance of this result is that it allows us to concentrate our efforts on the much harder problem of computing MSNE.

Proposition 2. *No single-attack IDD game in which Assumptions 1 and 2 hold has a PSNE.*

The proof is in Appendix B.3.

3.3. Mixed Strategies in IDD Games

We do not impose Assumption 3 in this entire subsection. We do define some notation that will become useful when dealing with single-attack IDD games later in the manuscript (Section 3.4).

For each player i, denote by x_i the *mixed strategy of player i*: the probability that player i invests. Let $\mathbf{x} \equiv (x_i)_{i\in[n]}$ be the *joint mixed strategy*. Consider any subset $I \subset [n]$ of the internal players. Denote by $\mathcal{P}_{\{0,1\}^{|I|}}$ the set of all *joint-subset marginal probability mass functions (PMFs)* over $\{0,1\}^{|I|}$. For instance, $\mathcal{P}_{\{0,1\}^n}$ is the set of *joint PMFs* over the joint pure-strategy space of the attacker $\{0,1\}^n$, which is by definition the *set of all possible mixed strategies of the attacker.* Denote by $P_{\mathbf{B}} \in \mathcal{P}_{\{0,1\}^n}$ the *joint PMF* over $\{0,1\}^n$ corresponding to the *attacker's mixed strategy* so that for all $\mathbf{b} \in \{0,1\}^n$,

$$P_{\mathbf{B}}(\mathbf{b}) \equiv \mathbf{P}(\mathbf{B} = \mathbf{b}) \tag{22}$$

is the *probability that the attacker executes joint-attack vector* $\mathbf{b}$. Denote the *(subset) marginal PMF* over a subset $I \subset [n]$ of the internal players by $P_{\mathbf{B}_I} \in \mathcal{P}_{\{0,1\}^{|I|}}$, such that for all $\mathbf{b}_I \in \{0,1\}^{|I|}$, $P_{\mathbf{B}_I}(\mathbf{b}_I) \equiv \sum_{\mathbf{b}_{-I}} P_{\mathbf{B}}(\mathbf{b}_I, \mathbf{b}_{-I})$ is the *(joint marginal) probability that the attacker chooses a joint-attack vector in which the sub-component decisions corresponding to players in I are as in* $\mathbf{b}_I$. For simplicity of presentation, it will be convenient to let $\mathcal{P} \equiv \mathcal{P}_{\{0,1\}^n}$, $P \equiv P_{\mathbf{B}}$, $P_I \equiv P_{\mathbf{B}_I}$, and

$$y_i \equiv P_i(1) \equiv P_{B_i}(1) = P_{\mathbf{B}_{\{i\}}}(1)\ , \tag{23}$$

the *marginal probability that the attacker chooses an attack vector in which player i is directly targeted.*

Slightly abusing notation, we redefine the function e_{ij} (i.e., how safe i is from j), s_i and r_i (i.e., the overall transfer safety and risk, respectively), originally defined in Equations (10) and (11), as

$$\begin{aligned} e_{ij}(x_j, b_j) \equiv& x_j + (1 - x_j)(1 - b_j\widehat{q}_{ji})\ , \\ s_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) \equiv& \prod_{j\in\mathrm{Pa}(i)} e_{ij}(x_j, b_j)\ , \\ s_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{Pa}(i)}) \equiv& \sum_{\mathbf{b}_{\mathrm{Pa}(i)}\in\{0,1\}^{|\mathrm{Pa}(i)|}} P_{\mathrm{Pa}(i)}(\mathbf{b}_{\mathrm{Pa}(i)})\ s_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})\ , \text{ and} \end{aligned} \tag{24}$$

$$r_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{Pa}(i)}) \equiv 1 - s_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{Pa}(i)})\ , \tag{25}$$

where $\widehat{q}_{ji}$ is as defined in Equation (9).

In general, we can express the *expected* cost of protection to site i, with respect to a mixed-strategy profile $(\mathbf{x}, P)$, as

$$M_i(x_i, \mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) \equiv x_i[C_i + \alpha_i r_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{Pa}(i)})L_i] + (1 - x_i)[\widehat{p}_i f_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) + r_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)})]L_i\ , \tag{26}$$

where $\widehat{p}_i$ and r_i are as in Equations (8) and (25), respectively,

$$f_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) \equiv \mathbf{E}_{\mathbf{b}_{\mathrm{PF}(i)}\sim P_{\mathrm{PF}(i)}}[b_i s_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})] = \sum_{\mathbf{b}_{\mathrm{PF}(i)}\in\{0,1\}^{|\mathrm{PF}(i)|}} P_{\mathrm{PF}(i)}(\mathbf{b}_{\mathrm{PF}(i)})\ b_i\ s_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})\ , \tag{27}$$

and $s_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})$ is as defined in Equation (24).

The *expected payoff of the attacker* is

$$U(\mathbf{x}, P) \equiv \sum_{i=1}^{n} M_i(x_i, \mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) - x_i C_i - y_i C_i^0\ , \tag{28}$$

where M_i is as defined in Equation (26). Let

$$\widehat{s}_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) \equiv f_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) + \frac{1 - \alpha_i}{\widehat{p}_i} r_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{Pa}(i)})\ , \tag{29}$$

where $\widehat{p}_i$, r_i, and f_i are as defined in Equations (8), (25), and (27), respectively. The *mixed-strategy best-response correspondence of defender i* is then

$$\mathcal{BR}_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) \equiv \begin{cases} \{1\}, & \text{if } \widehat{s}_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) > \widehat{\Delta}_i, \\ \{0\}, & \text{if } \widehat{s}_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) < \widehat{\Delta}_i, \\ [0,1], & \text{if } \widehat{s}_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) = \widehat{\Delta}_i \end{cases}, \tag{30}$$

where $\widehat{\Delta}_i$ and $\widehat{s}_i$ are as defined in Equations (13) and (29), respectively.

The *best-response correspondence for the attacker* is simply

$$\mathcal{BR}_0(\mathbf{x}) \equiv \arg\max_{P \in \mathcal{P}_{\{0,1\}^n}} U(\mathbf{x}, P)\ , \tag{31}$$

where U is as defined in Equation (28).

Definition 8. *A mixed-strategy profile $(\mathbf{x}^*, P^*)$ is an MSNE of an IDD game if (1) for all $i \in [n]$, $x_i^* \in \mathcal{BR}_i(\mathbf{x}^*_{\mathrm{Pa}(i)}, P^*_{\mathrm{PF}(i)})$ and (2) $P^* \in \mathcal{BR}_0(\mathbf{x}^*)$, where $\mathcal{BR}_i$ and $\mathcal{BR}_0$ are as defined in Equations (30) and (31), respectively.*

3.3.1. A Characterization of the MSNE: Compact Representation of Attacker's Mixed Strategies

Recall that the space of pure strategies of the aggressor is, in its most general form, exponential in the number of internal players. This is an obstacle to tractable computational representations in large-population games. In the following result, we establish an equivalence class over the aggressor's mixed strategy that allows us to only consider "simpler" mixed strategies in terms of their probabilistic structure.

Proposition 3. *For any mixed strategy $(\mathbf{x}^*, P^*)$ of an IDD game, there exists another mixed strategy $(\mathbf{x}^*, \widetilde{P})$, such that*

1. *the joint PMF $\widetilde{P}$ decomposes as* [10]
$$\widetilde{P}(\mathbf{b}) \propto \prod_{i=1}^{n} \Phi_{\mathrm{PF}(i)}(\mathbf{b}_{\mathrm{PF}(i)})$$
for some non-negative functions $\Phi_{\mathrm{PF}(i)} : \{0,1\}^{k_i} \to [0,\infty)$, and all $\mathbf{b} \in \{0,1\}^n$,
2. *for all $i \in [n]$, the parent-family marginal PMFs $\widetilde{P}_{\mathrm{PF}(i)} = P^*_{\mathrm{PF}(i)}$ agree, and*
3. *the sites and the aggressor achieve the same expected cost and utility, respectively, in $(\mathbf{x}^*, \widetilde{P})$ as in $(\mathbf{x}^*, P^*)$: for all $i \in [n]$,*
$$M_i(\mathbf{x}^*_{\mathrm{PF}(i)}, \widetilde{P}_{\mathrm{PF}(i)}) = M_i(\mathbf{x}^*_{\mathrm{PF}(i)}, P^*_{\mathrm{PF}(i)}),$$
and
$$U(\mathbf{x}^*, \widetilde{P}) = U(\mathbf{x}^*, P^*).$$

Proof. *(Sketch)* The proof of the proposition follows closely a similar argument used by Kakade et al. [43] to characterize the probabilistic structure of *correlated equilibria* [44,45] in arbitrary graphical games. The core of the argument is to realize that the *maximum-entropy (MaxEnt) distribution* [46], over the aggressor's pure strategies, with the same parent-family marginals as those of P^* satisfy all the conditions above. We refer the reader to Appendix B.4 for the formal proof. □

[10] In other words, $\widetilde{P}$ is a Gibbs distribution with respect to the undirected "moralized" graph that results from adding an (undirected) edge among every pair of parents of every node to the original directed graph of the game and ignoring the directions of the edges in the original game graph. We refer the reader to Koller and Friedman [42] for a textbook introduction to concepts from probabilistic graphical models.

Thus, in the sense given by Proposition 3, even though in principle there may be equilibria in which the aggressor's mixed strategy is an arbitrarily complex distribution, we can restrict our attention to aggressor's mixed strategies that respect the decomposition in terms of functions over the parent-families of each site *only*. The proposition has important implications for the representation of the mixed strategies of the aggressor, and hence, the representation size of any MSNE, modulo expected-payoff equivalence. In particular, within the equivalence class of mixed strategies achieving the same expected-payoff, the representation needed to represent a mixed strategy in an IDD game is exponential only in the size of the *largest parent-family* k_i, not in the number of sites n. If the size of the largest parent-family $k_{\max} \equiv \max_{i \in [n]} k_i$ is *bounded*, then the representation is polynomial in the size of the representation size of the game N. Otherwise, because $N = O(n + |E|)$ is linear in the size of the game graph $G = ([n], E)$, while being exponential in the $k_{\max}$ is an exponential reduction in general from being exponential in n, it is still technically intractable with respect to N. The following corollary summarizes the discussion.

Corollary 1. *For any IDD game, let $k_{\max} \equiv \max_{i \in [n]} k_i$ be the size of the largest parent-family in the game graph. The representation size of any mixed strategy of the aggressor in the game is $O\left(2^{k_{\max}}\right)$, modulo expected-payoff equivalence.*

Theoretically establishing the existence of such compact representations for the attacker's mixed strategy in general may also have computational and algorithmic implications. This is because those results suggest that computing MSNE in arbitrary IDD games with arbitrary attacker's mixed strategies (i.e., arbitrary multiple simultaneous attacks) may be at least feasible in terms of the representation of the output MSNE itself. This is despite the fact that the attacker's mixed strategy in the MSNE is over an exponentially-sized set (i.e., $\{0,1\}^n$), as we discussed briefly at the beginning of this subsection, and Sections 1.2 and 3.2.

3.4. MSNE of IDD Games with at Most One Simultaneous Attack and Full Transfer Vulnerability

In this subsection, we impose Assumption 3. We first consider IDD games in which the players' investments cannot reduce the overall risk (i.e., $\alpha_i = 1$). This is the same setting used in the original IDS games (see Definitions 1 and 2, and the discussion on model semantics, in Section 2).

Assumption 4. *For all internal players $i \in [n]$, the probability that player i's investment in security does not protect the player from transfers, α_i, is* 1.

For convenience, we introduce the following definition.

Definition 9. *We say an IDD game is* fully transfer-vulnerable *if Assumption 4 holds.*

Before continuing, we remind the reader that the Definition 7 does not explicitly preclude multiple attacks, just that they cannot occur simultaneously. In terms of the attacker's mixed strategy P, as defined in Equation (22), Definition 7, via Assumption 3, implies that $\sum_{i=0}^{n} y_i = 1$, where each y_i is as defined in Equation (23) and y_0 is *the probability of no attack*:

$$y_0 \equiv P(\mathbf{0}) \equiv \mathbf{P}(\mathbf{B} = \mathbf{0}) = 1 - \sum_{i=1}^{n} y_i \,. \tag{32}$$

Thus, in what follows, when dealing with single-attack IDD games, we denote the joint mixed strategy $(\mathbf{x}, P)$ simply as $(\mathbf{x}, \mathbf{y})$ when P is compactly represented by $\mathbf{y}$ as defined in Equations (23) and (32); hence, for such games, we denote any MSNE $(\mathbf{x}^*, \mathbf{y}^*) \equiv (\mathbf{x}^*, P^*)$.

In addition, the sites' cost functions and the attacker's utility of fully transfer-vulnerable single-attack IDD games are simpler than their most general versions for arbitrary IDD games. This is

because Assumptions 3 and 4 greatly simplify the condition of the best-response correspondence of the internal players.

The following technical results extend Lemma 1 and Proposition 1 for single-attack IDD games from pure strategies to mixed strategies.

Lemma 2. *The following holds in any single-attack IDD game* $([n], G, \mathbf{C}, \mathbf{L}, \widehat{\mathbf{p}}, \widehat{\mathbf{Q}}, \boldsymbol{\alpha})$*: given joint mixed-strategy* $(\mathbf{x}, \mathbf{y}) \equiv (\mathbf{x}, P)$*, with* P *represented using* $\mathbf{y}$ *as defined in Equations (23) and (32), for all players* $i \in [n]$*, for all* $\mathbf{b}_{\mathrm{PF}(i)}$*,*

$$r_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) \equiv r_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) = \sum_{j \in \mathrm{Pa}(i)} b_j(1 - x_j)\widehat{q}_{ji} \,, \tag{33}$$

which implies

$$r_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)}) \equiv r_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{Pa}(i)}) = \sum_{j \in \mathrm{Pa}(i)} y_j(1 - x_j)\widehat{q}_{ji} \,, \tag{34}$$

$$s_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)}) \equiv s_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{Pa}(i)}) = 1 - r_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)}) \tag{35}$$

$$f_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) \equiv f_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) = y_i \text{ , and} \tag{36}$$

$$\widehat{s}_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) \equiv y_i + \frac{1 - \alpha_i}{\widehat{p}_i} \sum_{j \in \mathrm{Pa}(i)} y_j(1 - x_j)\widehat{q}_{ji} \,. \tag{37}$$

The proof is in Appendix B.5.

Proposition 4. *In any single-attack IDD game, we have, for all players* $i \in [n]$*,*

$$M_i(x_i, \mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) \equiv M_i(x_i, \mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}), \tag{38}$$

where $M_i(x_i, \mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)})$ *results from the corresponding substitution of the expressions given in Equations (34) and (36) into Equation (26); and*

$$U(\mathbf{x}, \mathbf{y}) \equiv U(\mathbf{x}, P) = \sum_{i=1}^{n} y_i U_i(\mathbf{x}) \,, \tag{39}$$

where for all $i = 1, \ldots, n$*,*

$$U_i(\mathbf{x}) \equiv (1 - x_i) \left(\widehat{p}_i L_i + \sum_{j \in \mathrm{Ch}(i)} \widehat{q}_{ij}(x_j \alpha_j + (1 - x_j)) L_j \right) - C_i^0 \,. \tag{40}$$

The proof is in Appendix B.6.

The expressions for sites' costs and attacker's payoff of fully transfer-vulnerable single-attack IDD games are even simpler, as we discuss in detail in the remaining of this section.

To start, from Equation (37), for this class of games we have $\widehat{s}_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) = y_i$.

Let $L_i^0(x_i) \equiv (1 - x_i)(\widehat{p}_i L_i + \sum_{j \in \mathrm{Ch}(i)} \widehat{q}_{ij} L_j)$. It will also be convenient to denote by

$$\overline{L}_i^0 \equiv L_i^0(0) = \widehat{p}_i L_i + \sum_{j \in \mathrm{Ch}(i)} \widehat{q}_{ij} L_j \,, \tag{41}$$

so that we can express $L_i^0(x_i) = (1 - x_i)\overline{L}_i^0$, to highlight that L_i^0 is a linear function of x_i.

Similarly, it will also be convenient to let $M_i^0(x_i) \equiv L_i^0(x_i) - C_i^0$, and denote by

$$\overline{M}_i^0 \equiv M_i^0(0) = \overline{L}_i^0 - C_i^0 \,. \tag{42}$$

Let

$$\eta_i^0 \equiv C_i^0 / \overline{L}_i^0 \ . \tag{43}$$

The best-response condition of the attacker also simplifies under the same assumptions because now

$$U(\mathbf{x}, \mathbf{y}) \equiv U(\mathbf{x}, P) = \sum_{i=1}^{n} y_i M_i^0(x_i) \ . \tag{44}$$

Assumption 2 is reasonable in our new context because, under Assumption 4, if there were a player i with $\eta_i^0 > 1$, the attacker would never attack i, and as a result player i would never invest. In that case, we can safely remove j from the game, without any loss of generality.

As graphical multi-hypermatrix games [40], fully transfer-vulnerable single-attack IDD games have a considerably simpler graph structure than for arbitrary single-attack IDD games. In particular, from Equation (38) in Proposition 4 (via Equations (30) and (37) with $\alpha_i = 1$ for all i), the local hypergraph of each site has a single hyperedge, i.e., $\{0, i\}$, while, from Equation (44), the local hypergraph of the attacker has n hyperedges, one for each site, of size 2 and the set of local hyperedges for the attacker (i.e., player 0) equals $\bigcup_{i\in[n]}\{\{0, i\}\} = \{\{0, 1\}, \{0, 2\}, \ldots, \{0, n\}\}$. Recall that in single-attack IDD games, given a fixed $\mathbf{y}$, only the subgame over just the sites is a graphical polymatrix game, not the whole game. Hence, adding full transfer vulnerability makes the whole game a graphical polymatrix game with a simple star graph in which the attacker is the center node (i.e., the single internal node), and each site node is a leaf. We exploit this property in the next subsection to provide a characterization of all MSNE in any such game.

3.4.1. Characterizing the MSNE of Fully Transfer-Vulnerable Single-Attack IDD Games

We now characterize the space of MSNE in fully transfer-vulnerable single-attack IDD games under Assumptions 1 and 2. Our characterization will immediately lead to a polynomial-time algorithm for computing *all* MSNE in that subclass of games (Section 4.1).

The characterization starts by partitioning the space of games into three, based on whether $\sum_{i=1}^{n} \widehat{\Delta}_i$ is (1) $<$, (2) =, or (3) $>$ than 1, where $\widehat{\Delta}_i$ is as defined in Equation (13). The rationale behind this is that now the players are indifferent between investing or not investing when $y_i = \widehat{\Delta}_i$, where y_i is as defined in Equation 23, by the resulting best-response correspondence for the attacker's mixed strategy in this case. The following result completely characterizes the set of MSNE in fully transfer-vulnerable single-attack IDD games.

Proposition 5. *Consider any fully transfer-vulnerable single-attack IDD game* $\mathcal{G} \equiv (n, G, \mathbf{C}, \mathbf{L}, \mathbf{C}^0, \widehat{\mathbf{p}}, \widehat{\mathbf{Q}}, \mathbf{1})$*, whose parameters satisfy Assumptions 1 and 2. Let* $\widehat{\Delta}_i$*,* $\overline{L}_i^0$*,* $\overline{M}_i^0$*, and* η_i^0 *be as defined in Equations (13), (41)–(43), respectively. The mixed-strategy profile* $(\mathbf{x}^*, \mathbf{y}^*)$ *is an MSNE of* $\mathcal{G}$ *in which*

1. $\sum_{i=1}^{n} \widehat{\Delta}_i < 1$ *if and only if*
   - (a) $1 > y_0^* = 1 - \sum_{i=1}^{n} \widehat{\Delta}_i > 0$*, and*
   - (b) *for all* i*,* $y_i^* = \widehat{\Delta}_i > 0$ *and* $0 < x_i^* = 1 - \eta_i^0 < 1$.
2. $\sum_{i=1}^{n} \widehat{\Delta}_i = 1$ *if and only if*
   - (a) $y_0^* = 0$*, and*
   - (b) *for all* i*,* $y_i^* = \widehat{\Delta}_i > 0$ *and* $x_i^* = 1 - \frac{v + C_i^0}{\overline{L}_i^0}$ *with* $0 \leq v \leq \min_{i\in[n]} \overline{M}_i^0$.
3. $\sum_{i=1}^{n} \widehat{\Delta}_i > 1$ *if and only if*
   - (a) $y_0^* = 0$*, and*

(b) there exists a non-singleton, non-empty subset $I \subset [n]$*, such that* $\min_{i\in I} \overline{M}_i^0 \geq \max_{k\notin I} \overline{M}_k^0$ *if* $I \neq [n]$*, and the following holds:*

i. *for all* $k \notin I$, $x_k^* = 0$ *and* $y_k^* = 0$,

ii. *for all* $i \in J \equiv \arg\min_{i\in I} \overline{M}_i^0$, $x_i^* = 0$ *and* $0 \leq y_i^* \leq \widehat{\Delta}_i$*, and in addition,* $\sum_{i\in J} y_i^* = 1 - \sum_{t\in I-J} \widehat{\Delta}_i$*; and*

iii. *for all* $i \in I - J$, $y_i^* = \widehat{\Delta}_i$ *and* $0 < x_i^* = 1 - \frac{\min_{t\in I} \overline{M}_t^0 + C_i^0}{\overline{L}_i^0} < 1$.

The proof of Proposition 5 is in Appendix B.7. As a proof sketch, we briefly state that the proposition follows from the restrictions imposed by the model parameters and their implication to indifference and monotonicity conditions. We also mention that the third case in the proposition implies that if the $\overline{M}_l^0$'s form a complete order, then the last condition stated in that case allows us to search for an MSNE by exploring only $n-2$ sets, vs. 2^{n-2} if done naively.

It turns out that a complete order is unnecessary. The following claim allows us to safely move all the internal players with the same value of $\overline{M}_i^0$ in a group as a whole inside or outside I. This technical result is important because of its algorithmic implications, as we discuss in Section 4.1.

Claim 1. *Let* $I \subset [n]$*, such that* $I' \subset I$, $|I'| < |I| < n-1$*. Suppose we find an MSNE* (x, y) *such that* $I' = \{i \mid y_i > 0\}$*, with the property that* $\min_{l\in I'} \overline{M}_l^0 = \max_{k\notin I'} \overline{M}_k^0$*. In addition, suppose* I *satisfies* $\min_{l\in I'} \overline{M}_l^0 = \min_{l\in I} \overline{M}_l^0 \geq \max_{k\notin I} \overline{M}_k^0$*. Then, we can also find* (x, y) *using the partition imposed by* I*.*

The proof of the claim is in Appendix B.8.

3.4.2. Some Remarks on the MSNE of Fully Transfer-Vulnerable Single-Attack IDD Games

We begin by pointing out that, under the conditions of Proposition 5, for *almost every* setting of the free parameters of the system, subject to their respective constraints, every IDD games have a corresponding unique MSNE, which we denote by $(\mathbf{x}^*, \mathbf{y}^*)$. We consider that MSNE to be the equilibrium or stable outcome of the system.

Security Investment Characteristics

At equilibrium, if $x_i^* > 0$, the probability of *not* investing is proportional to C_i^0 and *inversely* proportional to $\widehat{p}_i L_i + \sum_{j\in \mathrm{Ch}(i)} \widehat{q}_{ij} L_j$. It is kind of reassuring that, at an equilibrium, the probability of investing increases with the potential loss a site's non-investment decision could cause to the system. Hence, behavior in a stable system implicitly "forces" all sites to indirectly account for or take care of their own children. This may sound a bit paradoxical at first given that we are working within a "noncooperative" setting and each sites's cost function is only dependent on the investment decision of the player's *parents*, in general. However, in the case of fully transfer-vulnerable single-attack IDD games, the site's cost is only a function of its mixed strategy x_i and the probability y_i that the attacker will directly target site i. Said differently, any site's best response is independent of their parents, the source of transfer risk, if investment in security does nothing to protect that player from transfers (i.e., $\alpha_i = 1$). Interestingly, even in such circumstances, the existence of the attacker in the system is inducing an (almost-surely) unique stable outcome in which an implicit form of "cooperation" occurs. In retrospect, this makes sense because no site can control the transfer risk. Said differently, there is nothing any site can do to prevent the transfer, even though the original potential for transfers does depend on the parents' investment strategies. In short, rational/optimal noncooperative behavior for each site is not only to protect the player's own losses but also to "cooperate" to protect the player's children.

Relation to Network Structure

How does the network structure of the given input game model relates to its equilibrium? As seen above, the values of the equilibrium strategy of each player depend on information from the attacker, the player and the player's children. From the discussion in the last paragraph, within the context of the given input model, a player's probability of investing at the equilibrium increases with the expected loss sustained from a "bad event" occurring as a result of a transfer from a player to the player's children.

Let us explore this last point further by considering the case of uniform-transfer probabilities (also studied by Kunreuther and Heal [3] and Kearns and Ortiz [47]). In that case, transfer probabilities are only a function of the source, not the destination: $\widehat{q}_{ij} \equiv \widehat{\delta}_i$. The expression for the equilibrium probabilities of those players who have a positive probability of investing would simplify to $x_i^* = 1 - \frac{v+C_i^0}{\widehat{p}_i L_i + \delta_i \sum_{j \in \mathrm{Ch}(i)} L_j}$, for some constant v. The last expression suggests that $\sum_{j \in \mathrm{Ch}(i)} L_j$ differentiates the probability of investing between sites. That would suggest that the larger the number of children in the given input game model, the larger the probability of investing at the corresponding equilibrium for the given input game model. A scenario that seems to further lead us to that conclusion is when we make the further assumption of an homogeneous system as first studied in the original IDS paper [3]: $L_i \equiv L$, $\widehat{p}_i \equiv \widehat{p}$, $\delta_i \equiv \delta$, and $C_i^0 \equiv C^0$ [11] for all players. Then, we would get $x_i^* = 1 - \frac{v+C^0}{L(\widehat{p}+\delta|\mathrm{Ch}(i)|)}$. Thus, at equilibrium, the probability of *not* investing, $1 - x_i^*$, is inversely proportional to the *number* of children player i has, which is implicit in the directed graph over sites in the model given as input.

On the Attacker's Equilibrium Strategy

The support of the attacker, $I^* \equiv \{i \mid y_i^* > 0\}$, at equilibrium has the following properties:

1. players for which the attacker's cost-to-expected-loss is higher are "selected" first in the algorithm;
2. if the size of that set is t, and there is a lower bound on $\widehat{\Delta}_i > \widehat{\Delta}$, and $\sum_{i=1}^n \widehat{\Delta}_i > 1$, then $t < 1/\widehat{\Delta}$ is an upper-bound on the number of players that could potentially be attacked;
3. if we have a game with homogeneous parameters, then the probability of an attack will be uniform over that set I^*; and
4. all but one of the players in that set I^* invest in security with some non-zero probability, for almost every parameter setting for IDD games satisfying the conditions of Proposition 5.

4. On the Complexity of Computing an MSNE in Single-Attack IDD Games

Here, we consider the computational complexity of computing MSNE in single-attack IDD games.

4.1. Computing All MSNE of Fully Transfer-Vulnerable Single-Attack IDD Games in Polynomial Time

We now describe an algorithm to compute *all* MSNE in a fully transfer-vulnerable single-attack IDD games that falls off Proposition 5. We begin by noting that the equilibrium in the case of IDD games with $\sum_{i=1}^n \widehat{\Delta}_i \leq 1$, corresponding to cases 1 and 2 of the proposition, has essentially an analytic closed-form. Hence, we concentrate on the remaining and most realistic case in large-population games, for which we expect $\sum_{i=1}^n \widehat{\Delta}_i > 1$. We start by sorting the indices of the internal players in descending order based on the $\overline{M}_i^0$'s. Let $\mathrm{Val}(l)$ and $\mathrm{Idx}(l)$ be the lth value and index in the resulting sorted list, respectively. Find t such that $1 - \widehat{\Delta}_{\mathrm{Idx}(t)} \leq \sum_{l=1}^{t-1} \widehat{\Delta}_{\mathrm{Idx}(l)} < 1$. Let $k = \max\{l \mid l \geq t \text{ and } \mathrm{Val}(l) = \mathrm{Val}(t)\}$ (i.e., continue down the sorted list of values until a change occurs). For $i = 1, \ldots, t-1$, let $l = \mathrm{Idx}(i)$ and set $x_l^* = 1 - \frac{\mathrm{Val}(t)+C_l^0}{\overline{L}_l^0}$ and $y_l^* = \widehat{\Delta}_l$. For $i = k+1, \ldots, n$, let $l = \mathrm{Idx}(i)$ and set $x_l^* = 0$ and $y_l^* = 0$.

[11] Note that this does not mean that the expected loss caused by a player that does not invest but is attacked, $L\,(\widehat{p} + \delta|\mathrm{Ch}(i)|)$, is the same for all players.

For $i = t, \ldots, k$, let $l = \text{Idx}(i)$ and set $x_l^* = 0$. Finally, represent the simplex defined by the following constraints: for $i = t, \ldots, k$, let $l = \text{Idx}(i)$ and $0 \leq y_l^* \leq \widehat{\Delta}_l$; $\sum_{i=t}^{k} y_{\text{Idx}(i)}^* = 1 - \sum_{i=1}^{t-1} \widehat{\Delta}_{\text{Idx}(i)}$. The running time of the algorithm is $O(n \log n)$ (because of sorting).

Theorem 1. *There exists a polynomial-time algorithm to compute all MSNE of any fully transfer-vulnerable single-attack IDD game with parameters that satisfy Assumptions 1 and 2.*

In cases in which the equilibria is not unique, it can be generated via simple sampling of either a simple linear system or a simplex. In either case, one can compute a single MSNE from that infinite set in polynomial time [48].

Let us revisit the types of games that may have an infinite MSNE set. Note that the case in which $\sum_{i=1}^{n} \widehat{\Delta}_i = 1$ has (Borel) measure zero and is quite brittle (i.e., adding or removing a player breaks the equality). For the case in which $\sum_{i=1}^{n} \widehat{\Delta}_i > 1$, if the value of the $\overline{M}_i^0$'s are distinct, [12] then there is a unique MSNE. Algorithm A1 in Appendix C provides pseudocode of the exact algorithm just described in this subsection.

In what remains of this section, we will continue to study the problem of computing MSNE in single-attack IDD games under Assumptions 1 and 2 as stated and discussed in Section 3.1.

4.2. Hardness Results on Computing MSNE in General Single-Attack IDD Games

For the purpose of studying the computational complexity of single-attack IDD games, it is natural to view the computation of an MSNE as a two-part process. Given an attacker's strategy, we need to determine the MSNE of the underlying game of the sites, or *sites-game* for short. The sites-game could have many MSNE and each MSNE could yield a different utility for the attacker (and the sites). Naively, the attacker can verify whether each of the MSNE is in the attacker's best response. Clearly, doing so depends on whether we can efficiently compute all MSNE in the sites-game, which of course depends on the given attacker's strategy. For example, if $\sum_{i=1}^{n} y_i = 0$, then the sites-game would have 'none invest' as the only outcome, because of Assumption 1 in Section 3.1.

Our goal in this subsection is to formally prove that there is an instance of a single-attack IDD game, and an attacker's strategy in that instance, with the property that, if we fix that attacker's strategy, we cannot compute all of the MSNE efficiently in the underlying sites-game, unless $P = NP$. The implication is that the existence of an efficient algorithm to compute an MSNE of IDD games based on the natural two-part process described in the previous paragraph, (i.e., checking whether each attacker's strategy can be part of an MSNE), is unlikely.

To formally prove that it is unlikely that we can always tractably compute all of the MSNE in an instance of the sites-games, as induced by an IDD game and an attacker's strategy, we consider the PURE-NASH-EXTENSION computational problem [47] for binary-action n-player games, which is NP-complete for graphical IDS games (see Definition 2). In the PURE-NASH-EXTENSION problem, the input is given by a description of the game and a *partial* assignment to a joint pure strategy $\mathbf{a} \in \{0, 1, *\}^n$, where '$*$' is our way of indicating that some components of the joint pure strategy $\mathbf{a}$ do not have assignments yet. In fact, the computational problem is precisely to determine whether there exists a *complete* assignment, i.e., a joint pure strategy, $\mathbf{a}' \in \{0, 1\}^n$ consistent with $\mathbf{a}$ in the following sense: for all i, if $a_i \in \{0, 1\}$, then we must have $a_i = a_i'$, but if $a_i = *$, then we are free to assign a value to $a_i' \in \{0, 1\}$, as long as the resulting joint pure strategy $\mathbf{a}'$ is a PSNE of the given input game model. [13]

---

[12] Distinct $\overline{M}_i^0$'s for the set of defenders at which the sum goes over one is sufficient to guarantee unique MSNE.

[13] We note that proving that computing an MSNE in IDD games is PPAD-complete would be more appropriate, since there always exists an MSNE in IDD games, but we will leave that question for future work.

Theorem 2. *Consider a variant of single-attack IDD games in which Assumptions 1 and 2 hold, and* $\sum_{i=1}^{n} \widehat{\Delta}_i \leq 1$. *There is an attacker's strategy P defined by* $\mathbf{y}$ *such that if we fix* $\mathbf{y}$, *then the* PURE-NASH EXTENSION *problem for the induced n-player sites-game is NP-complete.*

The proof of the theorem is in Appendix B.9. In the worst case, we need to consider the $\mathbf{y}$ just described, should other strategies fail to be a part of any MSNE. Another challenge is that even if we can compute all exact MSNE, there could be exponentially many of them to check. In the next section, we provide provably efficient algorithms to compute an approximate MSNE in tree-like subgraph structures over the sites only.

4.3. FPTAS to Compute Approximate MSNE of Tree-Like Single-Attack IDD Games

In this section, we focus on the question of computing *approximate* MSNE, a concept which the upcoming Definition 10 formalizes in our context, in a subclass of IDD games. In particular, we focus our study to the case of single-attack IDD games in which the game subgraph composed of the sites is a directed tree, although our technical result holds for slightly more general tree-like graph structures among the sites. Despite the apparent simplicity of the subgraph over the sites, one can envision very important real-world applications such as protection of supply chains and other hierarchical structures (e.g., see Agiwal and Mohtadi [23]). Recent work on optimization problems related to the power grid uses a directed-tree graphical model as the underlying structure of the electricity distribution network [24]. We note that the attacker is connected to all of the sites even if we do not point it out explicitly.

Given that there is no PSNE in any IDD games, under reasonable conditions (Proposition 2), we shift our focus to computing an MSNE. In Section 4.1, we provided an algorithm to compute all exact MSNE in an instance of IDD games where $\alpha_i = 1$ for all sites i (i.e., investment cannot protect the sites from indirect risk). The result we present in this subsection is for computing approximate MSNE, but holds for general $\boldsymbol{\alpha}$. We now formally define approximate MSNE in the context of this subsection.

Definition 10. *A mixed strategy* $(\mathbf{x}^*, \mathbf{y}^*)$ *is an* ϵ-MSNE *of a single-attack IDD game if*

1. *for all* $i \in [n]$, $M_i(x_i^*, \mathbf{x}^*_{\mathrm{Pa}(i)}, \mathbf{y}^*_{\mathrm{PF}(i)}) \leq \min_{x_i} M_i(x_i, \mathbf{x}^*_{\mathrm{Pa}(i)}, \mathbf{y}^*_{\mathrm{PF}(i)}) + \epsilon$, *and*
2. $U(\mathbf{x}^*, \mathbf{y}^*) \geq \max_{\mathbf{y}} U(\mathbf{x}^*, \mathbf{y}) - \epsilon = \max\left(\max_{i\in[n]} U_i(\mathbf{x}), 0\right) - \epsilon$,

where the M_i*'s,* U, *and* U_i*'s are as defined in Equations (38)–(40) in Proposition 4, respectively.*

An exact MSNE $\equiv$ 0-MSNE. Moreover, we assume that all the cost and utility functions are individually normalized to $[0, 1]$ and $\epsilon \in [0, 1]$; otherwise ϵ is not truly well-defined.

The following is one of our main technical results about computing approximate MSNE in IDD games with arbitrary $\boldsymbol{\alpha}$ with a directed tree network structure over the sites.

Theorem 3. *There exists an FPTAS to compute an* ϵ*-MSNE in single-attack IDD games with directed tree-like graphs G over the sites, under Assumptions 1 and 2 on the game parameters.*

The proof is in Appendix D.

Note that Theorem 3 is nontrivial within the context of the state-of-the-art in computational game theory. We are working with a graph structure where there is one node (the attacker) connected to *all* the nodes of the tree-like graph G (the sites). Naively applying the traditional well-known *dynamic programming (DP)* algorithms of [37] and [49] to our problem would not give us any FPTAS. In fact, their game representation size is exponential in the number of neighbors instead of our *linear* representation size. Moreover, finding ϵ-MSNE in general degree-3 graphical games is PPAD-hard [49], even if the payoff is additive [29], or more generally a graphical polymatrix game in parametric form [50], as is the case for our model under at most one simultaneous attack (Assumption 3).

5. Experiments

In the previous section (Section 4), we established the theoretical characteristics and computational tractability of single simultaneous attack IDD games with the highest transfer vulnerability parameter: $\alpha_i = 1$. We also established hardness results related to computing all MSNE in single-attack IDD games with arbitrary graph structures and $\boldsymbol{\alpha}$ values. We then considered approximate MSNE in the same class of games and provided an FPTAS for cases in which the game subgraph over the sites is a directed tree-like graph. In this section, partly motivated by security problems in cyberspace, we concentrate instead on empirically evaluating the other extreme of transfer vulnerability: games with low α_i values (i.e., near 0), so that investing in security considerably reduces the transfer risk. We also consider a complex graph structure found in the real-world Internet corresponding to the AS-level network, as measured in March 2010 by DIMES.

Our main objectives for the experiments presented here are

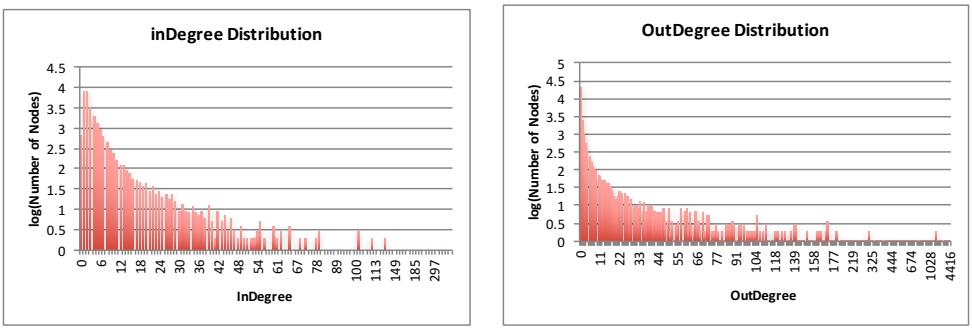
1. to demonstrate that a simple heuristic, *best-response-gradient dynamics (BRGD)*, is practically effective in computing an ϵ-MSNE, up to $\epsilon = \Omega(10^{-3})$, in a very large class of IDD games with realistic Internet-scale network graphs in a reasonable amount of time for cases in which the transfer vulnerabilities α_i's are low;
2. to explore the general structural and computational characteristics of (approximate) MSNE in such IDD games, including their dependence on the underlying network structure of the game (and approximation quality); and
3. to evaluate and illustrate the effectiveness of an improved version of the simple heuristics, which uses the concept of *smooth best-response dynamics (SBRD)* for the attacker, for computing ϵ-MSNE for ϵ values that are an order of magnitude lower (i.e., $\epsilon = \Omega(10^{-4})$).

BRGD is a well-known technique from the literature on learning in games [35]. We refer the reader to Singh et al. [51] for more information on properties of BRGD, and to Kearns and Ortiz [47], Heal and Kunreuther [4], and Kearns [52] for examples of its application within an IDS context. Here, we use BRGD as a tool to *compute* an ϵ-approximate MSNE (Definition 10), as was the case for the other previous applications of the technique in IDS contexts. Our particular implementation of BRGD begins by initializing x_i and y_i in $[0,1]$ for all sites i such that $\sum_{i=1}^{n} y_i \leq 1$. At each round, BRGD updates, simultaneously for each i,

$$\begin{aligned} x_i \leftarrow & x_i - 10 \times (M_i(1, \mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) - M_i(0, \mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) \text{ and} \\ y_i \leftarrow & y_i + 10 \times (U_i(\mathbf{x}) - U(\mathbf{x}, \mathbf{y})) \,, \end{aligned}$$

where the U_i's are as defined in Equation (40); the M_i's (Equation (38)) and U (Equation (39)) functions are normalized to $[0,1]$; and the constant value 10 is the *learning-rate/step-size* in our case.

We obtained the latest version (March 2010 at the time) of the real structure and topology of the AS-level Internet from DIMES (netdimes.org) [34]. The AS-level network has 27,106 nodes (683 isolated) and 100,402 directed edges; the graph length (diameter) is 6253, the density (number of edges divided by number of possible edges) is 1.9920×10^{-5}, and the average (in and out) degree is 3.70, with ≈76.93% and 2.59% of the nodes having zero indegree and outdegree, respectively. Figure 1 shows the indegree and outdegree distribution and Figure 2 shows the scatter plot of indegree and outdegree of the graph.



(a) **(b)**

Figure 1. Histograms of Indegree and Outdegree of the Nodes of the Internet Graph from DIMES at the Level of *Autonomous Systems* (*AS*). The bar graphs show (the logarithm, base 10, of) the number nodes with a particular outdegree (**a**) and indegree (**b**) value. (The graphs only show the in/out degrees with a non-zero number of nodes.)

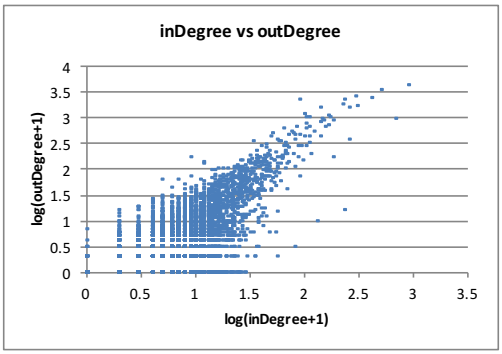


Figure 2. Indegree and Outdegree of the Nodes of the AS-level Intenet Graph from DIMES. The scatter plot shows the indegree and outdegree pairs of the AS nodes in logarithmic (base 10) scale.

All the IDD games in the experiments presented in this section have this network structure.

For simplicity, we call *Internet games* the class of IDD games with the AS-level network graph and low α_i values. We considered various settings for model parameters of Internet games: a single instance with specific fixed values; and several instances generated at random (see Table 1 for details). The attacker's cost-to-attack parameter for each node i is always held constant: $C_i^0 = 10^6$. For each run of each experiment, we ran BRGD with randomly-generated initial conditions (i.e., random initializations of the players' mixed strategies): $x_i \sim \text{Uniform}([0,1])$, i.i.d. for all i, and y is a probability distribution generated uniformly at random, and independent of $\mathbf{x}$, from the set of all probability mass functions over $n+1$ events. [14] The initialization of the transfer-probability parameters of a node essentially gives higher transfer probability to children with high (total) degree (because they are potentially "more popular"). The initialization also enforces $\widehat{p}_i + \sum_{j \in \text{Ch}(i)} \widehat{q}_{ij} = 0.9$. Other initializations are possible but we did not explore them here.

[14] Recall the probability of no attack $y_0 = 1 - \sum_{i=1}^{n} y_i$.

Table 1. Internet Games' Model Parameters.

Model Parameters	Fixed: $\mathrm{U} = 0.5$ Random: $\mathrm{U} \sim \mathbf{Uniform}([0,1])$
α_i	$\mathrm{U}/20$
L_i	$10^8 + (10^9) \times \mathrm{U}$
C_i	$10^5 + (10^6) \times \mathrm{U}$
$\widehat{p}_i$	$0.9 \times \frac{\widetilde{p}_i}{\widetilde{p}_i + \sum_{k \in \mathrm{Ch}(i)} \widetilde{q}_{ik}}$
$\widehat{q}_{ij}$	$0.9 \times \frac{\widetilde{q}_{ij}}{\widetilde{p}_i + \sum_{k \in \mathrm{Ch}(i)} \widetilde{q}_{ik}}$
z_i	$0.2 + \mathrm{U}/5$
$\widetilde{p}_i$	$0.8 + \mathrm{U}/10$
$\widetilde{q}_{ij}$	$z_i \frac{\lvert\mathrm{Ch}(j)\rvert + \lvert\mathrm{Pa}(j)\rvert}{\sum_{k \in \mathrm{Ch}(i)} \lvert\mathrm{Ch}(k)\rvert + \lvert\mathrm{Pa}(k)\rvert}$
C_i^0	10^6

5.1. Computing an ϵ-MSNE Using BRGD

Given the lack of theoretical guarantees on the convergence rate of BRGD, we began our empirical study by evaluating the convergence and computation/running-time behavior of BRGD on Internet games. We ran ten simulations for each ϵ value and recorded the number of iterations until convergence (up to 2000 iterations). Figure 3 presents the number of iterations taken by BRGD to compute an ϵ-MSNE as a function of ϵ. All simulations in this experiment converged (except for $\epsilon = 0.001$, for which two of the runs on the single instance and all those on randomly-generated instances did not converge). Each iteration took roughly 1–2 s. (wall clock). Hence, we can use BRGD to consistently compute an ϵ-MSNE of a 27 K-players Internet game in a few seconds.

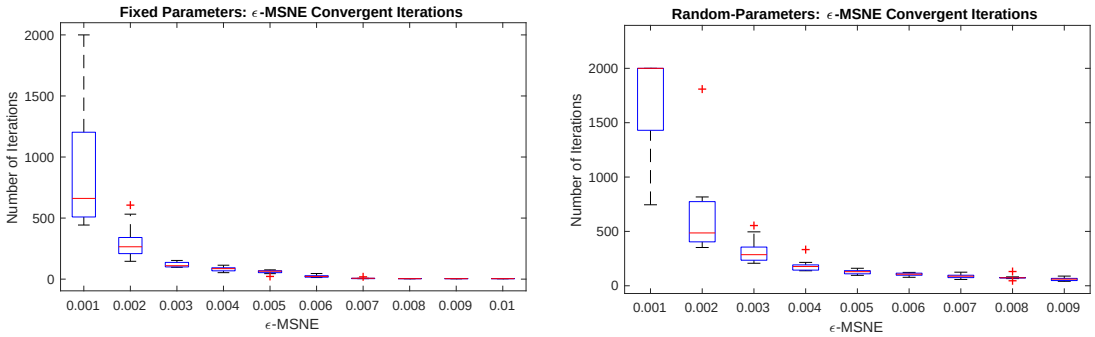


Figure 3. Convergence Rate of *Best-Response Gradient Dynamics* (*BRGD*) Heuristic for Computing an Approximate *Mixed Strategy Nash Equilibrium* (*MSNE*). The plots in this figure present the number of iterations of BRGD as a function of ϵ under the two experimental conditions: Internet games with fixed (**a**) and randomly-generated parameters (**b**). Applying *mean-squared-error* (*MSE*) regression to the left-hand and right-hand graphs, we obtain a functional expression for the number of iterations $N^F(\epsilon) = 0.00003\ \epsilon^{-2.547}$ ($R^2 = 0.90415$) and $N^R(\epsilon) = 0.0291\ \epsilon^{-1.589}$ ($R^2 = 0.9395$), respectively (i.e., low-degree polynomials of $1/\epsilon$).

We now concentrate on the empirical study of the *structural* characteristics of the ϵ-MSNE found by BRGD. We experimented on both the single and randomly-generated Internet game instances. We discuss the typical behavior of the attacker and the sites in an ϵ-MSNE, and the typical relationship between ϵ-MSNE and network structure.

5.1.1. A Single Internet Game

We first studied the characteristics of the ϵ-MSNE of a single Internet game instance. The only source of randomness in these experiments comes from BRGD's initial conditions (i.e., the initialization of the mixed strategies $\mathbf{x}$ and $\mathbf{y}$). BRGD consistently found *exact* MSNE (i.e., $\epsilon = 0$) in *all* runs.

Players' Equilibrium Behavior

In fact, we consistently found that the attacker always displays only two types of "extreme" equilibrium behavior, corresponding to the two kinds of MSNE BRGD found for the single Internet game: place positive probability of a direct attack to either *almost all* nodes (Strategy A) or a *small subset* (Strategy B). Figure 4 shows a plot of the typical probability of direct attack for those two equilibrium strategies for the attacker when BRGD stops. In both strategies, a relatively small number of nodes (about 1K out of 27K) have a reasonably *high* (and near *uniform*) probability of direct attack. In Strategy A, however, *every* node has a positive probability of being the target of a direct attack, albeit relatively very low for most; this is in contrast to Strategy B where *most* nodes are fully immune from a direct attack. Interestingly, *none* of the nodes invests in either MSNE: $x_i^* = 0$ for all nodes i. Thus, in this particular Internet game instance, *all* site nodes are willing to risk an attack.

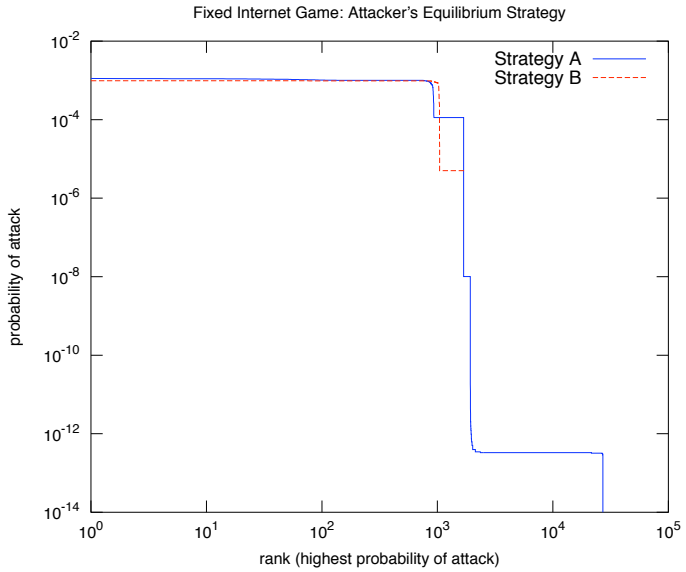


Figure 4. Attacker's Equilibrium Strategy on an Internet Game Instance (Fixed). The graph shows the values of $y_i^* > 0$ for each node i, sorted in decreasing order (in log-log scale), for attacker's Strategy A (blue/denser-dots line) and Strategy B (red/sparser-dots line) at an MSNE of the single instance of the Internet game.

Relation to Network Structure

We found that the nodes with (relatively) high probability of direct attack are at the "fringe" of the graph (i.e., have low or no degree). In Strategy A, fringe nodes (with mostly 0 or 1 outdegree) have relatively higher probability of direct attack than nodes with higher outdegree. Similarly, in Strategy B, the small subset of nodes that are potential target of a direct attack have relatively low outdegree (mostly 0, and 0.0067 on average; this is in contrast to the average outdegree of 3.9639 for the nodes immune from direct attack). Figure 5 shows the relation between the probability of attack and outdegree and the relation between the indegree and outdegree of a typical simulation runs for strategy A and for Strategy B as described above, respectively. We emphasize that these observations

are consistent throughout all runs of the experiment. In short, we consistently found that the nodes with low outdegree are more likely to get attacked directly in the single Internet game instance.

5.1.2. Randomly-Generated Internet Games

We now present results from experiments on randomly-generated instances of ten Internet games, a single instance for each $\epsilon \in \{0.001, 0.002, \ldots, 0.009\}$. For simplicity, we present the result of a single BRGD run on each instance. [15]

Behavior of the Players

Figure 6 shows plots of the attacker's probability of direct attack and histograms of the nodes's probability of investment in a typical run of BRDG on each randomly-generated Internet game instance for each ϵ value.

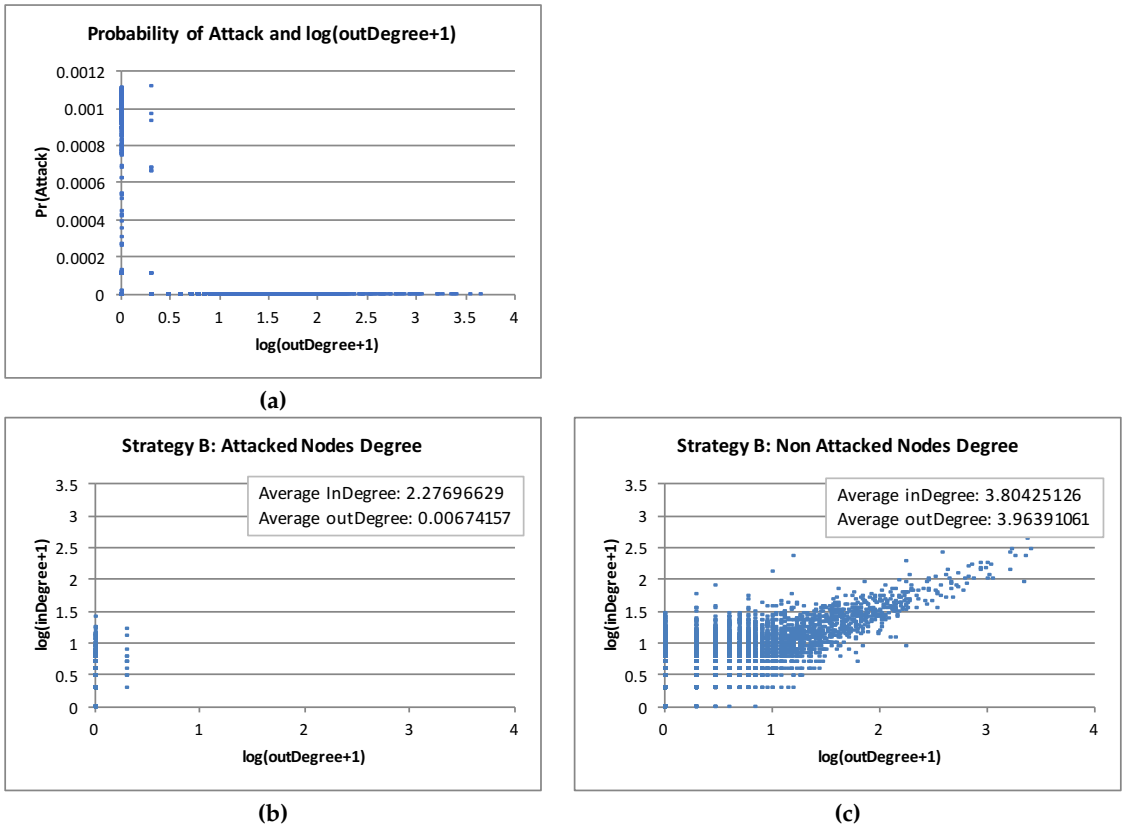


Figure 5. Attacker's Equilibrium Strategy and the Degrees of the Nodes. The top graph (**a**), which depicts Strategy A (all 27106 nodes), shows the probability of attack (y-axis) of a node and its corresponding outdegree (x-axis) in logarithmic (base 10) scale. The bottom graphs, (**b**) and (**c**), show the indegree (y-axis) of a node and its corresponding outdegree (x-axis) in logarithmic (base 10) scale of Strategy B: the graphs on the left and right consist of the (1780) nodes with nonzero probability of attack and the (25326) nodes with zero probability of attack, respectively.

[15] While some results presented here are for a single instance of the Internet game for each ϵ, the results are typical of multiple instances. Our observations are robust to the experimental randomness in both the Internet game parameters and the initialization of BRGD. For the sake of simplicity of presentation, we discuss results based on a single instance of the Internet game, and in some cases based on a single BRGD run. Note that, for each ϵ value we considered, the Internet game parameters remain constant within different BRGD runs. BRGD always converged within 2000 iterations (except 6 runs for $\epsilon = 0.001$).

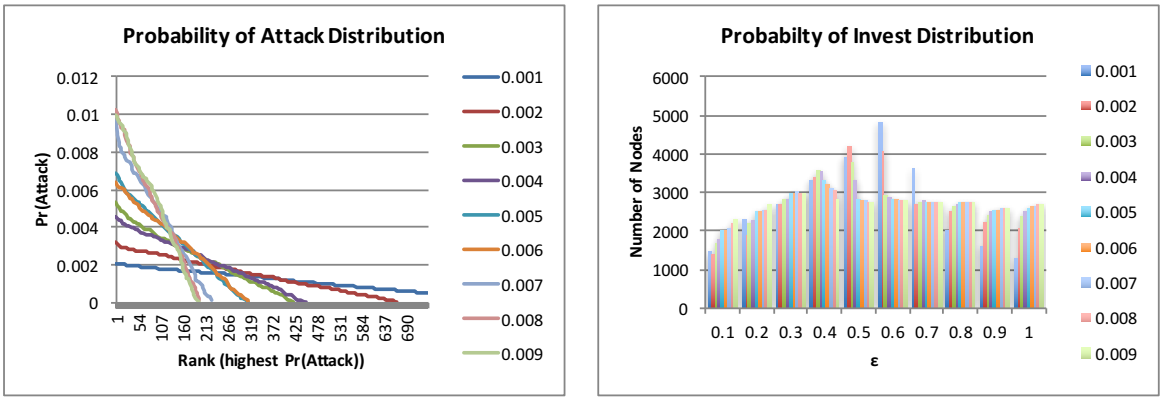


Figure 6. Attacker's and Site's ϵ-mixed-strategy Nash equilibria (MSNE) Strategies for a Randomly-Generated Internet Game. The graphs show the empirical distributions of the probability of attack (**a**) and histograms of the probability of investment (**b**), for different ϵ-value conditions encoded in the right-hand side of the plots (i.e., from 0.001 to 0.009). In both graphs, the distributions and histograms found for each ϵ value considered are drawn in the same corresponding graph superimposed. The left-hand graph plots the distribution of y_i where the nodes are ordered decreasingly based on the y_i value. The right-hand bar graph shows histograms of the probability of investing in defense/security measures based on the following sequence of 10 ranges partitioning the unit interval: $([0, 0.1], (0.1, 0.2], ..., (0.9, 1])$.

The plots suggest that approximate MSNE found by BRGD is quite sensitive to the ϵ value: as ϵ decreases, the attacker tends to "spread the risk" by selecting a larger set of nodes as potential targets for a direct attack, thus lowering the probability of a direct attack on any individual node; the nodes, on the other hand, tend to deviate from (almost) fully investing and (almost) not investing to a more uniform mixed strategy (i.e., investing or not investing with roughly equal probability).

A more thorough study confirms the above observation of the attacker and it is illustrated by Figure 7. Figure 7 shows: (a) the number of iterations taken by the smooth-best-response gradient-dynamics algorithm for ϵ-MSNE to converge (top left); (b) the number of nodes that are being targeted (top right); (c) the highest probability of attack (bottom left); and (d) the scatter plot of the nodes that are being targeted and the highest probability of attack (bottom right) for each of the ten simulations. From this figure, we observe that, as ϵ decreases, (1) the number of iterations takes for an ϵ-MSNE to converge increases (top left); (2) the number of nodes that are being targeted increases (top right); and (3) the highest probability of attack decreases (bottom left). From the bottom right graph of Figure 7, we observe that there is a negative correlation between the number of nodes that are being targeted and the highest probability of an attack: as the highest probability of an attack increases, the number of nodes that are being targeted decreases.

A possible reason to explain the behavior of the sites is that as more nodes become potential targets of a direct attack, more nodes with initial mixed strategies close to the "extreme" (i.e., very high or very low probabilities of investing) will move closer to a more uniform (and thus less "predictable") investment distribution.

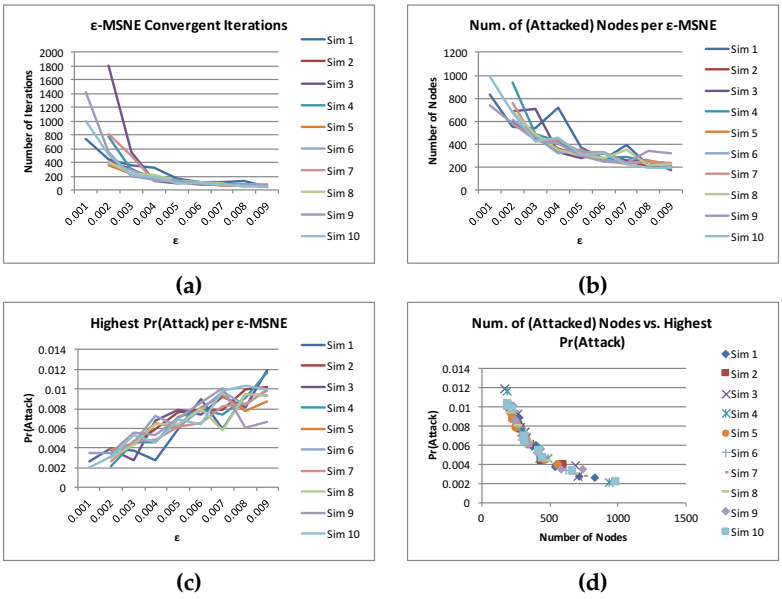


Figure 7. Attacker's Strategy at ϵ-MSNE. The x-axis of the graphs (**a**), (**b**), and (**c**) represents the ϵ value and their y-axis represents the number of iterations until convergence (or 2000 iterations max) to some ϵ-MSNE, the number of nodes that are being attacked, and the highest probability of attack, respectively. The scatter plot in graph (**d**) shows the relation between the number of nodes that are being attacked and the highest probability of attack in x-axis and y-axis, respectively.

Relation to Network Structure

Figure 8 presents some experimental results on the relationship between network structure and the attacker's equilibrium behavior. The graphs show, for each ϵ value, the average indegree and outdegree of those nodes that are potential targets of a direct attack at an ϵ-MSNE, across the BRGD runs of the ten randomly-generated IG instances. In general, both the average indegree and outdegree of the nodes that are potential targets of a direct attack tend to increase as ϵ decreases. One possible reason for this finding could be the fact that the values of α_i generated for each player are relatively low (i.e., uniformly distributed over $\left[0, \frac{1}{40}\right]$); yet, interestingly, such behavior and pattern, is the exact opposite of the theory for the case $\alpha_i = 1$.

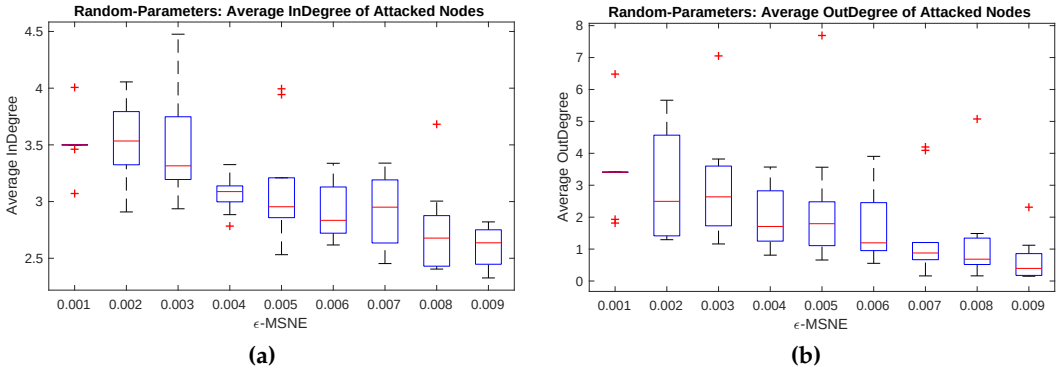


Figure 8. Attacker's ϵ-MSNE Strategy vs. Node Degrees. Average indegree (**a**) and outdegree (**b**) of nodes potentially attacked in terms of the ϵ-MSNE.

5.1.3. Case Study: A Randomly-Generated Instance of an IG at 0.005-MSNE

In this subsection, we provide a detailed topological study of a randomly-generated IG instance at 0.005-MSNE.

Topological Structure of an Attack to the Internet

In Figure 9, we plot the topological structure of the top sites (in this case 402) with the highest y_i and their immediate neighbors at 0.005-MSNE.

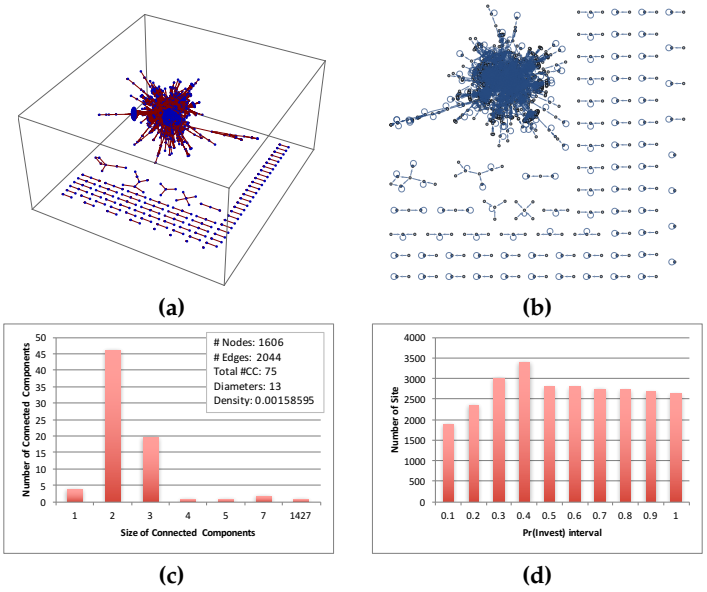


Figure 9. The Structure of an Attack to the Internet. The 3-d graph (**a**) corresponds to the top 402 Internet autonomous systems (AS)-level nodes most likely to be attacked according to our model at 0.005-MSNE, and their neighbors (i.e., both parent and children family). Graph (**b**) is a 2-d projection of the 3-d graph (**a**). The self-loops mark the nodes that are actually attacked. For the most part, the graph structures exhibit very dense clustering. Bar graph (**c**) corresponds to the number of *connected components* (*CCs*) of the top 402 Internet AS-level nodes that are most likely to be attacked. Bar graph (**d**) shows the number of nodes with the probability of investing in defense/security measures within the range of $([0, 0.1], (0.1, 0.2], ..., (0.9, 1])$. Some properties of the graph corresponding to the network structure are shown on the upper corner of graph (**c**). The graph consists of 1606 nodes, 2044 edges, and 75 CCs. Out of the 75 CCs, the largest CC contains 1427 nodes and the smallest CC consists of just 1 node (there are only 4 of them). There are 46 of 2-CC (CC with only 2 nodes), 20 of 3-CC, 1 of 4-CC, 1 of 5-CC, and 2 of 7-CC. The diameters and density of the graphs are 13 and 0.002, respectively.

Notice that there are a few isolated nodes and a few small "node-parent-children" networks, but in general, the largest network component tends to have a cluster-like structure. Figure 9 also shows the number of connected components of the network for the subgraph of the nodes most likely to be attacked (and their neighbors), as well as those of the network for the subgraph of the nodes with the highest probability of investing, along with some additional properties of the graphs.

Figures 10 and 11 show the indegree and outdegree of the (402) non-zero y_i nodes and the remaining (26704) zero y_i nodes, respectively. We did not observe in our experiments any strong relationship between the y_i's or x_i's in the ϵ-MSNE we found and the corresponding indegree or outdegree of the node i. However, we observed that, among the nodes with non-zero probability of attack, there was a slight tendency for those nodes with the lowest probability of attack to also have low outdegree and for those nodes with the highest probability of investing to also have low outdegree, but that tendency did not seem significant enough.

As mentioned earlier (Section 5.1.2), the behavior of the site players is quite sensitive to the ϵ value. Therefore, this could be one of the reasons that these nodes (with the highest y_i) have low outdegree.

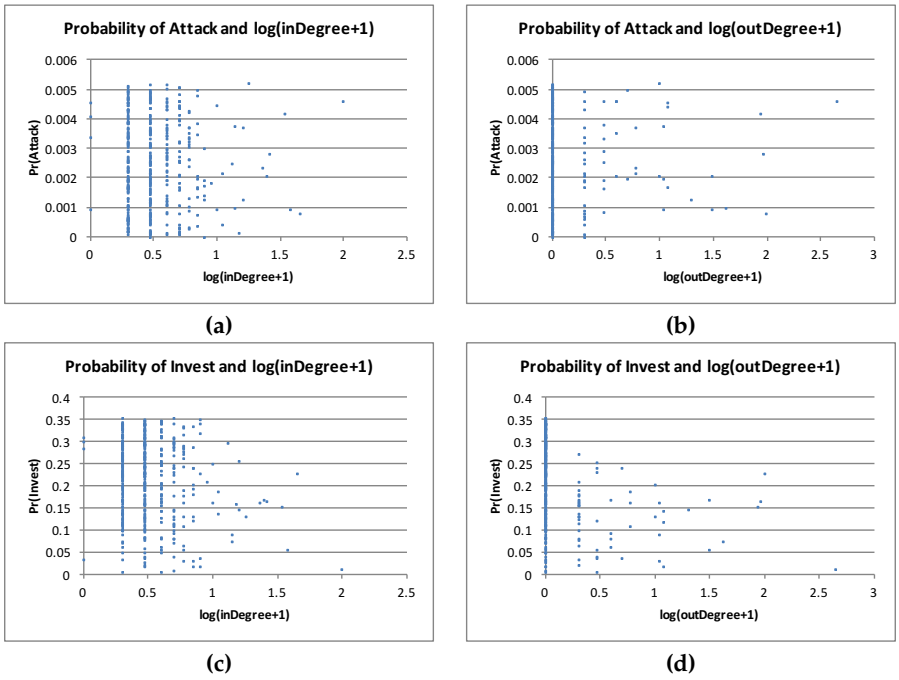


Figure 10. Attacker's Equilibrium Strategy vs Degree of the Nodes at 0.05-MSNE. These are plots on the 402 nodes with the highest y_i. The two graphs on top, (**a**) and (**b**), show the corresponding y_i (y-axis) and its indegree and outdegree in logarithmic (base 10) scale. Similarly, the two graphs at the bottom, (**c**) and (**d**), show the corresponding x_i (y-axis) and its indegree and outdegree in logarithmic (base 10) scale.

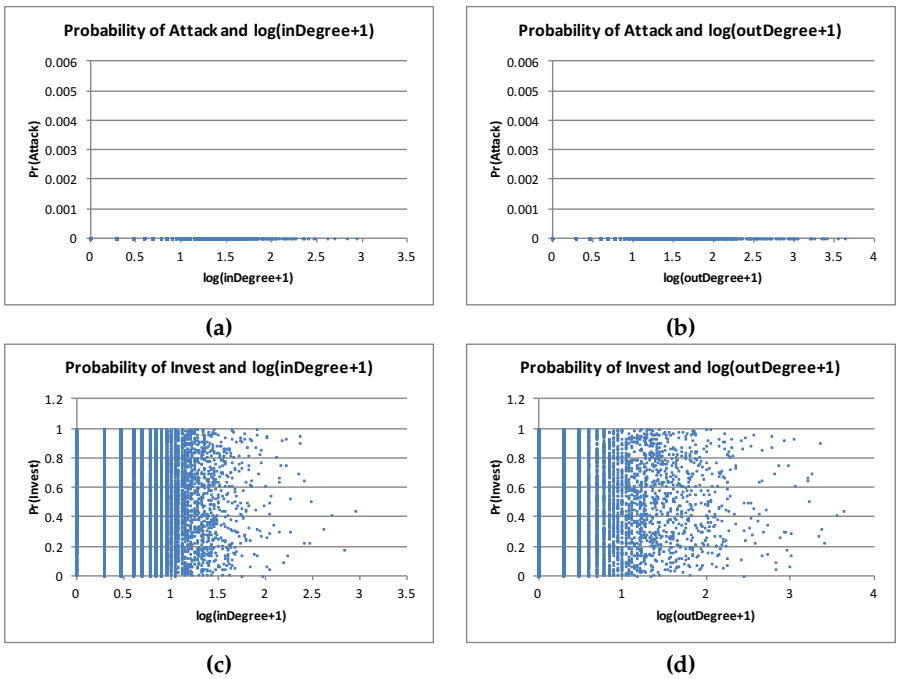


Figure 11. The Degrees and Strategies of Sites Not Directly Targeted. These are plots on the remaining 26,704 nodes with zero y_i. The two graphs on top, (**a**) and (**b**), show the corresponding y_i (y-axis) and its indegree and outdegree in logarithmic (base 10) scale. Similarly, the two graphs at the bottom, (**c**) and (**d**), show the corresponding x_i (y-axis) and its indegree and outdegree in logarithmic (base 10) scale.

5.2. A Heuristic to Compute ϵ-MSNE Based on Smooth Best-Response for the Attacker

In this subsection, we introduce an improved, simple heuristic to compute ϵ-MSNE on *arbitrary* graphs for lower ϵ values than those considered in the previous subsection (Section 5.1). Here, we evaluate the proposed hybrid BRGD-SBRD heuristic for computing ϵ-MSNE in IDD game using IGs randomly generated as described and used previously in this section.

We first look at the attacker's behavior at an ϵ-MSNE we obtain using BRGD in IGs. We generate a few IG instances and run BRGD until it converges to an ϵ-MSNE for $\epsilon \in \{0.001, 0.002, \ldots, 0.009\}$. We observe that in a 0.001-MSNE $(\mathbf{x}^*, \mathbf{y}^*)$, (1) there is a positive, almost-deterministic correlation between the probability of an attack y_i^* and each component $y_i^* U_i(\mathbf{x}^*)$ of the expected utility the attacker obtained for each site i, where $U_i(\mathbf{x}^*)$ is as defined in Equation 40, and (2) the attacker always target the sites with the highest potential utility $\max_i U_i(\mathbf{x}^*)$ (i.e., the maximum utility the attacker can get by attacking any site with probability 1). This observation is consistent with other IGs and holds across the different ϵ-MSNE for various ϵ values. Figure 12 shows evidence of this behavior. Indeed, the main take away is that the attacker tends to favor (or target) sites with highest expected utility. As observed, the attack seems to have some distributional form. Note that while one might expect this behavior given the way BRGD works, there is no theoretical guarantee for such behavior occuring at an approximate MSNE. This is because, in principle, any player may achieve a given approximation level without necessarily assigning a probability over each pure strategy in a way that is monotonically related to the expected payoff for executing that pure strategy deterministically, let alone the *linear* relationship we observe in the left-hand-side plot (**a**) of Figure 12. Similarly, given the last two statements, that the attacker is actually placing positive probability of attack only among those sites for which it would obtain highest maximum expected payoff, as the right-hand-side plot (**b**) of Figure 12 shows, is reassuring.

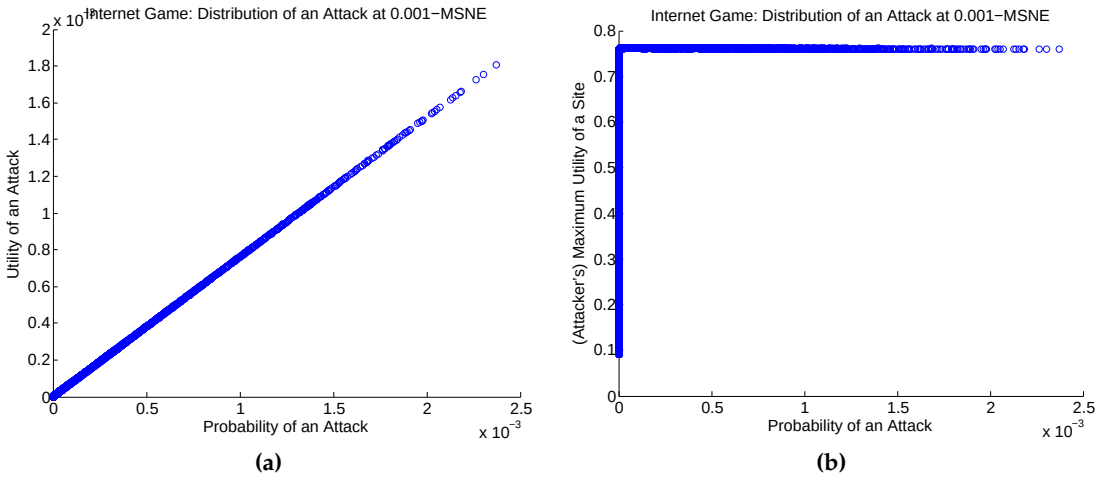


Figure 12. Attack Distribution of 0.001-MSNE $(\mathbf{x}^*, \mathbf{y}^*)$ Using best-response-gradient dynamics (BRGD) on Internet Game. Scatter plots of (y-axis) the component $y_i U_i(\mathbf{x}^*)$ of the (normalized) expected utility that the attacker obtained from attacking site i (**a**) and each expected utility $U_i(\mathbf{x}^*)$ that the attacker would obtain from fully targeting each site i (**b**) as a function of the probability y_i^* of attacking the corresponding site i (x-axis).

In what follows, we assume that the attacker is using *smoothed best-response* [35] and that the attack distribution is a *quantal-response mixed strategy* [53] (i.e., has the form of a *Gibbs-Boltzmann distribution*):

$$y_i^{\lambda}(\mathbf{x}) \equiv \frac{\exp(U_i(x_i, \mathbf{x}_{\mathrm{Ch}(i)}))/\lambda)}{\sum_{i=0}^{n} \exp(U_i(x_i, \mathbf{x}_{\mathrm{Ch}(i)})/\lambda)}\,, \tag{45}$$

where $U_0(x_i, \mathbf{x}_{\mathrm{Ch}(i)})$ is the normalized version of $U(\mathbf{x}, \mathbf{0}) = 0$ (i.e., after normalizing U to $[0, 1]$), and λ is a positive constant. Thus, the attacker's best-response correspondence is always a singleton, $\mathcal{BR}_0(\mathbf{x}) \equiv \{\mathbf{y}^\lambda(\mathbf{x})\}$, and thus essentially a strictly positive, vector-valued *function* of $\mathbf{x}$, for any given λ. This update is the result a slight modification of the attacker payoff function that adds a "controlled" entropic term to favor "diverse" attack probabilities, where λ controls diversity (i.e., the level of non-determinism of the attacker's best response): formally, given a positive real-value λ, the attacker's payoff is now $U(x, y) + \lambda H(y)$, where $H(y) \equiv \sum_{i=0}^{n} y_i \ln \frac{1}{y_i}$ is the standard *(Shannon) entropy function*. The interpretation of λ is that it controls the *precision* of the attacker and make the utility more distinct. The parameter λ is really the precision or *temperature parameter* of the Gibbs-Boltzmann distribution: increasing λ leads to the uniform distribution, while decreasing λ produces ϵ-MSNE with lower ϵ because λ restricts the effect of the entropic term in that case. In fact, at temperature $\lambda = 0$, we recover the original best-response for the attacker.

The form for the attacker's mixed strategy given in Equation (45) has several, natural attractive properties: (1) sites with high utility will have higher probability of an attack and (2) the respective expected utility and the probability of an attack are positively correlated (higher probability of attack implies higher expected utility gain). We observe these characteristics in our experiments (Figure 12).

Based on the previous discussion, we propose the following heuristic to compute ϵ-MSNE which refer to as the hybrid BRGD-SBRD heuristic. The heuristic starts by initializing all of the sites' investment level x_i to 0. It then updates the probability of attack for each site and increments the investment level of the site by a small amount (currently 0.001) for sites that do not satisfy the following condition: $R_i \geq y_i \hat{p}_i + (1 - \alpha_i) \sum_{j \in \mathrm{Pa}(i)} y_j (1 - x_j) \hat{q}_{ji}$. The algorithm terminates either when all of the sites satisfy the condition or when it reaches the maximum number of iterations. The condition, $R_i \geq y_i \hat{p}_i + (1 - \alpha_i) \sum_{j \in \mathrm{Pa}(i)} y_j (1 - x_j) \hat{q}_{ji}$, for site i is the threshold for i to not invest. A nice property of this is that given the attacker's Gibbs-Boltzmann distribution, for any site i, given the strategies of others, the attack decreases monotonically with x_i. As a result, no site has an incentive to unilaterally increase its investment to violate the constraint above. Consequently, to justify the use of the condition in the hybrid BRGD-SBRD heuristic in IGs, we observe that in all of the IGs we generated, the percentage of the sites at the 0.001-MSNE we obtained that satisfies the above condition is $\geq$98%. The quality of an ϵ-MSNE obtained by the hybrid BRGD-SBRD heuristic depends on the percentage of the sites that satisfy the condition at an ϵ-MSNE. Note that if a high percentage of the sites does not satisfy the condition at the ϵ-MSNE, we can reverse the heuristic by initializing all of the sites investment level x_i to 1 and lower the x_i's until all sites satisfy the opposite constraint.

Algorithm 1 provides pseudocode for the resulting hybrid BRGD-SBRD heuristic, in which the attacker uses smooth best-responses while the sites use best-response gradient, to compute an approximate MSNE in arbitrary IDD games as discussed.

5.3. Evaluation of the Hybrid BRGD-SBRD Heuristic on Internet Games

To evaluate the hybrid BRGD-SBRD heuristic, we randomly generated ten IGs and compare the results to those obtained using BRGD exclusively.

The first question we address is, what is the relation between the constant λ and the actual approximation quality ϵ achieved in practice? Table 2 shows the impact λ has on ϵ, for the smallest ϵ-MSNE we can obtain for an instance of the IGs (others are similar). The take-home message is that ϵ deceases with λ as expected. For the remaining of this section, we will fix $\lambda = 0.001$ when comparing to BRGD as BRGD cannot find ϵ-MSNE beyond 0.0009-MSNE within 10 K iterations (1 s per iteration).

Algorithm 1: Heuristic Based on Hybrid of BRGD and *Smooth Best-Response Dynamics (SBRD)* to Compute an ϵ-MSNE in Single-Attack IDD Games

Input : An instance of an n-player IDD game, $T_{\max}$
Output: $(\mathbf{x}, \mathbf{y})$ - An ϵ-MSNE
Let $x_i \leftarrow 0$ for all $i = 1, 2, ..., n$
Let iteration $\leftarrow 0$
Let increment $\leftarrow 0.001$
Let Converge $\leftarrow$ false
while *not Converge AND iteration* $< T_{\max}$ **do**
  Converge = true
  $\overline{y}_i \leftarrow \exp\left(\frac{U_i(\mathbf{x})}{\lambda}\right)$ for all $i = 0, 1, 2, ..., n$
  $y_i = \frac{\overline{y}_i}{\sum_{i=0}^{n} \overline{y}_i}$ for all $i = 1, 2, ..., n$
  foreach $i = 1, 2, ...n$ **do**
    if $R_i < y_i \hat{p}_i + (1 - \alpha_i) r_i(\mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)})$ **then**
      $x_i = x_i +$ increment (if $x_i > 1, x_i = 0$)
      Converge = false
    end
  end
  iteration = iteration + 1
end

Table 2. Selection of the Constant λ for the Hybrid BRGD-SBRD Heuristic.

λ	**Smallest** ϵ
0.05	0.06
0.01	0.008
0.005	0.004
0.001	0.0009
0.0005	0.0006
0.0001	0.0004

5.3.1. Comparing Running Time of BRGD and the Proposed Hybrid BRGD-SBRD Heuristic

Next we study the convergence time to an ϵ-MSNE on the ten IG instances when using (1) BRGD exclusively and (2) the hybrid heuristic. We consider the running time in terms of the number of iterations that the algorithm takes to achieve a particular ϵ-MSNE. Each iteration is roughly 1 s for both BRGD and the hybrid heuristic. Figure 13a shows that the running time of the hybrid heuristic is considerably faster than BRGD. The rate at which the number of iterations increases as ϵ decreases seems extreme for the hybrid heuristic—it is almost constant—relative to that for BRGD. Not only is the hybrid heuristic faster than BRGD but it can also find ϵ-MSNE with lower ϵ values.

As an application, we could run our heuristic until it reaches an ϵ-MSNE or converges. Then use the output of our ϵ-MSNE to initialize BRGD. Figure 14 shows the relative improvement over the hybrid heuristic on some IGs. It improves our 0.001/0.0009-MSNE to 0.0006-MSNE.

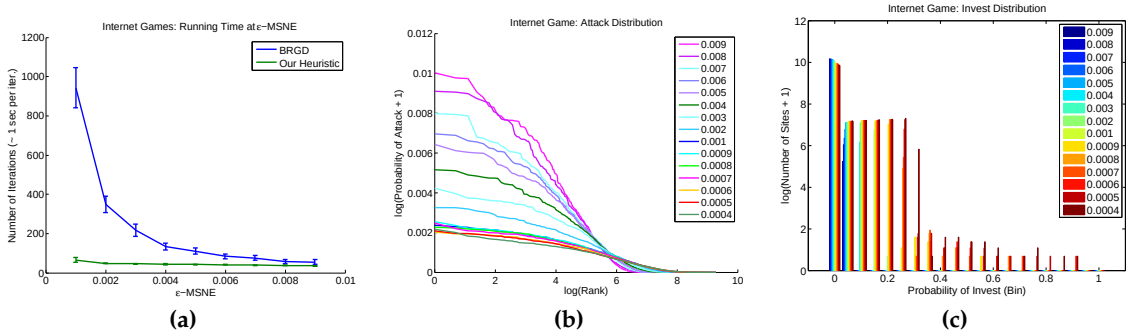


Figure 13. Properties of the Hybrid BRGD-Smooth Best-Response Dynamics (SBRD) Heuristic. BRGD vs. hybrid heuristic running time (**a**); Attacker's attack (**b**) and sites' (**c**) investment distribution on ϵ-MSNE.

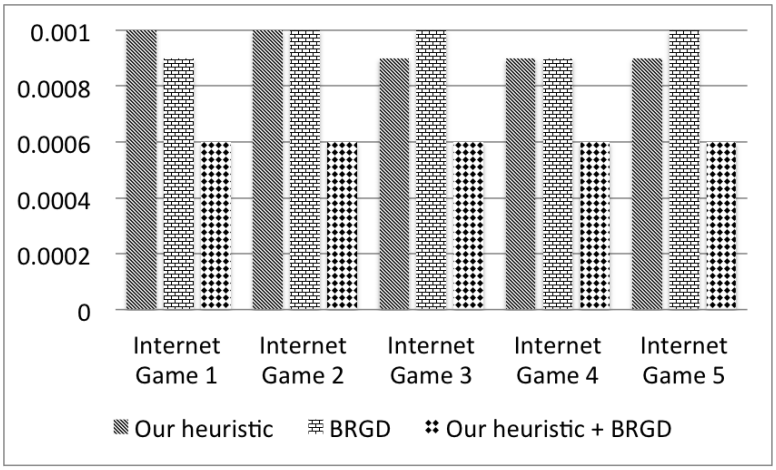


Figure 14. Combing BRGD and the Hybrid BRGD-SBRD Heuristic. Internet Games: BRGD Improvement (y-axis represents the ϵ values).

5.3.2. Attacker and Sites' Equilibrium Behavior

We study whether the equilibrium behavior of the attacker and sites that we observed and discussed in Section 5.1 remains the same, or is similar, as we lower ϵ. The following results are a direct output of our heuristic. Figure 13b shows the attack distribution (left) and the investment distribution (right) at ϵ-MSNE, for different ϵ values, on an IG instance. Our results are consistent with those in Section 5.1, and persist for lower ϵ values. We see that as ϵ decreases, the attacker targets more sites while lowering the probability of the direct attack, and more sites move from not invest to partially invest.

5.3.3. Network Structure of an Attack

Next, we present experimental results on the average indegree and outdegree of the targeted sites at ϵ-MSNE to understand the "network structure of the attack" as we did in Section 5.1. Figure 15 shows exactly this. To summarize our experimental results, we can clearly observe that as ϵ decreases both the average indegree and outdegree increase. The results for lower ϵ values indicate the average indegree and outdegree are stabilizing and converging as ϵ decreases. This is also consistent with the observations made in Section 5.1. This consistency also adds evidence to the effectiveness of our proposed heuristic for very low ϵ values.

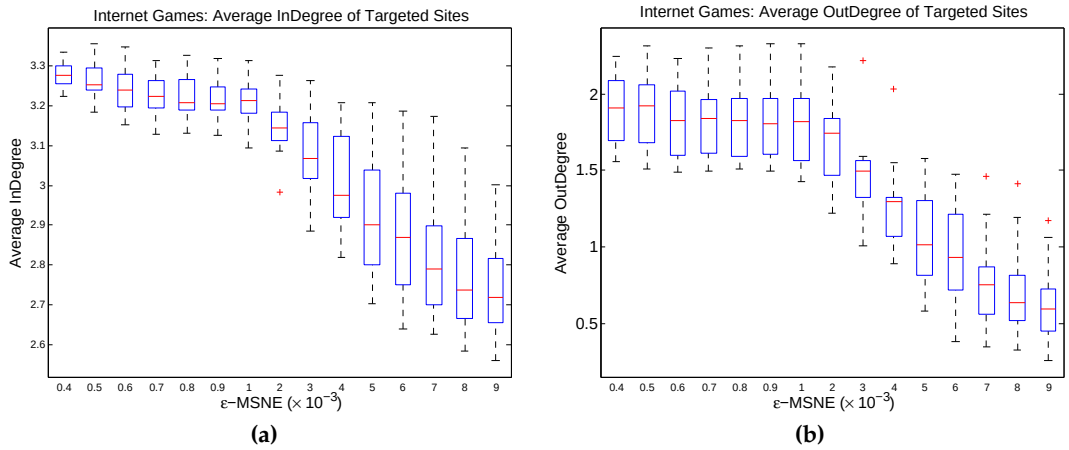


Figure 15. Degrees of the Targeted Site Nodes at ϵ-MSNE. Internet Games: average indegree (**a**) and average outdegree (**b**) of the targeted sites over ϵ-MSNE

6. Future Work, Open Problems, and a Summary of Our Contributions

We end by discussing future work and some open problems, and providing a brief summary of our contributions.

6.1. Future Work

In this subsection, we discuss potential avenues of future work.

6.1.1. Attackers Can Affect Transfer Probabilities

We could extend the strategy space of the attacker by allowing the attacker to affect transfer. One particular instantiation of this idea is to have the network graph *edges* represent the attacker's targets, as opposed to just the node. The attacker's pure strategies would now be based on the edges (i, j), such that binary action variable b_{ij} would now represent the attack, taking a value of one if the attacker wants to attack j but only via a transfer from i.

6.1.2. Multiple Attackers with Multiple Attacks

While dealing with multiple attackers is outside the scope of this paper, we have in fact extended the model in a natural way in that direction. We refer the reader to Appendix E for a discussion of this setting. In short, we were able to extend the representation results, but not the characterization or computational/algorithmic results. We leave that endeavor for future work. In principle, we can use BRGD and the hybrid BRGD-SBRD as heuristics in the case of multiple attackers.

6.1.3. Learning IDD Games

Another interesting research direction is the adaptation of machine-learning techniques to infer IDD games from data for the purpose of compactly representing stable outcomes and performing strategic inference.

6.1.4. Other Open Problems

A thorough characterization of the equilibria of interdependent defense games is lacking, specially for the case of multiple potential attacks by multiple aggressors. Also, we need a better understanding of the effect of network structure of the game and restrictions on the aggressors' available strategies on the equilibria of the game.

Many computational problems in the context of interdependent defense games remain open.

1. What is the computational complexity of the problem of computing equilibria of single-attack IDD games with arbitrary transfer vulnerability? (e.g., a single, multiple or all MSNE? MSNE with particular properties?)
2. What is the computational complexity of the problem of identifying "influential" agents, in the sense of Irfan and Ortiz [41] (see also, Kleinberg [54] and the references therein)?
3. How is the complexity affected by network structure or restrictions on the aggressors' available strategies? While we provide FPTAS for approximate MSNE in single-attack IDD games with arbitrary transfer vulnerability values and directed tree graphs over the sites, it is fair to say that there is more work to do in that direction. A particularly interesting question is whether we can establish PPAD-completeness results for arbitrary single-attack IDD games. That would strengthen the hardness (NP-complete) result we present in Section 4.2. The relationship between single-attack IDD games and graphical polymatrix games seems particularly close. Perhaps one may be able to apply existing results on the PPAD-completeness of certain classes of graphical polymatrix games to establish PPAD-completeness in our context. A promising direction is to pursue potential reductions using results such as those of (Cai and Daskalakis [50], Theorem 1.2) and (Daskalakis et al. [29], Lemma 6.3) on (graphical) polymatrix games with "strictly competitive games on the edges" and on 3-ADDITIVE GRAPHICAL NASH, respectively.

6.2. Summary of Contributions

In this paper, we propose IDD games, an adaptation of IDS games to the setting in which the attack is deliberate and the attacker is explicitly modeled. We consider the special case of the single attack scenario as a way to limit the attackers power, and prove that no PSNE exists in such subclass of games. We then consider randomized strategies and derive the appropriate expressions for the expected costs of the internal players and the expected payoff of the attacker, and consequently their respective best-response correspondences.

We study in depth the case in which only one attack is possible and investment in security does nothing to protect the players from the transfer risk (which is the same implicit assumption made in the original IDS work). We completely characterize the MSNE of such a subclass of IDD games. We prove that almost every game in that subclass has a unique MSNE, which can be almost-fully determined analytically.

That result immediately lead to a simple algorithm for computing the equilibrium that only requires a sorting of the cost-to-expected-loss ratio gain of the attacker for each player. Hence, the algorithm runs in $O(n \log n)$, where n is the number of internal players.

We then discuss some corollaries of the characterization and highlight the connection between the network structure and the investment of players at equilibrium. In particular, we show how investment probabilities at equilibrium essentially reflect some degree of "cooperation" (in a fully non-cooperative setting). It turns out that, at an equilibrium, players want to protect their own *children* in the network graph. That is, each player i wants to protect the set of players to which player i can *transfer*). Yet, somewhat counterintuitively at first, each player i's security investment level has no direct dependence on the player's *parents*, who are the true source of the risk to the player. In particular, we show how the probability of investment can is *directly proportional* to the fixed, but arbitrary number of children in the given directed network.

We also provided a hardness result for computing exact MSNE in arbitrary single-attack IDD games. We also designed an FPTAS to compute an ϵ-MSNE in graph with directed trees over the sites, despite the fact that the attacker's payoff is a function of all the sites.

Finally, inspired by problems in cyber and network security, we designed a generator of random instances of IDD games based on a real-world instance of the AS-level Internet graph. We call the resulting instances Internet games (IGs) . We studied simple heuristics based on BRGD, SBRD, and a hybrid, for computing approximate MSNE in IGs. We evaluated the running times of the different heuristics and found that BRGD, when run exclusively, seems to be effective at computing ϵ-MSNE for

ϵ values as low as 0.001. We studied the characteristics of the approximate MSNE that BRGD produces in terms of both sites' and attacker's behavior and the relationship to the IGs network structure. We then proposed and studied a hybrid BRGD-SBRD heuristic in which the attacker uses SBRD while the sites use BRGD. It allows us to compute ϵ-MSNE with an order of magnitude lower in ϵ values very quickly in IG instances. Using the hybrid heuristic we successfully verified some of the conclusions we had obtained using BRGD alone, but now consider considerably lower ϵ values. We also showed how we can improve the effectiveness both in terms of speed and quality when using BRGD alone, by first running the hybrid heuristic and then starting BRGD alone from the ϵ-MSNE values that the hybrid heuristic found.

We view our work as an initial illustration of the potential for the type of quantitative analysis that may be possible in network security settings, even at the real-world Internet scale.

Acknowledgments: This work was funded in part by a National Science Foundation (NSF) Graduate Research Fellowship, an Office of Naval Research's MINERVA Grant, and an NSF Faculty Early Career Development Program (CAREER) Award IIS-1643006 (transferred from IIS-1054541).

Author Contributions: H.C. and L.O. contributed equally to all aspects of this work, including the writing of the paper, model design and development, theoretical/mathematical analysis, algorithm design, conception, design, and execution of experiments, and evaluation and presentation of empirical results. M.C. helped to initiate work on the IDD model, to design the random generator of Internet Games using the DIMES AS-level Internet graph data, to implement the BRGD heuristics, and to run and collect preliminary data based on that heuristic. He did that as part of his Undergraduate Honor's Project, under the guidance of L.O. Even though M.C. did not write this submission, there are several parts of his original project report still embedded within this paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AS	autonomous systems
BRGD	best-response gradient dynamics
CC	connected components
DP	dynamic programming (or program)
FPTAS	fully polynomial-time approximation scheme
IDD	interdependent security games
IDS	interdependent security
IG	Internet game
MSE	mean-squared-error
MSNE	mixed-strategy Nash equilibrium (or equilibria)
PSNE	pure-strategy Nash equilibrium (or equilibria)
SBRD	smooth best-response dynamics

Appendix A

For the reader's convenience, Table A1 provides a summary of the most relevant notation used throughout the Appendices.

Table A1. Notation Legend.

Symbol	Semantics
n	number of sites
$[n]$	$\{1, 2, \ldots, n\}$
a_i	action (pure strategy) of site i: $a_i = 1$ ("invest") or $a_i = 0$ ("not invest")
$\mathbf{a}$	joint action (pure strategy) for all sites: $(a_i)_{i\in[n]}$
$\mathbf{b}$	pure strategy of attacker
b_i	component of attacker's pure strategy corresponding to site i (i.e., $\mathbf{b} \equiv (b_i)_{i\in[n]}$): $b_i = 1$ if attacker directly targets site i; $b_i = 0$ otherwise
$\mathcal{B}$	$\left\{\mathbf{b} \in \{0,1\}^n \;\middle|\; \sum_{i=1}^n b_i \leq 1\right\}$ (Assumption 3)
C_i	cost to site i of investing in security
$\mathbf{C}$	$(C_i)_{i\in[n]}$ (Definition 1)
C_i^0	cost to attacker for directly targeting site i
L_i	loss to site i should it experience the "bad event"
$\mathbf{L}$	$(L_i)_{i\in[n]}$ (Definition 1)
α_i	probability that the transfer of the "bad event" will *not* caught *given* site i invest in security (i.e., $a_i = 1$)
$\boldsymbol{\alpha}$	$(\alpha_i)_{i\in[n]}$
$\widehat{p}_i$	*conditional* probability that site i experience the "bad event" *given* that site i was a direct target (Equation (8))
$\widehat{\mathbf{p}}$	$(\widehat{p}_i)_{i\in[n]}$ (Definition 5)
$\widehat{q}_{ij}$	*conditional* probability that site j experience the "bad event" (Equation (9)) as result of a transfer from site i
$\widehat{\mathbf{Q}}$	matrix composed of the $\widehat{q}_{ij}$'s (Definition 5)
$\widehat{\Delta}_i$	ratio of cost to conditional expected loss (Equation (13)): $\frac{C_i}{\widehat{p}_i L_i}$
G	directed network graph of sites: $([n], E)$
$\mathrm{Pa}(i)$	set of sites that are parent of site i in G
$\mathrm{PF}(i)$	site i's parent family: $\mathrm{Pa}(i) \cup \{i\}$
k_i	$\lvert\mathrm{PF}(i)\rvert$
$k_{\max}$	$\max_{i\in[n]} k_i$
$\mathrm{Ch}(i)$	set of sites that are children of site i in G
$\mathrm{CF}(i)$	site i's children family: $\mathrm{Ch}(i) \cup \{i\}$
$e_{ij}(a_j, b_j)$	probability that site i is safe from j (Equation (10)): $a_j + (1 - a_j)(1 - b_j\widehat{q}_{ji})$
$s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})$	external overall *safety* of site i (Equation (11)): $\prod_{j\in\mathrm{Pa}(i)} e_{ij}(a_j, b_j)$
$r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})$	external overall *risk* of site i (Equation (11)): $1 - s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})$
$M_i(\mathbf{a}, \mathbf{b})$	cost function of site i (Equation (12)): $M_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)})$
$U(\mathbf{a}, \mathbf{b})$	payoff function of attacker (Equation (16))
x_i	site i's individual mixed strategy: probability of investing (i.e., probability assigned to $a_i = 1$)
$\mathbf{x}$	joint mixed strategy of all sites: $(x_i)_{i\in[n]}$
P	mixed strategy of attacker
y_i	probability that attacker directly targets site i: $P(b_i = 1)$
y_0	probability of no attack: $P(\mathbf{b} = \mathbf{0})$; under Assumption 3, $y_0 = 1 - \sum_{i=1}^n y_i$
$\mathbf{y}$	compact representation of attacker's mixed strategy under Assumption 3: $(y_i)_{i\in[n]}$, where $y_i = P(b_i = 1) = P(b_i = 1, \mathbf{b}_{-i} = \mathbf{0})$ and $\sum_{\mathbf{b}} P(\mathbf{b}) = \sum_{i=0}^n y_i = 1$
$L_i^0(x_i)$	$(1 - x_i)(\widehat{p}_i L_i + \sum_{j\in\mathrm{Ch}(i)} \widehat{q}_{ij} L_j)$
$\overline{L}_i^0$	$L_i^0(0) = \widehat{p}_i L_i + \sum_{j\in\mathrm{Ch}(i)} \widehat{q}_{ij} L_j$ (Equation (41))
$M_i^0(x_i)$	$L_i^0(x_i) - C_i^0$
$\overline{M}_i^0$	$M_i^0(0) = \overline{L}_i^0 - C_i^0$ (Equation (42))
η_i^0	$C_i^0 / \overline{L}_i^0$ (Equation (43))
$U(\mathbf{x}, \mathbf{y})$	expected payoff of attacker under Assumption 3 (Equation (44)): $\sum_{i=1}^n y_i M_i^0(x_i)$
$r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)})$	external overall *risk* of site i when the sites in $\mathrm{Pa}(i)$ use joint action $\mathbf{a}_{\mathrm{Pa}(i)}$ (Equation (34)): $\sum_{j\in\mathrm{Pa}(i)} y_j(1 - a_j)\widehat{q}_{ji}$

Appendix B. Proofs Missing from the Main Body of the Article

Here, we present all the proofs that we moved out of the main body of the article in order to maintain a fluid presentation.

Appendix B.1. Proof of Lemma 1

By the definition of single-attack IDD games (Definition 7), any attacker pure strategy $\mathbf{b}$ in $\mathcal{B}$, as defined in Assumption 3, is either a vector of all 0's, or exactly one 1. Because $e_{ij}(a_j, 0) = 1$ (Equation (10)), by the definition of s_i (Equation (11)), we have

$$\begin{aligned} s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) &= \prod_{j \in \mathrm{Pa}(i)} e_{ij}(a_j, b_j) \\ &= \prod_{j \in \mathrm{Pa}(i)} [e_{ij}(a_j, 1)]^{b_j} \\ &= \begin{cases} \sum_{j \in \mathrm{Pa}(i)} b_j e_{ij}(a_j, 1), & \text{if } b_k = 1 \text{ for some } k \in \mathrm{Pa}(i), \\ 1, & \text{if } b_k = 0 \text{ for all } k \in \mathrm{Pa}(i), \end{cases} \\ &= 1 - \sum_{j \in \mathrm{Pa}(i)} b_j (1 - a_j) \widehat{q}_{ji}, \end{aligned}$$

so that

$$r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) = \sum_{j \in \mathrm{Pa}(i)} b_j (1 - a_j) \widehat{q}_{ji},$$

and

$$b_i \, s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) = b_i.$$

Appendix B.2. Proof of Proposition 1

From Lemma 1, and the definition of U (Equation (16)), we obtain

$$\begin{aligned} U(\mathbf{a}, \mathbf{b}) &= \sum_{i=1}^{n} M_i(\mathbf{a}, \mathbf{b}) - a_i C_i - b_i C_i^0 \\ &= \sum_{i=1}^{n} (a_i \alpha_i r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) + (1 - a_i)(b_i \widehat{p}_i + (1 - b_i \widehat{p}_i) r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}))) L_i - b_i C_i^0 \\ &= \sum_{i=1}^{n} b_i (1 - a_i) \widehat{p}_i L_i + (a_i \alpha_i + (1 - a_i)(1 - b_i \widehat{p}_i)) r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) L_i - b_i C_i^0 \, . \end{aligned}$$

After distributing the sum over each term, for the second term inside the sum, by the expression for the r_i's (Equation (18)), we have

$$\begin{aligned}
&\sum_{i=1}^{n}(a_i\alpha_i + (1-a_i)(1-b_i\widehat{p}_i))r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})L_i \\
&= \sum_{i=1}^{n}(a_i\alpha_i + (1-a_i)(1-b_i\widehat{p}_i))\left(\sum_{j\in Pa(i)} b_j(1-a_j)\widehat{q}_{ji}\right)L_i \\
&= \sum_{i=1}^{n}\sum_{j\in Pa(i)}(a_i\alpha_i + (1-a_i)(1-b_i\widehat{p}_i))\left(b_j(1-a_j)\widehat{q}_{ji}\right)L_i \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{1}[j\in \mathrm{Pa}(i)](a_i\alpha_i + (1-a_i)(1-b_i\widehat{p}_i))b_j(1-a_j)\widehat{q}_{ji}L_i \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{1}[i\in \mathrm{Ch}(j)](a_i\alpha_i + (1-a_i)(1-b_i\widehat{p}_i))b_j(1-a_j)\widehat{q}_{ji}L_i \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n}\mathbb{1}[i\in \mathrm{Ch}(j)](a_i\alpha_i + (1-a_i)(1-b_i\widehat{p}_i))b_j(1-a_j)\widehat{q}_{ji}L_i \\
&= \sum_{j=1}^{n}\sum_{i\in \mathrm{Ch}(j)}(a_i\alpha_i + (1-a_i)(1-b_i\widehat{p}_i))b_j(1-a_j)\widehat{q}_{ji}L_i \\
&= \sum_{j=1}^{n}\sum_{i\in \mathrm{Ch}(j)}(b_ja_i\alpha_i + (1-a_i)(b_j-b_jb_i\widehat{p}_i))(1-a_j)\widehat{q}_{ji}L_i \\
&= \sum_{j=1}^{n}\sum_{i\in \mathrm{Ch}(j)}(b_ja_i\alpha_i + (1-a_i)(b_j-0))(1-a_j)\widehat{q}_{ji})L_i \\
&= \sum_{j=1}^{n}\sum_{i\in \mathrm{Ch}(j)}(b_ja_i\alpha_i + (1-a_i)b_j)(1-a_j)\widehat{q}_{ji}L_i \\
&= \sum_{j=1}^{n} b_j(1-a_j)\sum_{i\in \mathrm{Ch}(j)}(a_i\alpha_i + (1-a_i))\widehat{q}_{ji}L_i \\
&\quad \sum_{i=1}^{n} b_i(1-a_i)\sum_{j\in \mathrm{Ch}(i)}(a_j\alpha_j + (1-a_j))\widehat{q}_{ij}L_j\,.
\end{aligned}$$

The result follows by simple substitutions and some algebra.

Appendix B.3. Proof of Proposition 2

If the single-attack IDD game has a PSNE $(\mathbf{a}^*, \mathbf{b}^*)$, then, by Proposition 1, we can express the attacker's payoff in it as

$$U(\mathbf{a}^*, \mathbf{b}^*) = \left[\max_{i\in[n]}(1-a_i^*)\left(\widehat{p}_iL_i + \sum_{j\in \mathrm{Ch}(i)}\widehat{q}_{ij}(a_j^*\alpha_j + (1-a_j^*))L_j\right) - C_i^0\right]^+,$$

where for any real number $z \in \mathbb{R}$, the operator $[z]^+ \equiv \max(z, 0)$. In addition, if $b_l^* = 1$ for some $l \in [n]$, then

$$\begin{aligned}
&(1-a_l^*)\left(\widehat{p}_lL_l + \textstyle\sum_{j\in \mathrm{Ch}(l)}\widehat{q}_{lj}(a_j^*\alpha_j + (1-a_j^*))L_j\right) - C_l^0 \geq \\
&\left[\max_{i\in[n]}\,(1-a_i^*)\left(\widehat{p}_iL_i + \textstyle\sum_{j\in \mathrm{Ch}(i)}\widehat{q}_{ij}(a_j^*\alpha_j + (1-a_j^*))L_j\right) - C_i^0\right]^+.
\end{aligned} \tag{46}$$

The proof of the proposition is by contradiction. Consider an IDD game that satisfies the conditions of the proposition. Let $(\mathbf{a}^*, \mathbf{b}^*)$ be a PSNE of the game. We need to consider two cases at the PSNE: (1) there is some attack and (2) there is no attack.

1. If there is some attack, then $b_l^* = 1$ for some site $l \in [n]$, and for all $i \neq l$, $b_i^* = 0$. In addition, because $\mathbf{b}^*$ is consistent with the aggressor's best response to $\mathbf{a}^*$, we have, using the condition given in Equation (46) above,

$$(1 - a_l^*) \left(\widehat{p}_l L_l + \sum_{j \in \mathrm{Ch}(l)} \widehat{q}_{lj} (a_j^* \alpha_j + (1 - a_j^*)) L_j \right) \geq C_l^0 > 0 \,, \tag{47}$$

   The last condition in Equation (47) and Assumption 2 implies $a_l^* = 0$. Hence, by the best-response condition of site l, we have

$$C_l + \alpha_l r_l(\mathbf{a}^*_{\mathrm{Pa}(l)}, \mathbf{b}^*_{\mathrm{Pa}(l)}) L_l \geq \widehat{p}_l L_l + (1 - \widehat{p}_l) r_l(\mathbf{a}^*_{\mathrm{Pa}(l)}, \mathbf{b}^*_{\mathrm{Pa}(l)}) L_l \,.$$

   Because the attack occurs at l, the transfer risk $r_l(\mathbf{a}^*_{\mathrm{Pa}(l)}, \mathbf{b}^*_{\mathrm{Pa}(l)}) = r_l(\mathbf{a}^*_{\mathrm{Pa}(l)}, \mathbf{0}) = 0$ at the PSNE. Therefore, the last condition simplifies to

$$C_l \geq \widehat{p}_l L_l \,,$$

   which contradicts Assumption 1.
2. If there is no attack, then $\mathbf{b}^* = \mathbf{0}$. In this case, the site's best-response conditions imply $\mathbf{a}^* = \mathbf{0}$. From the attacker's best-response condition we obtain

$$\widehat{p}_l L_l + \sum_{j \in \mathrm{Ch}(l)} \widehat{q}_{lj} L_j \leq C_l^0 \,,$$

   which contradicts Assumption 2.

Appendix B.4. Proof of Proposition 3

Let $H(P) \equiv \sum_{\mathbf{b} \in \{0,1\}^n} P(\mathbf{b}) \ln \frac{1}{P(\mathbf{b})}$ be *Shannon's* entropy function, corresponding to a set of Bernoulli random variables with joint PMF P [55]. Consider the *maximum-entropy (MaxEnt) distribution* [46] resulting from the following optimization with respect to an MSNE $(\mathbf{x}^*, P^*)$ of an arbitrary IDD game.

$$\arg\max_{P \in \mathcal{P}} H(P)$$

such that, for all $i \in [n]$, and $\mathbf{b}_{\mathrm{PF}(i)} \in \{0,1\}^{k_i}$ with $P^*_{\mathrm{PF}(i)}(\mathbf{b}_{\mathrm{PF}(i)}) > 0$,

$$P(\mathbf{b}_{\mathrm{PF}(i)}) = P^*_{\mathrm{PF}(i)}(\mathbf{b}_{\mathrm{PF}(i)})$$

Note that the optimization problem is feasible because $P^* \in \mathcal{P}$. A standard duality argument [56] shows that we can express the solution to the optimization problem as

$$P^{\lambda^*}(\mathbf{b}) \propto \exp\left(\sum_{i=1}^{n} \sum_{\mathbf{b}_{\mathrm{PF}(i)} \in \{0,1\}^{k_i}, P^*_{\mathrm{PF}(i)}(\mathbf{b}_{\mathrm{PF}(i)}) > 0} \lambda^*_{i, \mathbf{b}_{\mathrm{PF}(i)}} \right) = \prod_{i=1}^{n} \exp\left(\sum_{\mathbf{b}_{\mathrm{PF}(i)} \in \{0,1\}^{k_i}, P^*_{\mathrm{PF}(i)}(b_{\mathrm{PF}(i)}) > 0} \lambda^*_{i, \mathbf{b}_{\mathrm{PF}(i)}} \right)$$

for some vector of real-valued parameters λ^* corresponding to the *dual variables* (also known as the *Lagrange multipliers*). We can now define PMF $\widetilde{P}$ over $\{0,1\}^n$ as

$$\widetilde{P}(\mathbf{b}) \propto \prod_{i=1}^{n} \left(\mathbb{1}\left[P^*_{\mathrm{PF}(i)}(\mathbf{b}_{\mathrm{PF}(i)}) > 0\right] \exp\left(\sum_{\mathbf{b}_{\mathrm{PF}(i)} \in \{0,1\}^{k_i}, P^*_{\mathrm{PF}(i)}(\mathbf{b}_{\mathrm{PF}(i)}) > 0} \lambda^*_{i,\mathbf{b}_{\mathrm{PF}(i)}} \right) \right) .$$

The optimization constraints and the fact that $P^*(\mathbf{b}) = 0$ if and only if $\mathbf{b}$ is such that $P^*_{\mathrm{PF}(i)}(\mathbf{b}_{\mathrm{PF}(i)}) = 0$ for some i, yield Parts 1 and 2 of the proposition. To see that Part 3 also holds, note that by the structure of the sites' costs, and thus the attacker's payoff, their expected value depends only on the joint marginal PMFs $P_{\mathrm{PF}(i)}$ for all i. Thus, Part 3 actually follows immediately from Part 2.

Appendix B.5. Proof of Lemma 2

Equation (33) follows from Lemma 1 (Equation (18)), and linearity of expectation. Equation (34) follows from Equation (33) and linearity of expectation. Equation (35) follows from Equation (34) and linearity of expectation. Equation (36) follows from Lemma 1 (Equation (19)) and the definition of expected value. Equation (37) follows from Equations (34) and (36).

Appendix B.6. Proof of Proposition 4

Equation (38) follows from Lemma 2, and applying Equations (34) and (36) to Equation (26). Equation (40) follows from Lemma 1 (Equation (21)), linearity of expectation, the definition of a joint mixed strategy (i.e., a product distribution), and the fact that each term in the sum is either the product of a linear function of a_i and either a constant or a linear function of a_j for $j \in \mathrm{Ch}(i)$. Equation (39) follows from Equation (40) and linearity of expectation.

Appendix B.7. Proof of Proposition 5

Throughout this proof, by the hypothesis of the proposition, we assume we are dealing with fully transfer-vulnerable single-attack IDD games. We also use the same notation as that introduced before the statement of the proposition in the main body of the manuscript. The same holds for the proof of Claim 1 in the next subsection of this appendix (Appendix B.8). We remind the reader that Table A1 provides a summary of the most relevant notation used throughout the Appendices.

We begin the proof by noting that Proposition 4 implies that the best-response $\mathcal{BR}_i$ of defender i *directly* depends on y_i *only*. Said differently, $\mathcal{BR}_i$ is *conditionally* independent of the mixed strategies $\mathbf{x}_{\mathrm{Pa}(i)}$ of its parent nodes $\mathrm{Pa}(i)$ of defender node i in the network *given* the probability y_i that the attacker's mixed-strategy assigns to a direct attack to i. Thus, in what follows, we abuse notation and define

$$\mathcal{BR}_i(y_i) \equiv \mathcal{BR}_i(\mathbf{x}_{\mathrm{Pa}(i)}, P_{\mathrm{PF}(i)}) = \begin{cases} \{1\}, & \text{if } y_i > \widehat{\Delta}_i, \\ \{0\}, & \text{if } y_i < \widehat{\Delta}_i, \\ [0,1], & \text{if } y_i = \widehat{\Delta}_i. \end{cases}$$

Next, we prove some useful properties of the MSNE. [16]

Claim 2. *In every MSNE* $(\mathbf{x}, \mathbf{y})$*, for all* $i \in [n]$*, if the probability of a direct attack to a defender* i *is* $y_i = 0$ *then the probability of investment of defender* i *is* $x_i = 0$*. In addition, if* $y_i = 0$ *for some* $i \in [n]$ *then the probability of no attack* $y_0 = 0$*.*

[16] Throughout the proof, to simplify notation, we drop the '*' superscript used in the main text to denote MSNE.

Proof. By $\mathcal{BR}_i$, $y_i = 0 < \widehat{\Delta}_i$ implies $x_i = 0$. For the second part, if $y_i = 0$ for some defender $i \in [n]$, then, by $\mathcal{BR}_0$, we have

$$\max_t M_t^0(x_t) \geq M_i^0(x_i) = \overline{M}_k^0 > 0,$$

and thus $y_0 = 0$. □

Proposition 6. *In every MSNE* $(\mathbf{x}, \mathbf{y})$*, an attack is always possible:* $y_0 < 1$.

Proof. The proof is by contradiction. Let $(\mathbf{x}, \mathbf{y})$ be an MSNE. Suppose there is no attack: $y_0 = 1$. Then, $\sum_{i=1}^n y_i = 1 - y_0 = 0$, so that $y_i = 0$ for all $i \in [n]$. Because $y_i = 0$ for some $i \in [n]$, Claim 2 yields $y_0 = 0$, a contradiction. □

Lemma 3. *In every MSNE* $(\mathbf{x}, \mathbf{y})$*, the probability* y_i *of direct attack to defender* i *is no larger than* $\widehat{\Delta}_i < 1$.

Proof. The proof is by contradiction. Suppose there is some MSNE in which $y_i > \widehat{\Delta}_i$ for some $i \in [n]$. Then, $x_i = 1$ and in turn $M_i^0(1) = -C_i^0 < 0$. Because the attacker can always achieve expected payoff 0 by not attacking anyone, the last condition implies $y_i = 0$, a contradiction. □

Claim 3. *Let* $\mathbf{y}$ *be the compact representation of the mixed-strategy of the attacker in some MSNE. If the probability of no attack* $y_0 > 0$*, then the probability of direct attack to defender* i *is equal to the cost-to-conditional expected-loss of defender* i*:* $y_i = \widehat{\Delta}_i$ *for all* $i \in [n]$.

Proof. The proof is by contradiction. By Lemma 3 $y_i \leq \widehat{\Delta}_i$ for all $i \in [n]$. Suppose $y_i < \widehat{\Delta}_i$ for some i. Then, by $\mathcal{BR}_i$, we have $x_i = 0$, and by $\mathcal{BR}_0$, we have $0 \geq \overline{M}_i^0 > 0$, a contradiction. □

Lemma 4. *In every MSNE* $(\mathbf{x}, \mathbf{y})$ *of an IDD game in which the total of cost-to-conditional expected-loss of all defenders is* $\sum_{i=1}^n \widehat{\Delta}_i < 1$*, there may not be an attack:* $y_0 > 0$.

Proof. By Lemma 3, $y_i \leq \widehat{\Delta}_i$ for all $i \in [n]$. Using the last statement, note that

$$1 - y_0 = \sum_{i=1}^n y_i \leq \sum_{i=1}^n \widehat{\Delta}_i < 1,$$

from which the lemma immediately follows. □

As stated in the main text, we partition the class of IDD games into three subclasses, based on whether $\sum_{i=1}^n \widehat{\Delta}_i$ is (1) less than, (2) equal to, or (3) greater than 1. We consider each subclass in turn.

Proposition 7. *The joint mixed-strategy* $(\mathbf{x}, \mathbf{y})$ *is an MSNE of an IDD game in which the total cost-to-conditional expected-loss of all defenders is* $\sum_{i=1}^n \widehat{\Delta}_i < 1$ *if and only if it satisfies the following properties.*

1. *There may not be an attack with probability of no attack equal to one minus the cost-to-conditional expected-loss of all defenders: for all defenders* i $1 > y_0 = 1 - \sum_{i=1}^n \widehat{\Delta}_i > 0$.
2. *Every defender has non-zero chance of being attacked directly, and this probability equals the respective defender's cost-to-conditional expected-loss of defender: for all defenders* $i \in [n]$, $y_i = \widehat{\Delta}_i > 0$.
3. *Every defender invests some but none does fully, and in particular, the probability a defender does* not *invest equals the respective cost-to-loss ratio to the attacker: for all defenders* $i \in [n]$, $0 < x_i = 1 - \eta_i^0 < 1$.

Proof. Suppose the joint mixed-strategy $(\mathbf{x}, \mathbf{y})$ satisfies the properties above. Then, every defender is indifferent (i.e., for all $i \in [n]$, $\mathcal{BR}_i(y_i) = [0, 1]$, because $y_i = \widehat{\Delta}_i$), as is also the attacker (i.e., $\mathcal{BR}_0(\mathbf{x})$ equals the set of *all* probability distributions over $n + 1$ events because $M_i^0(x_i) = 0$ for all $i \in [n]$). Hence, $(\mathbf{x}, y)$ is an MSNE.

Now suppose $(\mathbf{x}, y)$ is an MSNE of the game. By Lemma 4, $y_0 > 0$. Hence, for all $i \in [n]$, we have $y_i = \widehat{\Delta}_i > 0$ by Claim 3. Both of the previous sentences together imply $M_i^0(x_i) = 0$ for all $i \in [n]$, because of $\mathcal{BR}_0$. Simple algebra yields that $x_i = 1 - \eta_i^0$. Finally, because $y_0 + \sum_{i=1}^n y_i = 1$, we have $y_0 = 1 - \sum_{i=1}^n \widehat{\Delta}_i$. □

Proposition 8. *The joint mixed-strategy* $(\mathbf{x}, \mathbf{y})$ *is an MSNE of an IDD game in which* $\sum_{i=1}^n \widehat{\Delta}_i = 1$ *if and only if it satisfies the following properties.*

1. *There is always an attack:* $y_0 = 0$.
2. *Every defender has non-zero chance of being attacked directly, and this probability equals the respective defender's cost-to-conditional expected-loss of defender i: for all defenders* $i \in [n]$, $y_i = \widehat{\Delta}_i > 0$.
3. *No defender invests fully, and the possible investment probabilities are connected by a 1-d line segment in* $\mathbb{R}^n$*:*

$$x_i = 1 - \frac{v + C_i^0}{\overline{L}_i^0} \text{ for all } i \in [n]$$

with $0 \leq v \leq \min_{i \in [n]} \overline{M}_i^0$.

Proof. Suppose the joint mixed-strategy $(\mathbf{x}, \mathbf{y})$ satisfies the properties above. Then, every defender is indifferent: for all $i \in [n]$, $\mathcal{BR}_i(y_i) = [0, 1]$, because $y_i = \widehat{\Delta}_i$. To test $y \in \mathcal{BR}_0(\mathbf{x})$, note $0 \leq (1 - x_i)\overline{L}_i^0 - C_i^0 = M_i^0(x_i) = \max_{t \in [n]} M_t^0(x_t)$ for all $i \in [n]$, and

$$\sum_{i=1}^n y_i M_i^0(x_i) = \sum_{i=1}^n y_i \max_{t \in [n]} M_t^0(x_t) =$$

$$\left(\sum_{i=1}^n y_i \right) \max_{t \in [n]} M_t^0(x_t) = \max_{t \in [n]} M_t^0(x_t).$$

Let the joint mixed-strategy $(\mathbf{x}, \mathbf{y})$ be an MSNE of the game. Let $I \equiv I(y) \equiv \{i \in [n] \mid y_i > 0\}$. Note that $y_k = 0$ for all $k \notin I$. We first prove the following lemma.

Lemma 5. $I = [n]$.

Proof. The proof is by contradiction. Suppose $I \neq [n]$. By Proposition 6, $y_0 < 1 = y_0 + \sum_{i=1}^n y_i$ so that $y_i > 0$ for some $i \in [n]$, and therefore $I \neq \emptyset$. Also, there exists some $k \in [n] - I$, for which $y_k = 0$. By Claim 2, we then have for all $k \notin I$, $x_k = 0$. By $\mathcal{BR}_0$ and Assumption 2, for all $i, t \in I \neq \emptyset$ and $k \notin I$,

$$M_i^0(x_i) = M_t^0(x_t) \geq \overline{M}_k^0.$$

The condition above yields the following upper bound on the mixed strategies of the defenders in $i \in I$, after applying simple algebraic manipulations: for all $i \in I, k \notin I$,

$$x_i \leq 1 - \frac{\overline{M}_k^0 + C_i^0}{\overline{L}_i^0} < 1.$$

By $\mathcal{BR}_i$, this implies that $y_i \leq \widehat{\Delta}_i$ for all $i \in I$. Putting all of the above together, we have

$$1 = \sum_{i=0}^n y_i = \sum_{i=1}^n y_i = \sum_{i \in I} y_i \leq \sum_{i \in I} \widehat{\Delta}_i \leq \sum_{i=1}^n \widehat{\Delta}_i = 1.$$

Now, because $I \neq [n]$ (by the hypothesis assumed to obtain a contradiction), we have $\sum_{k \notin I} \widehat{\Delta}_k > 0$, and

$$\sum_{i \in I} y_i = \sum_{i=1}^{n} \widehat{\Delta}_i = \sum_{i \in I} \widehat{\Delta}_i + \sum_{k \notin I} \widehat{\Delta}_k > \sum_{i \in I} \widehat{\Delta}_i \geq \sum_{i \in I} y_i,$$

a contradiction. □

By the last lemma and $\mathcal{BR}_0$, we have

$$(1 - x_1)\overline{L}_1^0 - C_1 = \cdots = (1 - x_n)\overline{L}_n^0 - C_n \geq 0$$

Let $v \equiv (1 - x_1)\overline{L}_1^0 - C_1$. Then, $1 - x_i = \frac{v + C_i^0}{\overline{L}_i^0} > 0$. If $v > 0$ then $y_0 = 0$. Because $x_i < 1$, we have $y_i \leq \widehat{\Delta}_i$ for all $i \in [n]$. Thus, we have $y_i = \widehat{\Delta}_i$ for all $i \in [n]$ because otherwise if $y_t < \widehat{\Delta}_t$ for some $t \in [n]$, then $1 = y_0 + y_t + \sum_{i=1, i \neq t}^{n} y_i < \sum_{i=1}^{n} \widehat{\Delta}_i = 1$, a contradiction. If, instead, $v = 0$, for all i, we have $x_i = 1 - \eta_i^0 > 0$, which implies $y_i = \widehat{\Delta}_i$. Therefore, $y_0 = 1 - \sum_{i=1}^{n} y_i = 1 - \sum_{i=1}^{n} \widehat{\Delta}_i = 0$. □

Lemma 6. *In every MSNE* $(\mathbf{x}, \mathbf{y})$ *of an IDD game in which* $\sum_{i=1}^{n} \widehat{\Delta}_i > 1$, *the probability of no attack* $y_0 = 0$.

Proof. The proof is by contradiction. Suppose $y_0 > 0$. Then, by Claim 3, we have $y_i = \widehat{\Delta}_i$ for all $i \in [n]$, and $1 = \sum_{i=0}^{n} y_i = \sum_{i=1}^{n} \widehat{\Delta}_i > 1$, a contradiction. □

Proposition 9. *In every MSNE* $(\mathbf{x}, \mathbf{y})$ *of an IDD game, the probability of no attack* $y_0 > 0$ *if and only if the game has the property* $\sum_{i=1}^{n} \widehat{\Delta}_i < 1$.

Proof. The "if" part is Lemma 4. For the "only if" part, the case in which $\sum_{i=1}^{n} \widehat{\Delta}_i = 1$ follows from Proposition 8; the case in which $\sum_{i=1}^{n} \widehat{\Delta}_i > 1$ follows from Lemma 6. □

Proposition 10. *In every MSNE* $(\mathbf{x}, \mathbf{y})$ *of an IDD game in which* $\sum_{i=1}^{n} \widehat{\Delta}_i > 1$, *no defender is fully investing and some defender is not investing at all (i.e.,* $x_i = 0$ *for some* $i \in [n]$*).*

Proof. The proof is by contradiction. Proposition 9 yields $y_0 = 0$. Suppose $x_i = 1$ for some $i \in [n]$. Then, by $\mathcal{BR}_i$, $y_i \geq \widehat{\Delta}_i$, and by $\mathcal{BR}_0$ and the fact that $y_0 = 0$, we have $0 > -C_i^0 = M_i(x_i) \geq 0$, which implies $y_i = 0$, a contradiction.

Now suppose $0 < x_i < 1$ for all $i \in [n]$. Then, by $\mathcal{BR}_i$, we have $y_i = \widehat{\Delta}_i$ for all $i \in [n]$. Thus we have $1 = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \widehat{\Delta}_i > 1$, a contradiction. □

Proposition 11. *The joint mixed-strategy* $(\mathbf{x}, \mathbf{y})$ *is an MSNE of an IDD game in which* $\sum_{i=1}^{n} \widehat{\Delta}_i > 1$ *if and only if it satisfies the following properties.*

1. *There is always an attack:* $y_0 = 0$.
2. *There exists a non-singleton, non-empty subset* $I \subset [n]$, *such that* $\min_{i \in I} \overline{M}_i^0 \geq \max_{k \notin I} \overline{M}_k^0$, *if* $I \neq [n]$, *and the following holds.*

   (a) *No defender outside* I *invests or is attacked directly:* $x_k = 0$ *and* $y_k = 0$ *for all* $k \notin I$.

   (b) *Let* $J \equiv \arg\min_{i \in I} \overline{M}_i^0$. *No defender in* J *invests and the probability of that defender being attacked directly is at most the defender's cost-to-expected-loss ratio: for all* $i \in J$, $x_i = 0$ *and* $0 \leq y_i \leq \widehat{\Delta}_i$; *in addition,* $\sum_{i \in J} y_i = 1 - \sum_{t \in I - J} \widehat{\Delta}_i$.

(c) *Every defender in $I - J$ partially invests and has positive probability of being attacked directly equal to the defender's cost-to-expected-loss ratio: for all $i \in I - J$, $y_i = \widehat{\Delta}_i$ and*

$$0 < x_i = 1 - \frac{\min_{t \in I} \overline{M}_t^0 + C_i^0}{\overline{L}_i^0} < 1.$$

Proof. For the "if" part, we need to show $(\mathbf{x}, \mathbf{y})$ form mutual best-responses. For all $k \notin I$, $x_k = 0 \in \mathcal{BR}_k(y)$ because $y_k = 0 < \widehat{\Delta}_k$. For all $j \in J$, $x_j = 0 \in \mathcal{BR}_j(y)$ because $y_j \leq \widehat{\Delta}_j$. Finally, for all $i \in I - J$, $x_i \in \mathcal{BR}_i(y_i) = [0, 1]$ because $y_i = \widehat{\Delta}_i$. Hence, we have $x_i \in \mathcal{BR}_i(y_i)$ for all $i \in [n]$. For the attacker, let $v \equiv v(I) \equiv \min_{i \in I} \overline{M}_i^0$. We have for all $k \notin I$, $M_k(x_k) = \overline{M}_k^0 \leq \max_{l \notin I} \overline{M}_l^0 \leq \min_{i \in I} \overline{M}_i^0 = v$, where the first equality holds because $x_k = 0$ and the second inequality by the properties of I. We also have for all $j \in J$, $M_j(x_j) = \overline{M}_j^0 = \min_{i \in I} \overline{M}_i^0 = v$, where the first equality holds because $x_j = 0$ and the second follows from the definition of J. Finally, using simple algebra, we also have for all $i \in I - J$,

$$\begin{aligned} M_i(x_i) &= (1 - x_i)\overline{L}_i^0 - C_i^0 \\ &= \left(\frac{\min_{t \in I} \overline{M}_t^0 + C_i^0}{\overline{L}_i^0} \right) \overline{L}_i^0 - C_i^0 \\ &= \min_{t \in I} \overline{M}_t^0 + C_i^0 - C_i^0 = \min_{t \in I} \overline{M}_t^0 = v. \end{aligned}$$

Hence, we have for all $i \in [n]$, $M_i(x_i) \leq v$. The expected payoff of the attacker under the given joint mixed-strategy is

$$\begin{aligned} \sum_{i=1}^{n} y_i M_i(x_i) &= \sum_{j \in J} y_j M_j(x_j) + \sum_{i \in I - J} y_i M_i(x_i) \\ &= \sum_{j \in J} y_j v + \sum_{i \in I - J} y_i v \\ &= v \left(\sum_{j \in J} y_j + \sum_{i \in I - J} y_i \right) \\ &= v \left(\sum_{i=1}^{n} y_i \right) = v \geq M_i(x_i), \end{aligned}$$

for all $i \in [n]$. Hence, we also have $y \in \mathcal{BR}_0(\mathbf{x})$, and the joint mixed-strategy $(\mathbf{x}, \mathbf{y})$ is an MSNE.

We now consider the "only if" part of the proposition. Let $(\mathbf{x}, \mathbf{y})$ be an MSNE and let $I \equiv I(\mathbf{y}) \equiv \{i \in [n] \mid y_i > 0\}$ be the support of the aggressor's mixed strategy. We now show that I is a non-singleton and non-empty subset of $[n]$.

Claim 4. $1 < |I| \leq n$.

Proof. From Proposition 6, we have $I \neq \emptyset$. That I is not a singleton set follows from Lemma 3. □

By Proposition 9, we have $y_0 = 0$. Applying Proposition 10, let $t \in [n]$ be such that $x_t = 0$. Also by Proposition 10, the aggressor achieves a positive expected payoff: $\sum_{i=1}^{n} y_i M_i^0(x_i) = \max_{l=1}^{n} M_l^0(x_l) \geq M_t^0(x_t) = \overline{M}_t^0 > 0$. For all $k \notin I$, because $y_k = 0$, Claim 2 implies $x_k = 0$.

By $\mathcal{BR}_0$, if I is a strict, non-empty and non-singleton subset of $[n]$, we have, for all $i \in I$ and $k \notin I$,

$$\overline{M}_i^0 \geq M_i^0(x_i) = \max_{l \in I} M_l^0(x_l) \geq \overline{M}_k^0 > 0;$$

otherwise, if $I = [n]$, we have, for all $i \in [n]$,

$$M_i^0(x_i) = \max_{l \in [n]} M_l^0(x_l) = M_t^0(x_t) = \overline{M}_t^0 > 0.$$

Let $v \equiv v(I) \equiv \max_{l \in I} M_l^0(x_l)$. Then, the above expressions imply that for all $i \in I$, we have

$$0 < x_i = 1 - \frac{v + C_i^0}{\overline{L}_i^0} < 1.$$

In addition, we have that if I is a strict, non-empty and non-singleton subset of $[n]$, we have,

$$v = \overline{M}_t^0 \geq \min_{i \in I} \overline{M}_i^0 \geq v \geq \max_{k \notin I} \overline{M}_k^0;$$

and if, instead, $I = [n]$, then

$$v = \overline{M}_t^0 = \min_{i \in [n]} \overline{M}_i^0.$$

Hence, we have $v = \min_{i \in I} \overline{M}_i^0$.

Let $J \equiv J(I) \equiv \arg\min_{i \in I} \overline{M}_i^0$. For all $i \in J$, we have $\overline{M}_i^0 = v$, and thus

$$x_i = 1 - \frac{v + C_i^0}{\overline{L}_i^0} = 1 - \frac{\overline{M}_i^0 + C_i^0}{\overline{L}_i^0} = 1 - \frac{\overline{L}_i^0 - C_i^0 + C_i^0}{\overline{L}_i^0} = 0,$$

and by $\mathcal{BR}_i$, we have $0 \leq y_i \leq \widehat{\Delta}_i$.

For all $i \in I - J$, we have $\overline{M}_i^0 > v$, and thus

$$0 = 1 - \frac{\overline{M}_i^0 + C_i^0}{\overline{L}_i^0} < x_i = 1 - \frac{v + C_i^0}{\overline{L}_i^0} < 1,$$

and by $\mathcal{BR}_i$, we have $y_i = \widehat{\Delta}_i$.

Finally, we have $\sum_{i \in J} y_i = 1 - \sum_{i \in I - J} \widehat{\Delta}_i$, because y is a mixed strategy (i.e, a probability distribution). □

Proposition 5 stated in the main text follows by combining Propositions 7, 8 and 11.

Appendix B.8. Proof of Claim 1

To simplify the notation, let $v \equiv \min_{l \in I} \overline{M}_l^0 = \min_{l \in I'} \overline{M}_l^0$, $J' \equiv \arg\min_{l \in I'} \overline{M}_l^0$ and $J \equiv \arg\min_{i \in I} \overline{M}_i^0$. The hypothesis implies that $(\mathbf{x}, \mathbf{y})$ satisfies the following properties.

$$\begin{aligned}
&\text{for all } i \notin I' \text{: } x_i = y_i = 0 \\
&\text{for all } i \in J' \text{: } x_i = 0 \text{ and } 0 \leq y_i \leq \widehat{\Delta}_i \text{; also } \sum_{i \in J'} y_i = 1 - \sum_{i \in I' - J'} \widehat{\Delta}_i \\
&\text{for all } i \in I' - J' \text{: } x_i = 1 - \frac{v + C_i^0}{\overline{L}_i^0} \text{ and } y_i = \widehat{\Delta}_i
\end{aligned}$$

We now show that $(\mathbf{x}, \mathbf{y})$ also satisfies the constraints when using I with the properties stated in the claim. For that, it needs to satisfy the same expressions as above, but with I' and J' replaced by I and J, respectively.

The first condition is satisfied because $I' \subset I$. The second condition is satisfied for all $i \in J - I'$, because $i \notin I'$ satisfies $x_i = 0$ and $0 \leq y_i = 0 \leq \widehat{\Delta}_i$. It is also satisfied for all $i \in J \cap I'$ because $i \in J$

implies $\overline{M}_i^0 = v$ and, because $i \in I'$, $i \in J'$. For the third condition, note that $I - J \subset I' - J'$ because $i \in I - J$ implies the inequality $\overline{M}_i^0 > v = \max_{k \notin I'} \overline{M}_k^0$; hence, the first inequality in the last expression implies $i \notin J'$, while the equality implies $i \in I'$.

Appendix B.9. Proof of Theorem 2

First, we construct a graph structure and set the values of the parameters to define the IDD game based on an NP-complete problem. Next, we show that if $\mathbf{y}$ exists, then the induced sites-game solves the NP-complete problem. Finally, we show that such a $\mathbf{y}$ exists.

We first define the notations that will be used in the proof. In particular, we consider the problem of determining whether there is a PSNE in the sites-games while fixing some strategies of some players. More specifically, we denote the instances with PSNE as

$$\begin{aligned} \text{sites-game} \quad = \quad & \{\ ((n, G, \mathbf{C}, \mathbf{L}, \widehat{\mathbf{p}}, \widehat{\mathbf{Q}}, \boldsymbol{\alpha}), \\ & \mathbf{a}_S \subset \{0,1\}^{|S|}, \mathbf{y} \subseteq [0,1]^{n+1}) : \text{there exists a PSNE in} \\ & \mathbb{G} \text{ with the players in } S \text{ playing according to } \mathbf{a}_S \\ & \text{and the attacker plays } \mathbf{y} \text{ such that } \sum_{i=0}^{n} y_i = 1\ \}. \end{aligned}$$

We will reduce our problem from MONOTONE 1-IN-3-SAT where each clause of the 3-SAT has exactly three variables and consists of (un-negated) variables. We use the term variable(s) by default for un-negated variable(s), unless stated otherwise. The solution to the MONOTONE 1-IN-3-SAT is to find a satisfiable assignment such that exactly one variable is true in each clause. The MONOTONE 1-IN-3-SAT is known to be NP-complete [57]. We denote the instances with satisfiable solutions as

$$\begin{aligned} \text{MONOTONE 1-IN-3-SAT} \quad = \quad & \{\ ((\mathrm{v}_i)_{i\in[m]}, \wedge_{i=1}^{c} \mathrm{F}_i, \mathrm{F}_i = (\vee_{j=1}^{3} \mathrm{v}_{i_j})) : \text{there exists a} \\ & \text{satisfiable assignment with exactly one} \\ & \text{variable true in each clause}\ \}, \end{aligned}$$

where there are m variables, c clauses, and each clause has three (un-negated) variables. A satisfiable assignment is defined to be an assignment of all variables i to zero or one, $\mathrm{v}_i \in \{0,1\}$, such that the boolean formula $\wedge_{i=1}^{c} \mathrm{F}_i$ is true or satisfied (i.e., each clause F_i is true or satisfied and has exactly one variable true).

Below, given an instance of MONOTONE 1-IN-3-SAT

$$\gamma \equiv \left((\mathrm{v}_i)_{i\in[m]}, \wedge_{i=1}^{c} \mathrm{F}_i, \mathrm{F}_i = (\vee_{j=1}^{3} \mathrm{v}_{i_j}) \right),$$

we are going to construct an instance of sites-games with partial assignments

$$\begin{aligned} \beta \equiv \Big(& (n, G, \mathbf{C}, \mathbf{L}, \widehat{\mathbf{p}}, \widehat{\mathbf{Q}}, \boldsymbol{\alpha}), \\ & \mathbf{a}_S \subseteq \{0,1\}^{|S|}, \mathbf{y} \subseteq [0,1]^n \text{ such that } \sum_{i=0}^{n} y_i = 1 \Big) \end{aligned}$$

that correspond to γ.

- There are $n = 2c + m$ players: two players for each clause and a player for each variable. The clause players and the variable players are indexed from 1 to $2c$ and $2c + 1$ to $2c + m$, respectively.
- First, we find $1 > L'' > C'' > 0$ and $1 > \widehat{p}'' > \frac{C''}{L''}$ such that $0 < \frac{C''}{L''\widehat{p}''} < 1$. Next, we find $\widehat{q} \in [0,1]$ such that $0 < \widehat{q} < \min\{\frac{L''\widehat{p}''}{3C''}, 1\}$. For completeness, we find $1 > \alpha'' > 0$. For each variable player $i \in \{2c+1, ..., 2c+m\}$, let $C_i = C''$, $\alpha_i = \alpha''$, $L_i = L'$, $\widehat{p}_i = \widehat{p}''$, and $y_i = \frac{C_i}{L_i \widehat{p}_i}$.

The variable players are indifferent from playing the action "invest" or "not invest."

- Next, using the values of the parameters defined above, we find $0 < C < L < 1, 1 > \widehat{p} > \frac{C}{L} > 0$, $0 < y < \frac{C}{L\widehat{p}}$, and $1 > \alpha > 0$ such that $\frac{3C''\widehat{q}}{L''\widehat{p}''} > \frac{1}{1-\alpha}\left(\frac{C}{L} - y\widehat{p}\right) > \frac{2C''\widehat{q}}{L''\widehat{p}''}$. Indeed, such value is always possible as we can make α and y to be arbitrarily small so that $\frac{1}{1-\alpha}\left(\frac{C}{L} - y\widehat{p}\right) \approx \frac{C}{L}$.

  For each clause player $i \in [c]$ such that $\mathrm{F}_i = (\vee_{j=1}^{3} \mathrm{v}_{i_j})$, $q_{(i_j+2c)i} = \widehat{q}$ for all j. To set the remaining parameters, for each clause player $i \in [c]$, set $C_i = C$, $L_i = L$, $\alpha_i = \alpha$, $p_i = p$, and $y_i = y$.
- Then, using the same values of the parameters defined for the variable players, we find $0 < C' < L' < 1, 1 > \widehat{p}' > \frac{C'}{L'} > 0, 0 < y' < \frac{C'}{L'\widehat{p}'}$, and $1 > \alpha' > 0$ such that $\frac{2C''\widehat{q}}{L''\widehat{p}''} > \frac{1}{1-\alpha'}\left(\frac{C'}{L'} - y'\widehat{p}'\right) > \frac{C''\widehat{q}}{L''\widehat{p}''}$.

  For each clause player $i \in \{c+1, ..., 2c\}$ such that $\mathrm{F}_{i-c} = \left(\vee_{j=1}^{3} \mathrm{v}_{(i-c)_j}\right)$, $q_{((c-i)_j+2c)i} = q$ for all j. To set the remaining parameters, for each clause player $i \in \{c+1, ..., 2c\}$, set $C_i = C'$, $L_i = L'$, $\alpha_i = \alpha'$, $p_i = p'$, and $y_i = y'$.
- Here, we construct a partial action profile for some of the players. In particular, for each clause player $i \in [c]$, $a_i = 0$ and $a_{i+c} = 1$. Thus, we are giving a partial action profile of all clause players. For completeness, let $y_0 = 1 - \sum_{i=1}^{n} y_i$.

Lemma 7. $\gamma \in$ MONOTONE 1-IN-3-SAT $\implies$ $\beta \in$ *sites-game.*

Proof. Given a satisfiable assignment for γ, we show how to construct a PSNE for β. Let $I^{(1)} \equiv \{i \in [m] \ : \ \mathrm{v}_i = 1\}$ be the indices of the variables that are assigned a value of one in the satisfiable assignment. For consistence, we denote as a_i the action of any player $i \in [n]$ and construct a PSNE as follows. For each of the variable player $i \in \{2c+1, ..., 2c+m\}$, $a_i = 1$ if $(i - 2c) \in I^{(1)}$ and $a_i = 0$ otherwise. Together with the partial action profile of the clauses, we will call this constructed pure-strategy profile $\mathbf{a} = (a_1, ..., a_n)$.

To show that $\mathbf{a}$ is a PSNE, we argue that each player is playing its best-response. First, we consider the clause players. Recall that best-response correspondence of a clause player $i \in [c]$ is

$$\mathcal{BR}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) \equiv \begin{cases} \{1\}, & \text{if } \widehat{s}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) > \widehat{\Delta}_i, \\ \{0\}, & \text{if } \widehat{s}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) < \widehat{\Delta}_i, \\ [0,1], & \text{if } \widehat{s}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) = \widehat{\Delta}_i. \end{cases}$$

where $\widehat{\Delta}_i \equiv \frac{C_i}{L_i\widehat{p}_i}$, $\widehat{s}_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{PF}(i)}) \equiv y_i + \frac{1-\alpha_i}{\widehat{p}_i} r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)})$. Notice that, to determine the best-response strategy of player i, without loss of generality, we can also compare the values of $\frac{1}{1-\alpha_i}\left(\frac{C_i}{L_i} - y_i\widehat{p}_i\right)$ and $r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)})$. By our construction, $\mathrm{Pa}(i) = \{i_1, i_2, i_3\}$ (which corresponds to variables $\mathrm{v}_{i_1}, \mathrm{v}_{i_2}, \mathrm{v}_{i_3}$ of clause i) and $r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)}) = \sum_{j \in \mathrm{Pa}(i)} \frac{C''}{L''\widehat{p}}(1 - a_j)\widehat{q}$.

Moreover, by the satisfiable assignment, exactly one variable in $\mathrm{Pa}(i)$ is assigned to a value of 1 which corresponds to exactly one variable player that plays action 1. Therefore, $r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)}) = \frac{2C''\widehat{q}}{L''\widehat{p}}$. By our construction, $\frac{3C''\widehat{q}}{L''\widehat{p}''} > \frac{1}{1-\alpha_i}\left(\frac{C_i}{L_i} - y_i\widehat{p}_i\right) > \frac{2C''\widehat{q}}{L''\widehat{p}''}$. It follows that $r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)}) < \frac{1}{1-\alpha_i}\left(\frac{C_i}{L_i} - y_i\widehat{p}_i\right)$, and the i's best-response is zero. This holds for all clause players $i \in [c]$. On the other hand, for the clause player $i \in \{c+1, ..., 2c\}$, $r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)}) = \frac{2C''\widehat{q}}{L''\widehat{p}}$ as well. By our construction, $\frac{2C''\widehat{q}}{L''\widehat{p}''} > \frac{1}{1-\alpha_i}\left(\frac{C_i}{L_i} - y_i\widehat{p}_i\right) > \frac{C''\widehat{q}}{L''\widehat{p}''}$, it follows that $\frac{1}{1-\alpha_i}\left(\frac{C_i}{L_i} - y_i\widehat{p}_i\right) < r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)})$ and $a_i = 1$ is the best-response.

For each variable player $i \in \{2c+1, ..., 2c+m\}$, i has no parent and i's overall risk is 0. To determine whether i plays the action invest or not invest, we only need to compare the value of $\frac{C_i}{L_i}$ and $y_i\widehat{p}_i$. By construction, $\frac{C_i}{L_i\widehat{p}_i} = y_i$ for all variable players i, we have that the variable players are indifferent between playing 0 or 1. Hence, the pure-strategy profile $\mathbf{a}$ is a PSNE. □

Lemma 8. $\beta \in$ *sites-game* $\implies \gamma \in$ MONOTONE 1-IN-3-SAT.

Proof. Now we show how to construct a satisfiable assignment for γ given a PSNE of β. Let $\mathbf{a} = (a_1, ..., a_n)$ be a PSNE of β. For each variable $i \in [m]$, if $a_{2c+i} = 1$ then $v_i = 1$ and if $a_{2c+i} = 0$ then $v_i = 0$. To show that each clause, say $i \in [c]$, has exactly one variable that is true, we observe the best-response of clause players i and $c+i$ that correspond to clause i. Given the fixed action of $a_i = 0$ and $a_{c+i} = 1$ at a PSNE, it implies that $r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{y}_{\mathrm{Pa}(i)}) < \frac{1}{1-\alpha_i}\left(\frac{C_i}{L_i} - y_i\widehat{p}_i\right)$ and $r_{c+i}(\mathbf{a}_{\mathrm{Pa}(c+i)}, \mathbf{y}_{\mathrm{Pa}(c+i)}) > \frac{1}{1-\alpha_{c+i}}\left(\frac{C_{c+i}}{L_{c+i}} - y_{c+i}\widehat{p}_{c+i}\right)$. Because $\frac{3C''\widehat{q}}{L''\widehat{p}''} > \frac{1}{1-\alpha_i}\left(\frac{C_i}{L_i} - y_i\widehat{p}_i\right) > \frac{2C''\widehat{q}}{L''\widehat{p}''}$, $\frac{2C''\widehat{q}}{L''\widehat{p}''} > \frac{1}{1-\alpha_{c+i}}\left(\frac{C_{c+i}}{L_{c+i}} - y_{c+i}\widehat{p}_{c+i}\right) > \frac{C''\widehat{q}}{L''\widehat{p}''}$, $\mathrm{Pa}(c+i) = \mathrm{Pa}(i)$, $|\mathrm{Pa}(i)| = 3$, and the transfer risks are the same, we have $r_{c+i}(\mathbf{a}_{\mathrm{Pa}(c+i)}, \mathbf{y}_{\mathrm{Pa}(c+i)}) = \frac{2C''\widehat{q}}{L''\widehat{p}''}$. This implies that exactly one of the variables is true. □

It is easy to see that given a (partial) pure-strategy profile, we can verify whether it is a PSNE of a sites-game in polynomial time. This fact, together with Lemma 7 and Lemma 8, yields our hardness result.

Appendix C. Pseudocode for Exact Algorithm to Compute All MSNE in Single-Attack Fully-Transfer-Vulnerable IDD Games

Algorithm A1 provides pseudocode of the exact algorithm described in Section 3.4. It uses several sub-routines provided in Algorithms A2, A3, and A4, corresponding to the different cases of the characterization of all MSNE for this class of games as also stated in Section 3.4.

Algorithm A1: Compute All MSNE of a Fully Transfer-Vulnerable Single-Attack IDD Game.

Input : A Fully Transfer-Vulnerable Single-Attack IDD game
$\mathcal{G} = (n, G = ([n], E), \mathbf{L}, \mathbf{C}, \widehat{\mathbf{p}}, \widehat{\mathbf{Q}}, \mathbf{C}^0)$
Output: The set $\mathcal{NE}$ of All MSNE of $\mathcal{G}$
foreach $i = 1$ **to** n **do**
  $\widehat{\Delta}_i \leftarrow \frac{C_i}{\widehat{p}_i L_i}$
  $\mathrm{Ch}(i) \leftarrow \{j \in [n] \mid (i,j) \in E\}$
  $\overline{L}_i^0 \leftarrow \widehat{p}_i L_i + \sum_{j \in \mathrm{Ch}(i)} \widehat{q}_{ij} L_j$
  $\eta_i^0 \leftarrow C_i^0 / \overline{L}_i^0$
  $\overline{M}_i^0 \leftarrow \overline{L}_i^0 - C_i^0$
end
if $\sum_{i=1}^n \widehat{\Delta}_i < 1$ **then**
  Assign to $\mathcal{NE}$ the output of call to subroutine for this case given in Algorithm A2 with input $n, \boldsymbol{\eta}^0, \widehat{\boldsymbol{\Delta}}$
end
if $\sum_{i=1}^n \widehat{\Delta}_i = 1$ **then**
  Assign to $\mathcal{NE}$ the output of call to subroutine for this case given in Algorithm A3 with input $n, \widehat{\boldsymbol{\Delta}}, \overline{\mathbf{L}}^0, \mathbf{C}^0$
end
if $\sum_{i=1}^n \widehat{\Delta}_i > 1$ **then**
  Assign to $\mathcal{NE}$ the output of call to subroutine for this case given in Algorithm A4 with input $n, \widehat{\boldsymbol{\Delta}}, \overline{\mathbf{L}}^0, \mathbf{C}^0, \overline{\mathbf{M}}^0$
end
return $\mathcal{NE}$

Algorithm A2: Subroutine to Compute the Unique MSNE of a Fully Transfer-Vulnerable Single-Attack IDD Game with $\sum_{i=1}^{n} \widehat{\Delta}_i < 1$.

Input : $n, \widehat{\boldsymbol{\Delta}}, \boldsymbol{\eta}^0$
Output: The Unique MSNE for this Case as the Set $\mathcal{NE}$
$S \leftarrow 0$
foreach $i = 1$ **to** n **do**
  $x_i \leftarrow 1 - \eta_i^0$
  $y_i \leftarrow \widehat{\Delta}_i$
  $S \leftarrow S + y_i$
end
$y_0 \leftarrow 1 - S$
$\mathcal{NE} \leftarrow \{(\mathbf{x}, \mathbf{y})\}$
return $\mathcal{NE}$

Algorithm A3: Subroutine to Compute (a Simple Linear Representation of) All MSNE of a Fully Transfer-Vulnerable Single-Attack IDD Game with $\sum_{i=1}^{n} \widehat{\Delta}_i = 1$.

Input : $n, \widehat{\boldsymbol{\Delta}}, \overline{\mathbf{L}}^0, \mathbf{C}^0$
Output: The set $\mathcal{NE}$ of All MSNE for this Case
foreach $i = 1$ **to** n **do**
  $y_i \leftarrow \widehat{\Delta}_i$
end
$y_0 \leftarrow 0$
$\mathcal{X}_0 \leftarrow \{\mathbf{x} \geq 0 \mid (1 - x_1)\overline{L}_1^0 - C_1^0 = \cdots = (1 - x_n)\overline{L}_n^0 - C_n^0 \geq 0\}$
$\mathcal{NE} \leftarrow \mathcal{X}_0 \times \{\mathbf{y}\}$
return $\mathcal{NE}$

Algorithm A4: Subroutine to Compute (a Simple Simplex Representation of) All MSNE of a Fully Transfer-Vulnerable Single-Attack IDD Game with $\sum_{i=1}^{n} \widehat{\Delta}_i > 1$.

Input : $n, \widehat{\boldsymbol{\Delta}}, \overline{\mathbf{L}}^0, \mathbf{C}^0, \overline{\mathbf{M}}^0$
Output: The Set $\mathcal{NE}$ of All MSNE for this Case
$(\mathbf{Val}, \mathbf{Idx}) \leftarrow \text{sort}(\overline{\mathbf{M}}^0, \text{'descending'})$
$t \leftarrow 0$
$S \leftarrow 0$
while $t = n$ **or** $S \geq 1$ **do**
  $t \leftarrow t + 1$
  $S \leftarrow S + \widehat{\Delta}_{\text{Idx}(t)}$
end
$k \leftarrow t$
$S \leftarrow S - \widehat{\Delta}_{\text{Idx}(t)}$
$v \leftarrow \text{Val}(t)$ **while** $\text{Val}(t) = v$ **and** $t < n$ **do**
  $t \leftarrow t + 1$
end
foreach $i = 1$ **to** $k - 1$ **do**
  $l \leftarrow \text{Idx}(i)$
  $x_l \leftarrow 1 - \frac{v + C_l^0}{\overline{L}_l^0}$
  $y_l \leftarrow \widehat{\Delta}_l$
end
$O \leftarrow \emptyset$
foreach $i = k$ **to** $t - 1$ **do**
  $l \leftarrow \text{Idx}(i)$
  $x_l \leftarrow 0$
  $O \leftarrow O \cup \{l\}$
end
foreach $i = t$ **to** n **do**
  $l \leftarrow \text{Idx}(i)$
  $x_l \leftarrow 0$
  $y_l \leftarrow 0$
end
$\mathcal{Y}_O \leftarrow \{\mathbf{y}_O \mid 0 \leq y_i \leq \widehat{\Delta}_i$, for all $i \in O$, and $\sum_{i \in O} y_i = 1 - S\}$
$\mathcal{NE} \leftarrow \{\mathbf{x}\} \times \{\mathbf{y}_{-O}\} \times \mathcal{Y}_O$
return $\mathcal{NE}$

Appendix D. FPTAS for Computing an ϵ-MSNE in IDD Games with Directed-Tree Graphs

We will start off simple by considering a *directed star (DS)* graph structure. We refer to single-attack IDD games with such structure simply as *DS-IDD games*. We show that there is a *fully polynomial-time approximation scheme (FPTAS)* to compute an ϵ-MSNE in DS-IDD games. Roughly speaking, an FPTAS's running time is some polynomial of the input and $\frac{1}{\epsilon}$ for $1 < \epsilon < 1$ [58]. Then we generalize the result to *directed trees (DT)*.

Appendix D.1. Directed Stars

Let the source node correspond to player n, and the remaining $n-1$ sink nodes correspond to players' $1, \ldots, n-1$. The directed star (DS) is equivalent to a directed tree with a single root at n with $n-1$ leaves and no internal nodes.

Since the domain of the variables (i.e., mixed strategies) is $[0,1]$, a direct optimization method to compute an MSNE would require solving a highly non-linear optimization problem: cubic objective function for the attacker with quartic constraints for the sites. *An alternative is to discretize the continuous space of the x_i's and y_i's.*

Given two integers $\rho_x > 1$ and $\rho_y > 1$, let $\mathcal{X} \equiv \mathcal{X}(\rho_x) \equiv \{0, \tau_x, 2\tau_x, \ldots, (\rho_x - 2)\tau_x, 1\}$ and $\mathcal{Y} \equiv \mathcal{Y}(\rho_y) \equiv \{0, \tau_y, 2\tau_y, \ldots, (\rho_y - 2)\tau_y, 1\}$ be the respective *discretization* of the interval $[0,1]$ for the sites and the attacker, where $\tau_x \equiv \frac{1}{\rho_x - 1}$ and $\tau_y \equiv \frac{1}{\rho_y - 1}$ are the respective *lengths between points in the discretization grid of* $[0,1]$, and ρ_x and ρ_y are the respective *discretization sizes*. The discretization defines the domains of x_i and y_i to be $\mathcal{X}$ and $\mathcal{Y}$, respectively. Moreover, $|\mathcal{X}| = \rho_x$ and $|\mathcal{Y}| = \rho_y$. Of course, there is an extra constraint for the y_i's in $\mathcal{Y}$: $\sum_{i=1}^{n} y_i \leq 1$ for $y \in \mathcal{Y}^n$. We will determine the values of ρ_x and ρ_y to guarantee an ϵ-MSNE later in the section, but for now, assume they are given. A simple brute-force algorithm to compute an ϵ-MSNE is to check all possible discrete combinations and would take $O\left((\rho_x \rho_y)^n\right)$ time to run in the worst case.

Indeed, we can apply the principle of *dynamic programming (DP)* [59] and design an efficient algorithm to compute ϵ-MSNE that is provably an FPTAS. The key idea is to realize that given a strategy (x_n, y_n) of the root n, the leaves' decisions are independent of each other. However, there is a sum less than or equal to one constraint for the attacker (i.e., $\sum_{i=1}^{n} y_i \leq 1$). Indeed, for each possible combination of (x_n, y_n), we can run a DP algorithm (to be presented later in Appendices D.1.1 and D.1.2) based on some ordering of the nodes and obtain a (best) value for each (x_n, y_n). Clearly, the best (x_n^*, y_n^*) that obtains the maximum value among all other (x_n, y_n)'s is the best possible strategy for the attacker. This guarantees that the attacker would not deviate to a different strategy. Moreover, the DP algorithm would produce solutions that ensure each leaf player is best-responding. More formally, we define the following mathematical expressions for the DP algorithm. This will give us an FPTAS for DS-IDD games.

Appendix D.1.1. Upstream Pass: Collection of Conditional ϵ-MSNE Computation

First, we impose an ordering on the leaves, that is, we order the leaves in increasing order. Let $\overline{M}_i(x_i, y_i, x_n, y_n) \equiv M_i(x_i, y_i, x_n, y_n) - x_i C_i - y_i C_i^0$ be the attacker's utility for attacking i. For each leaf $i = 1, \ldots, n-1$, we compute the set of individual conditional tables (in this order),

$$
\begin{aligned}
&\overline{T}_{i,n}(x_n, y_n, v_i, x_i, y_i, v_{i-1}) \equiv \\
&\overline{M}_i(x_i, y_i, x_n, y_n) + \\
&\log\left(\mathbb{1}[v_i = y_i + v_{i-1}]\right) + \\
&\log\left(\mathbb{1}\left[x_i \in \mathcal{BR}_{x_i}^{\epsilon}(y_i, x_n, y_n)\right]\right) + \\
&T_{i-1,n}(x_n, y_n, v_{i-1})
\end{aligned}
$$

$$T_{i,n}(x_n, y_n, v_i) \equiv \max_{(x_i, y_i, v_{i-1})} \overline{T}_{i,n}(x_n, y_n, v_i, x_i, y_i, v_{i-1})$$

$$W_{i,n}(x_n, y_n, v_i) \equiv \arg\max_{(x_i, y_i, v_{i-1})} \overline{T}_{i,n}(x_n, y_n, v_i, x_i, y_i, v_{i-1})$$

where $T_{0,n}(x_n, y_n, s_0) = 0$ for all (x_n, y_n, s_{i_0}). Each $T_{i,n}$ specifies the maximum possible utility an attacker can get by attacking all the leaves up to i given that the attacker will attack the root n with probability y_n, the root n to invest with probability x_n, and the allowable remaining probability of an attack v_i. The first and the second log-terms are to ensure that the overall probability of attack does not exceed the allowable limit and that site player i is playing best-respond strategies, respectively. Computing each "table of sets" T's and W's, given above, all take $O(\rho_x^2 \rho_y^4)$ each. For n, the *root* of the tree, we compute

$$\begin{aligned}\overline{R}_0(s_0, x_n, y_n, s_n) \equiv & \overline{M}_n(x_n, y_n) + \\ & \log\left(\mathbb{1}[s_0 = s_n + y_n]\right) + \\ & \log\left(\mathbb{1}[x_n \in \mathcal{BR}_n^{\epsilon}(y_n)]\right) + \\ & R_n(x_n, y_n, s_n)\end{aligned}$$

$$R_0(s_0) \equiv \max_{(x_n, y_n, s_n)} \overline{R}_0(s_0, x_n, y_n, s_n)$$

$$W_0(s_0) \equiv \arg\max_{(x_n, y_n, s_n)} \overline{R}_0(s_0, x_n, y_n, s_n)$$

Clearly, computing R_0 and W_0 takes $O(\rho_x \rho_y^3)$. As mentioned earlier, for each combination of (x_n, y_n), we are going to compute the best value an attacker can obtain. The computation of R_0 does exactly this.

Appendix D.1.2. Downstream Pass: Assignment Phase

The assignment phase is essentially the backtracking phase in the DP algorithm where we follow the "back pointers" to find the mixed-strategies for the players and the attacker. For the "downstream" or assignment pass, we are going to start with the root and find $s_0^* \in \arg\max_{s_0} R_0(s_0)$. Because of the discretization result of Theorem 4, there always exists an ϵ-MSNE, and thus, there is an s_0^* such that $R_0(s_0^*) < -\infty$. We set the mixed-strategy of the root to be some $(x_n^*, y_n^*, s_n^*) \in W_0(s_0^*)$. Starting from the opposite order of upstream pass (i.e., $n-1, ..., 1$), we set the mixed-strategies of the leaves according to $v_{n-1}^* \leftarrow s_n^*$, and for $i = n-1, \ldots, 1$,

$$(x_i^*, y_i^*, s_i^*, v_{i-1}^*) \in W_i(x_n^*, y_n^*, v_i^*) \,.$$

By construction the resulting $(\mathbf{x}^*, \mathbf{y}^*)$ is an ϵ-MSNE of the DS-IDD game.

The key to show that this DP algorithm produces an ϵ-MSNE for the DS-IDD games is the discretization sizes. The question is, how small can we make ρ_x and ρ_y and still guarantee an ϵ-MSNE in the discretized space? A more general result about sparse discretization for graphical games [40] provides the answer. Below, we formally state the result of Ortiz [40] for graphical games, but tailored to our context.

Theorem 4. *[40] For any single-attack IDD game, with graph $G = \{[n], E\}$, and any $\epsilon > 0$, a (individually-uniform) discretization of size*

$$\rho_i = \Theta\left(\frac{k_i}{\epsilon}\right)$$

for each site $i \in [n]$, and

$$\rho_0 = \Theta\left(\frac{n|E|}{\epsilon}\right)$$

for the attacker, is sufficient to guarantee that for every exact MSNE of the game, its closest mixed-strategy profile in the induced discretized space in terms of ℓ_∞ distance is also an ϵ-MSNE of the game.

In other words, to get an ϵ-MSNE, we need to set the discretization sizes as specified above for each player in the game.

Lemma 9. *Let $\rho_x = \Theta\left(\frac{1}{\epsilon}\right)$ and $\rho_y = \Theta\left(\frac{n^2}{\epsilon}\right)$. There is a DP algorithm that computes an ϵ-MSNE in DS-IDD games in time $O\left(n(\rho_x\rho_y^2)^2\right) = O\left(\frac{n^9}{\epsilon^6}\right)$.*

Proof. From Theorem 4, we need to set the appropriate discretization sizes for the sites and the attacker as given in the condition in order to guarantee the existence of an MSNE in the discretized grid. Also, in the case of directed stars, we have $k_{\max} = 2$ and $|E| = n - 1$. Moreover, the DP algorithm uses at most $O(\rho_x^2\rho_y^4)$ space and takes the same amount of time to run in the worst case. A simple substitution gives us the claimed running times. □

Our next corollary follows from the above lemma and the definition of FPTAS.

Corollary 2. *There is an FPTAS to compute an ϵ-MSNE in single-attack IDD games with a directed star graph over the sites.*

Appendix D.2. Directed Trees

We refer to a single-attack IDD game with a directed tree graph over the sites as a *DT-IDD game*. Let n denote a site/node in the directed tree with a single source (i.e., the root of the tree). Let $(i_1, \ldots, i_{l_n})$ be a sequence ordering the set of children of n, $\text{Ch}(n) \equiv \{i_1, \ldots, i_{l_n}\}$, where $l_n \equiv |\text{Ch}(n)|$. The following conditions provide the expressions corresponding to the "upstream pass" of the DP algorithm. For all n, *except the root* of the directed tree, we (recursively) define

$$R_n(x_n, y_n, s_n) \equiv T_{i_{k_n},n}(x_n, y_n, s_n)$$

such that, for all $j = 1, \ldots, l_n$, we define

$$\begin{aligned} T_{i_j,n}(x_n, y_n, v_{i_j}) \equiv \max_{(x_{i_j}, y_{i_j}, s_{i_j}, v_{i_{j-1}})} & \overline{M}_{i_j}(x_{i_j}, y_{i_j}, x_n, y_n) \\ & + \log\left(\mathbb{1}\left[v_{i_j} = s_{i_j} + y_{i_j} + v_{i_{j-1}}\right]\right) \\ & + \log\left(\mathbb{1}\left[x_{i_j} \in \mathcal{BR}^{\epsilon}_{x_{i_j}}(y_{i_j}, x_n, y_n)\right]\right) \\ & + R_{i_j}(x_{i_j}, y_{i_j}, s_{i_j}) \\ & + T_{i_{j-1},n}(x_n, y_n, v_{i_{j-1}}) \,, \end{aligned}$$

$W_{i_j,n}(x_n, y_n, v_{i_j})$ is the arg max of the same optimization (i.e., the set of "witnesses" containing the values of $(x_{i_j}, y_{i_j}, s_{i_j}, v_{i_j-1})$ that achieve the maximum values of the optimization given each

(x_n, y_n, v_{i_j})), and, to simplify the presentation, we use the boundary conditions (1) $T_{i_0,n}(x_n, y_n, s_{i_0}) = 0$ for all (x_n, y_n, s_{i_0}); and (2) if i_j is a *leaf* of the tree, then $R_{i_j}(x_{i_j}, y_{i_j}, s_{i_j}) = 0$ for all $(x_{i_j}, y_{i_j}, s_{i_j})$. If n is the *root* of the tree, we compute

$$\begin{aligned} R_0(s_0) \equiv \max_{(x_n, y_n, s_n)} & \overline{M}_n(x_n, y_n) \\ & + \log\left(\mathbb{1}[s_0 = s_n + y_n]\right) \\ & + \log\left(\mathbb{1}[x_n \in \mathcal{BR}_n^{\epsilon}(y_n)]\right) \\ & + R_n(x_n, y_n, s_n)\ , \text{ and} \end{aligned}$$

$W_0(s_0)$ is the arg max of the same optimization (i.e., the set of "witnesses" containing the values of (x_n, y_n, s_n) that achieve the maximum values of the optimization given each s_0 in the discretized grid of probability values in $[0, 1]$).

For the "downstream" or assignment pass, first find $s_0^* \in \arg\max_{s_0} R_0(s_0)$. Note that such s_0^* with $R_0(s_0^*) < -\infty$ because of the properties of the discretization (i.e., the existence of ϵ-MSNE in the appropriately sized grid). Set the values of the root node, denoted by n, to some $(x_n^*, y_n^*, s_n^*) \in W_0(s_0^*)$. Then (recursively) set the values of the children of n, in the reversed order in which the DP computes the maximizations: set $v_{i_{l_n}}^* \leftarrow s_n^*$, and for $j = l_n, \ldots, 1$.

$$(x_{i_j}^*, y_{i_j}^*, s_{i_j}^*, v_{i_{j-1}}^*) \in W_{i_j}(x_n^*, y_n^*, v_{i_j}^*)\ .$$

We repeat the same assignment process for all of the nodes in the tree. Recall that there will always be at least one witness value during the assignment phase because of the properties of the discretization and the existence of MSNE. By construction (i.e., the properties of the DP algorithm and the discretization of Theorem 4), the resulting $(\mathbf{x}^*, \mathbf{y}^*)$ is an ϵ-MSNE of the DT-IDD game. Just like for DS-IDD games, the "upstream phase" dominates the worst-case running time of the DP algorithm for DT-IDD games, which is now $O(n\rho_x^2\rho_y^5)$. Our result follows by applying the same analysis of Lemma 9, and noting that, just like for DS-IDD games, $k_{\max} = 2$ and $|E| = n - 1$.

Lemma 10. *Let $\rho_x = \Theta\left(\frac{1}{\epsilon}\right)$ and $\rho_y = \Theta\left(\frac{n^2}{\epsilon}\right)$. There is a DP algorithm that computes an ϵ-MSNE in DT-IDD games in time $O\left(n\rho_x^2\rho_y^5\right) = O\left(\frac{n^{11}}{\epsilon^7}\right)$.*

The proof of Theorem 3, stated in the main body of the paper (Section 4.3), follows from Lemma 10.

Appendix E. Multiple Attackers

In this section we extend the model further by considering the possibility of multiple attackers. While in the non-cooperative setting we considered the attackers *act* independently, some degree of mutual "cooperation" via *mutual interest* is induced because of the attackers' shared objectives, as encoded in their utility functions; for example, several attackers may all derive utility from attacking the same (set of) sites. We discuss other ways to induce cooperation at the end of Appendix E.3, where we briefly mention another solution concept: i.e., *correlated equilibria (CE)* [44,45].

Appendix E.1. Pure Strategies

First we consider the extension of the model in terms of pure strategies.

Let m be the number of attackers. Let $\mathcal{S}_l$ be the set of sites that can be a target of attacker $l \in [m]$. Let $\mathcal{T}_i$ be the set of attackers that can target site $i \in [n]$ and $\mathcal{T}_I = \cup_{i \in I} \mathcal{T}_i$ the set of all attackers that can target any player in $I \subset [n]$. Denote by $\mathbf{b}^l \equiv (b_i^l : i \in \mathcal{S}_l) \in \{0, 1\}^{|\mathcal{S}_l|}$ the action or pure-strategy (i.e., the "vector of attack") of attacker $l \in [m]$. Extending our previous notation, denote by $\mathbf{b}_I^l \equiv (b_i^l : i \in \mathcal{S}_l \cap I)$ the actions of player l with respect to the sites in $I \subset [n]$. (Note that we ignore any player in I that is not in $\mathcal{S}_l$.) Denote by $\mathbf{b}^D \equiv (\mathbf{b}^l : l \in D)$ the joint-actions or pure strategies (i.e., the joint "vector of attack")

of all attackers $l \in D \subset [m]$. For $I \subset [n]$ and $D \subset [m]$, denote by $\mathbf{b}_I^D \equiv (\mathbf{b}_I^l : l \in D)$. For simplicity, denote by $\mathbf{b}_i \equiv \mathbf{b}_i^{\mathcal{T}_i}$ and $\mathbf{b}_I \equiv \mathbf{b}_I^{\mathcal{T}_I}$. [17] Let

$$t_i \equiv t_i(\mathbf{b}_i) \equiv \mathbb{1}\left[\sum_{l \in \mathcal{T}_i} b_i^l > 0\right] = 1 - \prod_{l \in \mathcal{T}_i}(1 - b_i^l)$$

indicate whether i will be attacked.

A simple way to introduce t_i into the cost function for the sites i is to replace b_i in the previous single-attacker formulation with t_i: [18]

$$\begin{aligned} M_i &\equiv M_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) \\ &\equiv \begin{cases} C_i + \alpha_i r_i L_i, & \text{if } a_i = 1 \text{ (player } i \text{ invests)}, \\ t_i \widehat{p}_i L_i + (1 - t_i \widehat{p}_i) r_i L_i, & \text{if } a_i = 0 \text{ (player } i \text{ does not invest)}, \end{cases} \\ &= a_i[C_i + \alpha_i r_i L_i] + (1 - a_i)[t_i \widehat{p}_i L_i + (1 - t_i \widehat{p}_i) r_i L_i] \,, \end{aligned}$$

where

$$r_i \equiv r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) = 1 - s_i$$

and

$$s_i \equiv s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)}) = \prod_{j \in \mathrm{Pa}(i)} e_{ij}(a_j, t_j(\mathbf{b}_j)) \,.$$

Then, similar to the discussion for a single attacker, the utility function of each attacker l, after performing the replacement applied to the sites, becomes

$$U^l \equiv U^l(\mathbf{a}, \mathbf{b}) \equiv \sum_{i \in \mathcal{S}_l} u_i(\mathbf{a}, \mathbf{b}) - b_i^l C_i^l,$$

where

$$u_i(\mathbf{a}, \mathbf{b}) \equiv u_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) \equiv M_i - a_i C_i = w_i L_i$$

and

$$\begin{aligned} w_i \equiv w_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)}) &\equiv a_i \alpha_i r_i + (1 - a_i)(t_i \widehat{p}_i + (1 - t_i \widehat{p}_i) r_i) \\ &= (1 - a_i) \widehat{p}_i t_i s_i + (a_i \alpha_i + (1 - a_i)) r_i. \end{aligned}$$

Appendix E.2. Mixed Strategies

Here, we consider the extension of the interdependent defense model in the context of mixed strategies. Let $P^l \equiv P^l_{\mathbf{B}_{\mathcal{S}_l}}$ be the joint PMF corresponding to the mixed strategy of attacker l. Denote by

---

[17] Note that $\mathbf{b}^l = \mathbf{b}^l_{\mathcal{S}_l}$, thus consistent with the notation. Note also that, when clear from context, singleton sets are denoted without the set bracket.

[18] By defining the cost functions of each player i this way, i.e., based on the definition of the attack function t_i given, we are implicitly subscribing to the "you only die once" principle [38], because even if *multiple* attacks on any site i are successful, the loss L_i induced by the attack is the same as if a *single* attack were successful. Variations of this model that would make L_i depend on the number of successful attacks that are possible, but not pursued here. Also, in the case of multiple attackers, one may consider $\widehat{p}_i$ a function of $\mathbf{b}_i$, and more specifically, the number of attacks on site i.

$P \equiv (P^1, \ldots, P^m)$ the mixed strategies of all the attackers. The distribution of play induced by joint mixed-strategy $(\mathbf{x}, P)$ satisfies

$$\begin{aligned}\mathbf{P}(\mathbf{A}=\mathbf{a},\mathbf{B}=\mathbf{b}) &= \left[\prod_{i=1}^{n}\mathbf{P}(A_i=a_i)\right]\left[\prod_{l=1}^{m}\mathbf{P}(\mathbf{B}^l_{\mathcal{S}_l}=\mathbf{b}^l_{\mathcal{S}_l})\right]\\ &= \left[\prod_{i=1}^{n}x_i^{a_i}(1-x_i)^{1-a_i}\right]\left[\prod_{l=1}^{m}P^l(\mathbf{b}^l_{\mathcal{S}_l})\right] .\end{aligned}$$

The cost of each site i is now a random function of the decisions of site i's attackers as well as i's parents and i's parents' attackers. The expected cost is

$$\begin{aligned}M_i(\mathbf{x}_{\mathrm{PF}(i)}, P_{\mathrm{PF}(i)}) \equiv& \mathbf{E}[M_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)})]\\ =& x_iC_i + (x_i\alpha_i + (1-x_i))\mathbf{E}[r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}i}]L_i + (1-x_i)\mathbf{E}[t_i(\mathbf{b}_i)s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})]\widehat{p}_iL_i ,\end{aligned}$$

where

$$\mathbf{E}[r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})] \equiv 1 - \mathbf{E}[s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})] ,$$

and

$$\begin{aligned}\mathbf{E}[s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})] \equiv& \mathbf{E}\left[\prod_{j\in\mathrm{Pa}(i)} e_{ij}(a_j, t_j(\mathbf{b}_j))\right]\\ =& \mathbf{E}\left[\prod_{j\in\mathrm{Pa}(i)} e_{ij}(x_j, t_j(\mathbf{b}_j))\right]\\ =& \sum_{\mathbf{b}_{\mathrm{Pa}(i)}} P_{\mathrm{Pa}(i)}(\mathbf{b}_{\mathrm{Pa}(i)}) \prod_{j\in\mathrm{Pa}(i)} e_{ij}(x_j, t_j(\mathbf{b}_j))\\ =& \sum_{\mathbf{b}_{\mathrm{Pa}(i)}} \left[\prod_{l\in\mathcal{T}_{\mathrm{Pa}(i)}} P^l_{\mathrm{Pa}(i)}(\mathbf{b}^l_{\mathrm{Pa}(i)})\right] \prod_{\substack{j\in\mathrm{Pa}(i)\\ t_j(\mathbf{b}_j)=1}} e_{ij}(x_j, 1) .\end{aligned}$$

and

$$\mathbf{E}[t_i(\mathbf{b}_i)s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})] \equiv \sum_{\mathbf{b}_{\mathrm{PF}(i)}} \left[\prod_{l\in\mathcal{T}_{\mathrm{PF}(i)}} P^l_{\mathrm{PF}(i)}(\mathbf{b}^l_{\mathrm{PF}(i)})\right] t_i(\mathbf{b}_i) \prod_{\substack{j\in\mathrm{Pa}(i)\\ t_j(\mathbf{b}_j)=1}} e_{ij}(x_j, 1)$$

Similarly, for the attackers, the expected payoff of each attacker l becomes

$$\begin{aligned}U^l(\mathbf{x}_{\mathcal{S}_l}, \mathbf{y}_{\mathcal{S}_l}) \equiv& \mathbf{E}[U^l(\mathbf{a}, \mathbf{b})]\\ \equiv& \mathbf{E}\left[\sum_{i\in\mathcal{S}_l} w_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)})L_i - b^l_iC^l_i\right]\\ =& \sum_{i\in\mathcal{S}_l} \mathbf{E}[w_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)})]L_i - y^l_iC^l_i ,\end{aligned}$$

where

$$\mathbf{E}[w_i(\mathbf{a}_{\mathrm{PF}(i)}, \mathbf{b}_{\mathrm{PF}(i)})] = (1-x_i)\widehat{p}_i\mathbf{E}[t_i(\mathbf{b}_i)s_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})] + (x_i\alpha_i + (1-x_i))\mathbf{E}[r_i(\mathbf{a}_{\mathrm{Pa}(i)}, \mathbf{b}_{\mathrm{Pa}(i)})].$$

The following result extends Proposition 3 to the case of multiple attackers; and the proof is analogous.

Proposition 12. *For every mixed-strategy $(\mathbf{x}^*, P^*)$ of an IDD game with multiple attackers, and sites' costs and attackers' utilities defined as above, there exists a mixed strategy $(\mathbf{x}^*, \widetilde{P})$, such that for each attacker l, the following holds.*

1. *For all sites $i \in [n]$, the parent-family marginals $\widetilde{P}^l_{\mathrm{PF}(i)} = P^l_{\mathrm{PF}(i)}$ agree, and*
2. *the PMF $\widetilde{P}^l$ decomposes as*

$$\widetilde{P}^l(\mathbf{b}^l_{\mathcal{S}_l}) \propto \prod_{i \in \mathcal{S}_l} \Phi^l_{\mathrm{PF}(i)}(\mathbf{b}^l_{\mathrm{PF}(i)})$$

for some non-negative functions $\Phi^l_{\mathrm{PF}(i)} : \{0,1\}^{k_i} \to [0,\infty)$, and all $\mathbf{b}_{\mathcal{S}_l} \in \{0,1\}^{|\mathcal{S}_l|}$.

In addition, all sites $i \in [n]$ and attackers $l \in [m]$ achieve the same expected cost and utility, respectively, in $(\mathbf{x}^*, P^*)$ as in $(\mathbf{x}^*, \widetilde{P})$:

$$M_i(\mathbf{x}^*, \widetilde{P}) = M_i(\mathbf{x}^*, P^*)$$

and

$$U^l(\mathbf{x}^*, \widetilde{P}) = U^l(\mathbf{x}^*, P^*).$$

Similarly to the case of a single attacker, we have the following as a corollary of the last proposition.

Corollary 3. *For any IDD game with multiple attackers, and for every attacker $l \in [m]$ in the game, let $k_{\max}(l) \equiv \max_{i \in \mathcal{S}_l} k_i$ be the size of the largest parent-family in the game graph that attacker l can target. Then, for every attacker l in the game, the representation size of any of attacker l's mixed strategies is $O\left(2^{k_{\max}(l)}\right)$, modulo expected-payoff equivalence.*

Appendix E.3. Attackers with Limited Mixed Strategies

In this section we consider attackers with possibly limited randomization capabilities for attacks.

One of the simplest restrictions we can impose on the attackers is to assume that every attacker's decision to perform an attack on one of its target sites, is independent of the decision for other sites. Formally, for all attackers l, we would assume

$$P^l(\mathbf{b}^l) = \prod_{i \in \mathcal{S}_l} P^l_i(b^l_i) \equiv \prod_{i \in \mathcal{S}_l} \left(y^l_i\right)^{b^l_i} \left(1 - y^l_i\right)^{1 - b^l_i}. \tag{48}$$

The universal existence of MSNE in such form is still open. Note that Nash's Existence Theorem [60] no longer applies because we are not considering the nice, compact set of all possible probability distributions over each attacker's space of pure strategies, only the ones that are product distributions. *In the special case of a single attacker, given the structure of the attacker's mixed strategy as defined in the last equation (Equation (48)) and compactly represented by* $\mathbf{y}$*, the subgame induced over the sites only is a graphical* $\boldsymbol{\alpha}$*-IDS game* in which each $p_i = y_i \widehat{p}_i$ and $q_{ji} = y_j \widehat{q}_{ji}$ (see Definition 4 and Equation (7)). It is also important to note that, in general, Nash's Existence Theorem does imply that that subgame over the sites only, given the attackers' mixed-strategies, regardless of their structure or restrictions, always has an MSNE. This is because, given the attackers' mixed strategies, the subgame over the sites only is a 2-action finite game in parametric-form.

Another possibility is to restrict the mixed strategy P^l of attacker l to have some compact factored representation over its (factored) pure-strategy space such as those belonging to the class of probabilistic graphical models (i.e., a Markov or Bayesian network whose graph has the sites $\mathcal{T}_l$ as nodes). Unlike the previous restriction given in Equation (48) imposing complete independence (i.e., a product distribution over each site that each attacker could potentially attack), Theorem 12 establishes sufficient conditions for the existence of an MSNE in which the attackers' mixed strategies have such forms.

As a final remark, one could use the solution concept of *correlated equilibria (CE)* [44,45] to introduce potential "cooperation" jointly over both attackers and sites within a non-cooperative setting. Many

possible uses or combinations of CE and MSNE are possible. For instance, we could combine a CE over the attackers with another CE over the sites; or a CE over the attackers and MSNE over the sites; or an MSNE over the attackers and a CE over the sites. Of course, we would still have a problem of how to compactly represent a CE over either the attackers or the sites, or both. To do that, we would have to modify the representation result of Kakade et al. [43] because a naive application of that result would lead to a representation that is exponential over neighborhood sizes, while the IDD representation is only linear over each neighborhood size.

Appendix E.4. Brief Remarks on Computing Equilibria in Multi-Attacker Settings

While extending the *model* itself to handle multi-attackers is relatively straightforward, as we show in this appendix, we believe that extending the technical computational results would require non-trivial amount of work. We leave a comprehensive study of the computational question in multi-attacker IDD models for future work. Having said that, extending the learning-in-games heuristics such as BRGD and SBRD, or the hybrid BRGD-SBRD heuristic we propose in Section 5.2, to this setting is relatively straightforward. But once again, we leave a comprehensive evaluation of such heuristics for future work.

References

1. Bier, V.M.; Azaiez, M.N. *Game Theoretic Risk Analysis of Security Threats*; Springer: New York, NY, USA, 2009.
2. Cárceles-Poveda, E.; Tauman, Y. A Strategic Analysis of the War against Transnational Terrorism. *Games Eco. Behav.* **2011**, *7*, 49–65.
3. Kunreuther, H.; Heal, G. Interdependent Security. *J. Risk Uncertain.* **2003**, *26*, 231–249.
4. Heal, G.; Kunreuther, H. IDS Models of Airline Security. *J. Confl. Resolut.* **2005**, *49*, 201–217.
5. O'Connor, A.; Schmitt, E. Terror Attempt Seen as Man Tries to Ignite Device on Jet. The New York Times, 2009. Avaliable online: http://www.nytimes.com/2009/12/26/us/26plane.html (accessed on 31 August 2010).
6. Gkonis, K.; Psaraftis, H. Container transportation as an interdependent security problem. *J. Transp. Secur.* **2010**, *3*, 197–211.
7. Johnson, B.; Grossklags, J.; Christin, N.; Chuang, J. Uncertainty in Interdependent Security Games. In Proceedings of the First International Conference on Decision and Game Theory for Security, GameSec'10, Berlin, Germany, 22–23 November 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 234–244.
8. Fultz, N.; Grossklags, J. Blue versus Red: Towards a Model of Distributed Security Attacks. In *Financial Cryptography and Data Security*; Dingledine, R., Golle, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 167–183.
9. Roy, S.; Ellis, C.; Shiva, S.; Dasgupta, D.; Shandilya, V.; Wu, Q. A Survey of Game Theory as Applied to Network Security. In Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, HICSS '10, Honolulu, HI, USA, 5–8 January 2010; IEEE Computer Society: Washington, DC, USA, 2010; pp. 1–10.
10. Syverson, P.F.; Systems, A.C. A Different Look at Secure Distributed Computation. In Proceedings of the CSFW-10, Rockport, MA, USA, 10–12 June 1997; IEEE Computer Society Press: Washington, DC, USA, 1997; pp. 109–115.
11. Lye, K.W.; Wing, J. Game Strategies in Network Security. In Proceedings of the Workshop on Foundations of Computer Security, Cape Breton, NS, Canada, 24–26 June 2002; pp. 1–2.
12. Jain, M.; Korzhyky, D.; Vanek, O.; Conitzery, V.; Pechoucek, M.; Tambe, M. A Double Oracle Algorithm for Zero-Sum Security Games on Graphs. In Proceedings of the AAMAS, Taipei, Taiwan, 2–6 May 2011.
13. Kiekintveld, C.; Jain, M.; Tsai, J.; Pita, J.; Ordóñez, F.; Tambe, M. Computing Optimal Randomized Resource Allocations for Massive Security Games. In Proceedings of the AAMAS, Budapest, Hungary, 10–15 May 2009; pp. 689–696.
14. Korzhyk, D.; Conitzer, V.; Parr, R. Complexity of Computing Optimal Stackelberg Strategies in Security Resource Allocation Games. In Proceedings of the AAAI, Atlanta, GA, USA, 11–15 July 2010.

15. Korzhyk, D.; Conitzer, V.; Parr, R. Security Games with Multiple Attacker Resources. In Proceedings of the IJCAI, Catalonia, Spain, 16–22 July 2011; pp. 273–279.
16. Korzhyk, D.; Conitzer, V.; Parr, R. Solving Stackelberg Games with Uncertain Observability. In Proceedings of the AAMAS, Taipei, Taiwan, 2–6 May 2011; pp. 1013–1020.
17. Smith, A.; Vorobeychik, Y.; Letchford, J. MultiDefender Security Games on Networks. *SIGMETRICS Perform. Eval. Rev.* **2014**, *41*, 4–7.
18. Lou, J.; Vorobeychik, Y. Equilibrium Analysis of Multi-Defender Security Games. In Proceedings of the International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
19. Laszka, A.; Lou, J.; Vorobeychik, Y. Multi-Defender Strategic Filtering Against Spear-Phishing Attacks. In Proceedings of the AAAI, Phoenix, AZ, USA, 12–17 February 2016.
20. Liu, P. Incentive-Based Modeling and Inference of Attacker Intent, Objectives, and Strategies. In Proceedings of the 10th ACM Computer and Communications Security Conference (CCS'03), Washington, DC, USA, 27–30 October 2003; pp. 179–189.
21. Cremonini, M.; Nizovtsev, D. Understanding and Influencing Attackers' Decisions: Implications for Security Investment Strategies. In Proceedings of the Fifth Workshop on the Economics of Information Security (WEIS 2006), Cambridge, UK, 26–28 June 2006.
22. Merlevede, J.S.A.; Holvoet, T. Game Theory and Security: Recent History and Future Directions. In *Decision and Game Theory for Security*, Proceedings of the 6th International Conference, GameSec 2015, London, UK, 4–5 November 2015; Khouzani, M., Panaousis, E., Theodorakopoulos, G., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 334–345.
23. Agiwal, S.; Mohtadi, H. Risk Mitigating Strategies in the Food Supply Chain. In Proceedings of the American Agricultural Economics Assocation (Annual Meeting), Orlando, FL, USA, 27–29 July 2008.
24. Dvijotham, K.; Chertkov, M.; Van Hentenryck, P.; Vuffray, M.; Misra, S. Graphical models for optimal power flow. *Constraints* **2016**, 1–26, doi:10.1007/s10601-016-9253-y.
25. Kearns, M.; Ortiz, L.E. Algorithms for Interdependent Security Games. In Proceedings of the Neural Information Processing Systems (NIPS), Whistler, BC, Canada, 11–13 December 2003.
26. Gottlob, G.; Greco, G.; Scarcello, F. Pure Nash equilibria: Hard and easy games. *J. Artif. Intell. Res.* **2005**, *24*, 357–406.
27. Gilboa, I.; Zemel, E. Nash and correlated equilibria: Some complexity considerations. *Games Econ. Behav.* **1989**, *1*, 80–93.
28. Chen, X.; Deng, X.; Teng, S.H. Settling the complexity of computing two-player Nash equilibria. *J. ACM* **2009**, *56*, 1–57.
29. Daskalakis, C.; Goldberg, P.W.; Papadimitriou, C.H. The Complexity of Computing a Nash Equilibrium. *SIAM J. Comput.* **2009**, *39*, 195–259.
30. Papadimitriou, C.H. On the Complexity of the Parity Argument and Other Inefficient Proofs of Existence. *J. Comput. Syst. Sci.* **1994**, *48*, 498–532.
31. Daskalakis, C.; Goldberg, P.W.; Papadimitriou, C.H. The complexity of computing a Nash equilibrium. *Commun. ACM* **2009**, *52*, 89–97.
32. Conitzer, V.; Sandholm, T. New complexity results about Nash equilibria. *Games Econ. Behav.* **2008**, *63*, 621–641.
33. Shavitt, Y.; Shir, E. DIMES—Letting the Internet Measure Itself. http://www.arxiv.org/abs/cs.NI/0506099 (accessed on 25 January 2017).
34. Shavitt, Y.; Shir, E. DIMES: Let the Internet Measure Itself. *ACM SIGCOMM Comput. Commun. Rev.* **2005**, *35*, 71–74.
35. Fudenberg, D.; Levine, D. *The Theory of Learning in Games*; MIT Press: Cambridge, MA, USA, 1999.
36. Kearns, M. Graphical Games. In *Algorithmic Game Theory*; Nisan, N., Roughgarden, T., Éva T., Vaziran, V.V., Eds.; Cambridge University Press: Cambridge, UK, 2007; pp. 159–180.
37. Kearns, M.; Littman, M.; Singh, S. Graphical Models for Game Theory. In Proceedings of the Conference on Uncertainty in Artificial Intelligence, Seattle, WA, USA, 2–5 August 2001; pp. 253–260.
38. Heal, G.; Kunreuther, H. You Only Die Once: Managing Discrete Interdependent Risks. Technical Report W9885, NBER, 2003. Working Paper. Available online: http://ssrn.com/abstract=430599 (accessed on 25 January 2017).
39. Heal, G.; Kunreuther, H. Modeling Interdependent Risks. *Risk Anal.* **2007**, *27*, 621–634.

40. Ortiz, L.E. On Sparse Discretization for Graphical Games. *CoRR* **2014**, abs/1411.3320.
41. Irfan, M.T.; Ortiz, L.E. On influence, stable behavior, and the most influential individuals in networks: A game-theoretic approach. *Artif. Intell.* **2014**, *215*, 79–119.
42. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*; MIT Press: Cambridge, MA, USA, 2009.
43. Kakade, S.; Kearns, M.; Langford, J.; Ortiz, L. Correlated Equilibria in Graphical Games. In Proceedings of the 4th ACM conference on Electronic commerce, EC '03, San Diego, CA, USA, 9–12 June 2003; ACM: New York, NY, USA, 2003; pp. 42–47.
44. Aumann, R. Subjectivity and Correlation in Randomized Strategies. *J. Math. Econ.* **1974**, *1*, 67–96.
45. Aumann, R. Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica* **1987**, *55*, 1–18.
46. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
47. Kearns, M.; Ortiz, L.E. Algorithms for Interdependent Security Games. In *Advances in Neural Information Processing Systems 16*; Thrun, S., Saul, L.K., Schölkopf, B., Eds.; MIT Press: Cambridge, MA, USA, 2004; pp. 561–568.
48. Dyer, M.; Frieze, A.; Kannan, R. A random polynomial time algorithm for approximating the volume of a convex body. *JACM* **1991**, *38*, 1–17.
49. Elkind, E.; Goldberg, L.A.; Goldberg, P.W. Nash Equilibria in Graphical Games on Trees Revisited. In Proceedings of the 7th ACM Conference on Electronic Commerce, EC '06, Ann Arbor, MI, USA, 11–15 June 2006, pp. 100–109.
50. Cai, Y.; Daskalakis, C. On Minmax Theorems for Multiplayer Games. In Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '11, San Francisco, CA, USA, 23–25 January 2011; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2011; pp. 217–234.
51. Singh, S.P.; Kearns, M.J.; Mansour, Y. Nash Convergence of Gradient Dynamics in General-Sum Games. In Proceedings of the UAI, Stanford, CA, 30 June–3 July 2000; pp. 541–548.
52. Kearns, M. Economics, Computer Science, and Policy. *Issues Sci. Technol.* Available online: http://issues.org/21-2/kearns/ (accessed on 25 January 2017).
53. McKelvey, R.D.; Palfrey, T.R. Quantal Response Equilibria for Normal Form Games. *Games Econ. Behav.* **1995**, *10*, 6–38.
54. Kleinberg, J. Cascading Behavior in Networks: Algorithmic and Economic Issues. In *Algorithmic Game Theory*; Nisan, N., Roughgarden, T., Éva Tardos., Vazirani, V.V., Eds.; Cambridge University Press: Cambridge, UK, 2007; pp. 613–632.
55. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley & Sons: New York, NY, USA, 2006.
56. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2006.
57. Garey, M.R.; Johnson, D.S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*; W. H. Freeman & Co.: New York, NY, USA, 1979.
58. Vazirani, V.V. *Approximation Algorithms*; Springer: New York, NY, USA, 2001.
59. Bellman, R.E. *Dynamic Programming*; Dover Publications: Mineola, NY, USA, 2003.
60. Nash, J. Non-cooperative games. *Ann. Math.* **1951**, *54*, 286–295.