Role of Network Topology in Cybersecurity

Richard J. La

Abstract— **We study how an underlying network property affects network security when nodes are rational and have four choices – i) invest in protection, ii) purchase (incomplete coverage) insurance, iii) invest in protection and purchase insurance, or iv) do nothing. More specifically, using a population game model, we examine how the degree distribution of nodes influences their choices at Nash equilibria (NEs) and overall security level. We first show that there exists a degree threshold at NEs so that only the populations with degrees greater than or equal to the threshold invest in protection. Second, as the *weighted* degree distribution of nodes becomes stochastically larger, the risk or threat posed by a neighbor decreases, even though the aforementioned degree threshold tends to rise, hence only nodes with increasingly higher degrees invest in protection, at the same time. Third, we show that the social optimum also possesses similar properties. Finally, we derive an upper bound on the *price of anarchy*, which is an affine function of the *average* degree of nodes. This upper bound is tight in that it is achieved in some scenarios.**

Index Terms— **Cybersecurity, game theory, price of anarchy.**

I. Introduction

Consider a network consisting of nodes that represent, for instance, individual users or organizations. The edges in the network are not necessarily physical edges. Instead, they could be *logical* or *relational* edges. The degree of a node is defined to be the number of neighbors or incident edges it has in the network.[1]

We assume that there are malicious entities, called *attackers*, which launch attacks against the nodes, for example, in hopes of infecting the machines or gaining unauthorized access to information of victims. Not only can nodes suffer damages or losses due to *direct* attacks from the attackers, but the victims of successful direct attacks may also unknowingly help launch *indirect* attacks on their neighbors.

Faced with possible attacks, each node in the network can choose from a set of admissible actions: First, a node may choose to *protect* itself from malicious attacks by investing in security measures. Second, it may purchase *insurance* to transfer (some of) its risk. Third, it may both invest in protection and purchase insurance. Finally, it may take *no action* and assume all of the risk.

We use the following example of email viruses to motivate our problem. When a computer virus infects a user, the virus can scan the user's emails and disk and send the user's personal information to criminals interested in stealing the user's identify (ID). Moreover, the virus can scan the user's contact list and send bogus emails with a link or an attachment to those on the contact list. When a recipient clicks on the link or opens the attachment, it too becomes infected.

In order to reduce the risks or threats posed by such malicious viruses, users can purchase and install an anti-virus software on their devices. This also reduces the risk to those on their contact lists, thereby producing *positive network externalities* or *network effects* [17]. Also, a user may be able to insure against possible ID theft by purchasing a protection plan or ID theft insurance.

We assume that the nodes are rational and interested in minimizing their own costs. However, when the network is large and consists of many nodes, it is challenging to model the details of strategic interactions among all nodes. To skirt this difficulty, we employ a *population game* model [15]. A population game is often used to model the strategic interactions between many players, possibly from different populations. We adopt a well known solution concept, namely *Nash equilibrium* (NE) of the population game, as an approximation to nodes' behavior in practice.

Our goal is to understand a) the structure of NEs and b) the *network effects* and the resulting price of anarchy/stability (POA/POS) [7] as a function of node degree distribution. To the best of our knowledge, our work is the first study to (i) look at how node degree distribution influences network effects and resulting NEs with a large number of nodes and (ii) provide a (tight) bound on the POA/POS as a function of average node degree.

There exist several closely related studies. For instance, Lelarge and Bolot [10] use a model that captures epidemic propagation of viruses/worms and network effects. They show that organizations do not have a strong incentive to invest in security at equilibria. In addition, they derive the POA for large sparse random graphs and demonstrate the presence of free riding similar to that shown in [18]. In a follow-up study [11], they investigate the problem of incentivizing organizations to invest in security through taxation and insurance. They show that, in the absence of moral hazard, insurance can be used to encourage organizations to protect themselves.

In another interesting study [6], Jiang et al. examine a network security game and characterize the POA under *effective investment* and *bad traffic* models and shed some light on how the cost functions and mutual influence of players affect the POA. Then, they show that more efficient equilibria

This work was supported in part by the National Science Foundation under Grant CCF 08-30675 and a grant from National Institute of Standards and Technology.

Author is with the Department of Electrical & Computer Engineering (ECE) and the Institute for Systems Research (ISR) at the University of Maryland, College Park. E-mail: hyongla@umd.edu

[1] We assume that the network is modeled as an *undirected* graph in the paper. If the network was modeled as a *directed* graph instead, the degree distribution of players we are interested in would be that of *in-degrees*.

can be supported in repeated games through cooperation.

We note that none of these studies investigates the effects of node degree distribution on network-level security at resulting equilibria and/or the relation between the average node degree and inefficiency of equilibria, which are the foci of our study.

The rest of the paper is organized as follows. Section II describes the population game model we adopt for our analysis. Section III presents our main results on the properties of NEs and the POA/POS. We conclude in Section IV with a few remarks on future directions.

II. Model and problem formulation

We model the interaction among the nodes as a *noncooperative game*, in which players are the nodes in the network.[2] This is reasonable because, in many cases, it may be difficult for nodes to cooperate with each other and take coordinated countermeasures to attacks. In addition, even if they could coordinate their actions, they would be unlikely to do so in the absence of clear incentives for coordination.

We are interested in scenarios where the number of nodes is large. Unfortunately, modeling detailed microscale interactions among many nodes in a large network and analyzing ensuing games is difficult; the number of possible strategy profiles increases exponentially with the number of players and finding the NEs of noncooperative games is often challenging even with a moderate number of players.

For analytical tractability, we employ a *population game* model. Population games provide a unified framework and tools for studying *strategic interactions* among *a large number of agents* under following assumptions [15]. First, the choice of an individual agent has very little effect on the payoffs of other agents. Second, there are finitely many populations of agents, and each agent is a member of exactly one population. Third, the payoff of each agent depends only on the *distribution* of actions chosen by members of each population. For a detailed discussion of population games, we refer an interested reader to the manuscript by Sandholm [15].

Our population game does not capture the *edge level* interactions between every pair of neighbors assuming a fixed network. Instead, it is a simplification of complicated reality and only attempts to capture the *average* or *mean* behavior of different populations with varying degrees. A key advantage of this model is that it provides a *scalable* model that enables us to study the effects of network properties on *aggregate* behavior of the players and resulting NEs *regardless of* the network size.

A. Population game

We assume that the maximum degree among all players is $D_{\max} < \infty$. For each $d \in \{1, 2, \ldots, D_{\max}\} =: \mathcal{D}$, s_d denotes the *size* or *mass* of population with degree d, and $\mathbf{s} := \big(s_d;\ d \in \mathcal{D}\big)$ tells us the sizes of populations with varying degrees. All players have the same action space $\mathcal{A} := \{I, N, P, B\}$, where I, N, P and B denote *Insurance, No action, Protection* and *Both insurance and protection*, respectively.[3]

In our study, we assume that insurance companies are *passive* players in that they set the insurance premium and coverage level at the beginning and do not adjust them. The reason for this is that we wish to understand how node degrees influence their behavior and resulting equilibria in a *fixed* setting for a comparative study.

Population states and social state – We denote by $\mathbf{x}_d = \big(x_{d,a};\ a \in \mathcal{A}\big)$, where $\sum_{a\in\mathcal{A}} x_{d,a} = s_d$, the *population state* of population d. The elements $x_{d,a}$, $a \in \mathcal{A}$, represent the mass or size of population d which employs action a. For notational ease, we write $x_{d,PB}$ and $x_{d,U}$ for $x_{d,P} + x_{d,B}$ and $x_{d,I}+x_{d,N}$, respectively. Define $\mathbf{x} := \big(\mathbf{x}_d;\ d \in \mathcal{D}\big)$ to be the *social state*. Let $\mathcal{X}_d := \big\{\mathbf{x}_d \in \mathbb{R}_+^4 \mid \sum_{a\in\mathcal{A}} x_{d,a} = s_d\big\}$, where $\mathbb{R}_+ := [0, \infty)$, and $\mathcal{X} := \prod_{d\in\mathcal{D}} \mathcal{X}_d$.

Costs – The cost function of the game is denoted by $\mathbf{C} : \mathcal{X} \to \mathbb{R}^{4D_{\max}}$. For each social state $\mathbf{x} \in \mathcal{X}$, the cost of a player from population d playing action $a \in \mathcal{A}$ is equal to $\mathbf{C}_{d,a}(\mathbf{x})$. In addition to the cost of investing in protection or purchasing insurance, as we will show shortly, our cost function reflects (i) (expected) losses from attacks and (ii) insurance coverage when a player is insured.

As mentioned earlier, we are interested in exploring how network properties affect the preferences of players. To this end, we model two different types of attacks players suffer from – *direct* and *indirect*. While the first type of attacks are not dependent on the network, the latter depends critically on the underlying network, allowing us to capture the desired *network effects* on players' choices.

a) Direct attacks: We assume that attacker(s) launch an attack against each player with probability τ_A, independently of other players.[4] We call this a *direct* attack. When a player experiences a direct attack, its cost depends on whether or not it is *protected*; if the player is protected (resp. unprotected), it is infected with probability p_P^i (resp. p_U^i). When infected, the node suffers an (average) cost of L. Thus, the expected cost from a (direct) attack is $L_P := L \cdot p_P^i$ (resp. $L_U := L \cdot p_U^i$) when it is protected (resp. unprotected). We assume $0 \le p_P^i < p_U^i \le 1$ and denote $L_U - L_P$ by $\Delta L > 0$.

b) Indirect attacks: Besides the direct attacks by the attackers, a player may also experience *indirect* attacks from its neighbors that have sustained successful direct attacks and are infected. We assume that each infected neighbor will launch an indirect attack against each of its neighbors with probability $\beta_I \in [0, 1]$, independently of each other.

The infections from direct attacks spread only to immediate neighbors and do not propagate any further. While this may not be true with some attacks, we introduce this assumption to avoid the difficulty of accurately modeling the propagation of infections in a network and to simplify

[2]We will use the words *nodes* and *players* interchangeably from here on.

[3]There are other studies where the investment in security is restricted to a binary case, e.g., [2], [11]. In addition, many scenarios including the earlier example of email viruses are well approximated by our model.

[4]Our model can be altered to capture the intensity or frequencies of attacks instead, with appropriate changes to cost functions of the players.

our analysis. Moreover, we suspect that, although relaxing this assumption will generally increase the risks and threats seen by nodes from their neighbors, it will not change the qualitative nature of our findings in Section III.

We denote the mapping that yields the degree distribution of populations by $\mathbf{f} : \mathbb{R}_+^{D_{\max}} \to [0,1]^{D_{\max}}$, where

$$f_d(\mathbf{s}) = \frac{s_d}{\sum_{d' \in \mathcal{D}} s_{d'}},\ \mathbf{s} \in \mathbb{R}_+^{D_{\max}} \text{ and } d \in \mathcal{D},$$

is the fraction of total population with degree d. Similarly, define $\mathbf{w} : \mathbb{R}_+^{D_{\max}} \to [0,1]^{D_{\max}}$, where

$$w_d(\mathbf{s}) = \frac{d \cdot s_d}{\sum_{d' \in \mathcal{D}} d' \cdot s_{d'}},\ \mathbf{s} \in \mathbb{R}_+^{D_{\max}} \text{ and } d \in \mathcal{D}. \quad (1)$$

It is clear from the above definition that $\mathbf{w}$ gives us the *weighted* degree distribution of populations, where the weights are the degrees.

It is easy to show that both $\mathbf{f}$ and $\mathbf{w}$ are scale invariant. In other words, $\mathbf{f}(\mathbf{s}) = \mathbf{f}(\phi \cdot \mathbf{s})$ and $\mathbf{w}(\mathbf{s}) = \mathbf{w}(\phi \cdot \mathbf{s})$ for all $\phi > 0$. When there is no confusion, we write $\mathbf{f}$ and $\mathbf{w}$ in place of $\mathbf{f}(\mathbf{s})$ and $\mathbf{w}(\mathbf{s})$, respectively.

Let us explain the role of the mapping $\mathbf{w}$ briefly. Suppose that we fix a social state $\mathbf{x} \in \mathcal{X}$ and choose a player. The probability that a randomly picked neighbor of the player belongs to population $d \in \mathcal{D}$ can be approximated using w_d because it is proportional to the degree d [4].[5] Hence, the probability that the neighbor has degree d and plays action $a \in \mathcal{A}$ is approximately $w_d \cdot x_{d,a}/s_d$.

Based on this observation, we approximate the probability that a player experiences an indirect attack from a neighbor using $\gamma(\mathbf{x}) = \tau_A \cdot e(\mathbf{x})$, where

$$\begin{aligned} e(\mathbf{x}) &= \beta_I \left(\sum_{d \in \mathcal{D}} w_d \left[\frac{x_{d,PB}}{s_d} p_P^i + \frac{x_{d,U}}{s_d} p_U^i \right] \right) \\ &= \beta_I \left(p_U^i - \frac{\Delta p}{d_{\text{avg}}} \sum_{d \in \mathcal{D}} d \cdot x_{d,PB} \right), \end{aligned} \quad (2)$$

where $\Delta p := p_U^i - p_P^i > 0$ and $d_{\text{avg}} := \sum_{d \in \mathcal{D}} d \cdot s_d$ is the average degree of the populations. Note that $\sum_{d \in \mathcal{D}} w_d \left(\frac{x_{d,PB}}{s_d} p_P^i + \frac{x_{d,U}}{s_d} p_U^i \right)$ is the probability that a randomly selected neighbor of a node will be infected if it experiences a direct attack.

We call $e(\mathbf{x})$ the (risk) *exposure* from a neighbor at social state $\mathbf{x}$. It captures the conditional probability that a player faces an indirect attack from a neighbor *given* that the neighbor suffers a direct attack.

We assume that the costs of a player due to multiple successful attacks are additive and that the players are risk neutral.[6] Hence, the expected cost of a player from indirect attacks is proportional to $\gamma(\mathbf{x})$ and its degree. The additivity of costs is reasonable in many scenarios, including the earlier example of email viruses; each time a user is infected or its ID is stolen, the user will need to spend time and incur expenses to deal with the problem.

Based on this assumption, we adopt the following cost function for our population game: For any given social state $\mathbf{x} \in \mathcal{X}$, the cost of a player with degree $d \in \mathcal{D}$ playing $a \in \mathcal{A}$ is given by

$$\begin{aligned} &\mathbf{C}_{d,a}(\mathbf{x}) \\ &= \begin{cases} \tau_A \left(1 + d \cdot e(\mathbf{x})\right) L_P + c_P & \text{if } a = P, \\ \tau_A \left(1 + d \cdot e(\mathbf{x})\right) L_U & \text{if } a = N, \\ \tau_A \left(1 + d \cdot e(\mathbf{x})\right) L_U + c_I - I_U(\mathbf{x}, d) & \text{if } a = I, \\ \tau_A \left(1 + d \cdot e(\mathbf{x})\right) L_P + c_P & \\ \quad + c_I - I_P(\mathbf{x}, d) & \text{if } a = B, \end{cases} \end{aligned} \quad (3)$$

where $c_P > 0$ and $c_I > 0$ denote the cost of protection and insurance premium, respectively, and $I_P : \mathcal{X} \times \mathcal{D} \to \mathbb{R}$ and $I_U : \mathcal{X} \times \mathcal{D} \to \mathbb{R}$ are mappings that determine insurance payout to protected insured players and unprotected insured players, respectively, as a function of social state and degree. Note that $\tau(1 + d \cdot e(\mathbf{x}))$ is the expected number of direct and indirect attacks a node of degree d experiences.

We assume that the insurance payout to an insured player is proportional to its losses over a deductible amount and *approximate* it using[7]

$$\begin{aligned} I_P(\mathbf{x}, d) &= \xi \big(\tau_A (1 + d \cdot e(\mathbf{x})) L_P - ded \big)^+ \text{ and} \\ I_U(\mathbf{x}, d) &= \xi \big(\tau_A (1 + d \cdot e(\mathbf{x})) L_U - ded \big)^+,\ \mathbf{x} \in \mathcal{X}, \end{aligned} \quad (4)$$

where ded is the deductible amount, $\xi \in (0,1]$ is the *coverage level*, i.e., the fraction of total loss over the deductible amount covered by the insurance, and $(z)^+$ denotes $\max(0, z)$. It is clear from (2) - (4) that the cost of a player depends on both its own security level (i.e., protection) and those of other players via the risk exposure $e(\mathbf{x})$.

As mentioned in Section I, we focus on NEs of population games as an approximation to nodes' behavior in practice. A social state $\mathbf{x}^\star$ is an NE if it satisfies the condition that, for all $d \in \mathcal{D}$ and $a \in \mathcal{A}$,

$$x_{d,a}^\star > 0 \text{ implies } \mathbf{C}_{d,a}(\mathbf{x}^\star) = \min_{a' \in \mathcal{A}} \mathbf{C}_{d,a'}(\mathbf{x}^\star). \quad (5)$$

The existence of an NE in a population game is always guaranteed [15, Theorem 2.1.1, p. 24].

There is an important observation we should mention. From (2) - (4), the cost function also has a scale invariance property, i.e., $\mathbf{C}(\mathbf{x}) = \mathbf{C}(\phi \cdot \mathbf{x})$ for all $\phi > 0$. This scale invariance property of the cost function implies the following: Suppose that $\mathcal{NE}^\star$ denotes the set of NEs for a given population size vector $\mathbf{s}^1$. Then, the set of NEs for another population size vector $\mathbf{s}^2 = \phi \cdot \mathbf{s}^1$ for some $\phi > 0$ is given by $\left\{ \phi \cdot \tilde{\mathbf{x}} \mid \tilde{\mathbf{x}} \in \mathcal{NE}^\star \right\}$. This in turn means that the set of NEs scaled by the inverse of the total population size is the same for all population size vectors with the identical

[5] This implicitly assumes that the network is *neutral*. When the network is either assortative or disassortative, this assumption does not hold. However, we leave the effects of network assortativity for future study.

[6] While we assume that the players are risk neutral to facilitate our analysis, risk aversion can be modeled by altering the cost function and similar qualitative findings can be reached at the expense of more cumbersome proofs. We briefly explain this issue in [8].

[7] The exact insurance payout will likely be far more complicated in reality as it will depend on the realized losses of each insured player, which are given by random variables with some distribution.

degree distribution. As a result, it suffices to study the NEs for population size vectors whose sum is equal to one, i.e., $\sum_{d\in\mathcal{D}} s_d = 1$. We will make use of this observation in our analysis in the following section.

III. Main results

Before we state our main results, we first introduce the assumptions we impose throughout this section.

Assumption 1: We assume that the population size vectors are normalized so that the total population size is one.

Note that Assumption 1 implies that the population size vector $\mathbf{s}$ and its degree distribution $\mathbf{f}(\mathbf{s})$ are the same. Hence, a population size vector can also be viewed as its degree distribution.

Assumption 2: The following inequalities are assumed to hold.

a. $L_P < (1-\xi)\ L_U$;
b. $c_P > c_I + ded + \tau_A((1-\xi)L_U - L_P)$; and
c. $c_I \geq \xi(\tau_A\ L_P - ded)$.

Assumption 2-a states that when a player is attacked, its expected out-of-pocket cost from the attack is smaller when it is *protected* than when it is *insured*. This implies that the coverage level is less than 100 percent even when insured.[8]

We believe that these are reasonable assumptions: Without Assumption 2-a, players will opt for insurance over protection (unless the insurance premium is already too high), which in turn will likely drive up the premium charged by a private insurance company so high that the insurance will eventually become unattractive. If $c_P \leq c_I + ded + \tau_A((1-\xi)L_U - L_P)$, we can show that insurance would not be appealing to the players and would never be chosen. Assumption 2-c ensures that the premium should cover at least the expected payout to a protected insured player due to a direct attack.

We first examine the structure of NEs of the population games and the effects of node degree distribution on NEs in Section III-A. Then, we investigate the social optimum and the POA/POS in Sections III-B and III-C, respectively. Due to a space constraint, we omit the proofs of our main results, which can be found in [8].

A. Population games

Theorem 1: Let $\mathbf{s} \in \mathbb{R}_+^{D_{\max}}$ be a population size vector and $\mathbf{x}^\star \in \mathcal{X}$ be a corresponding NE. If $x^\star_{d_1,PB} > 0$ for some $d_1 \in \{1,2,\ldots,D_{\max}-1\} =: \mathcal{D}^-$, then $x^\star_{d,PB} = s_d$ for all $d > d_1$.

We note that Theorem 1 also implies the following: if $x^\star_{d_2,PB} < s_{d_2}$ for some $d_2 \in \{2,\ldots,D_{\max}\} =: \mathcal{D}^+$, then $x^\star_{d,PB} = 0$ for all $d < d_2$.

In practice, the exposure of a node to indirect attacks will depend on many factors, including not only its own degree, but also the degrees and protection levels of its neighbors. Therefore, even the nodes with the same degree may behave differently. However, it is reasonable to expect that the nodes with larger degrees tend to be more vulnerable and susceptible to indirect attacks and, as a result, have a stronger incentive to invest in protecting themselves against (indirect) attacks. Theorem 1 captures this intuition.

The following theorem suggests that, although there may not exist a unique NE, the size of each population $d \in \mathcal{D}$ investing in protection is identical at all NEs.

Theorem 2: Suppose that $\mathbf{x}^1$ and $\mathbf{x}^2$ are two NEs for a population size vector $\mathbf{s}$. Then, $x^1_{d,PB} = x^2_{d,PB}$ for all $d \in \mathcal{D}$.

One can easily construct an example where there are more than one NE. For instance, when the expected cost of playing I and N is the same and is smaller than that of playing either P or B for some population d at an NE, we have uncountably many NEs. This is a consequence of an observation that a purchase of insurance by a player does not affect the costs of other players, hence their (optimal) responses.

Because the populations choosing to protect remain the same at all NEs (when more than one NE exist) and our interests are in understanding the network security at NEs, with a little abuse of notation, we use $\mathbf{N}(\mathbf{s}) = (\mathbf{N}_{d,a}(\mathbf{s});\ d \in \mathcal{D}$ and $a \in \mathcal{A})$ to denote *any* arbitrary NE for a population size vector $\mathbf{s}$, where $\mathbf{N}_{d,a}(\mathbf{s})$ is the size of population d which plays action a at the NE.

Theorems 1 and 2 provide some insight into the properties and structure of NEs: For a given population size vector $\mathbf{s}$, define

$$d^\star(\mathbf{s}) = \min\{d \in \mathcal{D} \mid \mathbf{N}_{d,P}(\mathbf{s}) + \mathbf{N}_{d,B}(\mathbf{s}) > 0\}.$$

When the set on the right-hand side (RHS) is empty, we set $d^\star(\mathbf{s}) = D_{\max}+1$. Theorem 1 tells us that if $d^\star(\mathbf{s}) < D_{\max}$, $\mathbf{N}_{d,P}(\mathbf{s}) + \mathbf{N}_{d,B}(\mathbf{s}) = s_d$ for all $d \in \{d^\star(\mathbf{s})+1,\ldots,D_{\max}\}$. Also, if $d^\star(\mathbf{s}) > 1$, $\mathbf{N}_{d,P}(\mathbf{s}) + \mathbf{N}_{d,B}(\mathbf{s}) = 0$ for all $d \in \{1,\ldots,d^\star(\mathbf{s})-1\}$. Thus, there exists a threshold on degree such that only the populations with degree greater than or equal to the threshold would invest in protection at any NE.

Theorem 3: Let $\mathbf{s}^1$ and $\mathbf{s}^2$ be two population size vectors that satisfy

$$\sum_{\ell=1}^{d} w_\ell(\mathbf{s}^1) \leq \sum_{\ell=1}^{d} w_\ell(\mathbf{s}^2) \quad \text{for all } d \in \mathcal{D}. \tag{6}$$

Then,

1) $d^\star(\mathbf{s}^1) \geq d^\star(\mathbf{s}^2)$, and
2) $e\big(\mathbf{N}(\mathbf{s}^1)\big) \leq e\big(\mathbf{N}(\mathbf{s}^2)\big)$.

The condition (6) means that the *weighted* degree distribution $\mathbf{w}(\mathbf{s}^1)$ is larger than $\mathbf{w}(\mathbf{s}^2)$ in the usual stochastic order [16]. Therefore, Theorem 3 tells us that, as the degree distribution of a (randomly chosen) neighbor becomes stochastically larger, i) only the populations with increasingly larger degrees invest in protection and ii) the overall network security improves in the sense that the risk exposure from a neighbor diminishes.

[8] In addition to deductibles, coinsurance (i.e., $\xi < 1$) is often used to mitigate the issue of *moral hazard* [9] by sharing risk between both the insurer and the insured. Another way to deal with the issue of moral hazard is *premium discrimination* that ties the insurance premium directly with the security measures adopted by a player as suggested in [2], [11]

It turns out that the claims in Theorem 3 do not always hold when we replace the weighted degree distributions in (6) with the degree distributions of two population sizes. In other words, we can find two population size vectors $\tilde{\mathbf{s}}^1$ and $\tilde{\mathbf{s}}^2$ such that

$$\sum_{\ell=1}^{d} \tilde{s}^1_\ell \leq \sum_{\ell=1}^{d} \tilde{s}^2_\ell \ \text{ for all } d \in \mathcal{D},$$

but the claims in Theorem 3 fail to hold.

Lemma 1: Suppose that two population size vectors $\mathbf{s}^1$ and $\mathbf{s}^2$ satisfy

$$\frac{s^2_d}{s^1_d} \geq \frac{s^2_{d+1}}{s^1_{d+1}} \ \text{ for all } d \in \mathcal{D}^-. \tag{7}$$

Then, the condition (6) in Theorem 3 is satisfied.

The finding in Lemma 1 can be applied to several well known families of distributions. For example, consider a family of truncated power law degree distributions $\{\mathbf{s}^\alpha;\ \alpha \in \mathbb{R}_+\}$, where $s^\alpha_d \propto d^{-\alpha}$, $d \in \mathcal{D}$. Suppose that $\alpha_1 \leq \alpha_2$. Then, one can easily show that $\mathbf{s}^{\alpha_1}$ and $\mathbf{s}^{\alpha_2}$ satisfy (6) as follows:

$$\frac{d^{-\alpha_2}}{d^{-\alpha_1}} = d^{-(\alpha_2-\alpha_1)} \geq (d+1)^{-(\alpha_2-\alpha_1)} = \frac{(d+1)^{-\alpha_2}}{(d+1)^{-\alpha_1}}$$

because $\alpha_2 - \alpha_1 \geq 0$. Hence, the monotonicity properties of the degree threshold and the risk exposure shown in Theorem 3 are true for the family of power law degree distributions.

Since the degree threshold $d^\star$ decreases as the weighted degree distribution becomes (stochastically) smaller from the first claim in Theorem 3, one may suspect that the fraction of population choosing to invest in protection may increase as a result. However, our numerical studies suggest that the fraction of protected population in general tends to decrease instead, although no strict monotonicity property holds.

B. Social optimum

In this subsection, we consider a scenario where there is a single social player (SP) that is in charge of making the decisions for all populations. The SP is interested in minimizing the overall social cost given as the sum of (i) losses from attacks and (ii) the cost of protection. One can show that this social cost is also equal to the sum of the costs of all players including those of insurance companies. Thus, for a social state $\mathbf{x} \in \mathcal{X}$, the social cost is given by

$$\begin{aligned} C(\mathbf{x}) &= \sum_{d\in\mathcal{D}} \left(\sum_{a\in\mathcal{A}} x_{d,a} \cdot \mathbf{C}_{d,a}(\mathbf{x}) \right) \\ &\quad + \sum_{d\in\mathcal{D}} \Big(x_{d,I}\big(I_U(\mathbf{x},d) - c_I\big) + x_{d,B}\big(I_P(\mathbf{x},d) - c_I\big) \Big) \\ &= \sum_{d\in\mathcal{D}} \Big(x_{d,PB} \cdot \mathbf{C}_{d,P}(\mathbf{x}) + x_{d,U} \cdot \mathbf{C}_{d,N}(\mathbf{x}) \Big). \end{aligned} \tag{8}$$

It is clear from (8) that the social cost depends only on $x_{d,PB},\ d \in \mathcal{D}$, as $x_{d,U} = s_d - x_{d,PB}$ for all $d \in \mathcal{D}$. For this reason, we can limit the possible atomic actions of SP to $\{P, N\}$ and simplify the admissible action space of SP to $\mathcal{Y} := \prod_{d\in\mathcal{D}}[0, s_d]$. An SP action $\mathbf{y} = \big(y_d;\ d \in \mathcal{D}\big) \in \mathcal{Y}$ specifies the size of each population d that should invest in protection (i.e., y_d) with an understanding that the remaining population $s_d - y_d$ plays N.

Let us define a mapping $\mathbf{X} : \mathcal{Y} \to \mathcal{X}$, where

$$\mathbf{X}_{d,a}(\mathbf{y}) = \begin{cases} y_d & \text{if } a = P, \\ s_d - y_d & \text{if } a = N, \\ 0 & \text{if } a = I \text{ or } B. \end{cases}$$

Fix an SP action $\mathbf{y} \in \mathcal{Y}$. The social cost associated with $\mathbf{y}$ is then given by a mapping $\tilde{C} : \mathcal{Y} \to \mathbb{R}$, where

$$\begin{aligned} \tilde{C}(\mathbf{y}) &= C\big(\mathbf{X}(\mathbf{y})\big) \\ &= \sum_{d\in\mathcal{D}} \Big(y_d \cdot \mathbf{C}_{d,P}\big(\mathbf{X}(\mathbf{y})\big) + (s_d - y_d)\mathbf{C}_{d,N}\big(\mathbf{X}(\mathbf{y})\big) \Big). \end{aligned} \tag{9}$$

The goal of SP is then to solve the following constrained optimization problem:

$$\min_{\mathbf{y}\in\mathcal{Y}} \ \tilde{C}(\mathbf{y}) \tag{10}$$

Lemma 2: The cost function $\tilde{C}$ of SP given by (9) is a convex function of $\mathbf{y}$ over $\mathcal{Y}$.

Unfortunately, the cost function $\tilde{C}$ is not strictly convex and Lemma 2 alone is not enough to guarantee the uniqueness of minimizer. However, as we show below, the minimizer possesses structure similar to that of NEs which can be exploited to establish the uniqueness.

Let $\mathbf{y}^\star$ denote a minimizer of the social cost, i.e.,

$$\mathbf{y}^\star \in \arg\min_{\mathbf{y}\in\mathcal{Y}} \ \tilde{C}(\mathbf{y}).$$

When we wish to make the dependence of $\mathbf{y}^\star$ on the population size vector $\mathbf{s}$ clear, we use $\mathbf{y}^\star(\mathbf{s})$.

The following theorem reveals that any minimizer $\mathbf{y}^\star$ has a degree threshold so that only the populations with degree greater than or equal to the degree threshold should protect at the social optimum.

Let $d^\dagger = \min\{d \in \mathcal{D} \mid y^\star_d > 0\}$. As before, if the set on the RHS is empty, we set $d^\dagger = D_{\max} + 1$.

Theorem 4: If $d^\dagger < D_{\max}$, $y^\star_d = s_d$ for all $d > d^\dagger$.

We can make use of Theorem 4 to prove the uniqueness of the minimizer $\mathbf{y}^\star$.

Theorem 5: There exists a unique solution to the SP-OPT problem.

Define $d^+(\mathbf{s}) = \min\{d \in \mathcal{D} \mid y^\star_d(\mathbf{s}) > 0\}$ with an understanding that $d^+(\mathbf{s}) = D_{\max} + 1$ if the set on the RHS is empty. In addition to the existence of a degree threshold, monotonicity properties similar to those of NEs shown in Theorem 3 hold at the social optimum as well.

Theorem 6: Let $\mathbf{s}^1$ and $\mathbf{s}^2$ be two population size vectors that satisfy

$$\sum_{\ell=1}^{d} w_\ell(\mathbf{s}^1) \leq \sum_{\ell=1}^{d} w_\ell(\mathbf{s}^2) \ \text{ for all } d \in \mathcal{D}. \tag{11}$$

Then,

1) $d^+(\mathbf{s}^1) \geq d^+(\mathbf{s}^2)$, and
2) $e\big(\mathbf{X}(\mathbf{y}^\star(\mathbf{s}^1))\big) \leq e\big(\mathbf{X}(\mathbf{y}^\star(\mathbf{s}^2))\big)$.

The following theorem tells us that the protected population size is never smaller at the social optimum than at an NE under a mild technical condition.

Theorem 7: Fix a population size vector $\mathbf{s}$, and assume that $c_P \leq (\Delta L(c_I + \xi \cdot ded))/(\xi \cdot L_p)$. Let $\mathbf{x}^\star = \mathbf{N}(\mathbf{s})$ and $\mathbf{y}^\star = \mathbf{y}^\star(\mathbf{s})$. Then, $\sum_{d\in\mathcal{D}} x^\star_{d,PB} \leq \sum_{d\in\mathcal{D}} y^\star_d$.

Without the condition $c_P \leq (\Delta L(c_I + \xi \cdot ded))/(\xi \cdot L_p)$, no player will find P more appealing than N or B in *any* scenario no matter what the risk exposure is. In practice, e.g., the example of email viruses, it is likely that some players will invest in protection without purchasing insurance at least in some cases.

We note that Theorem 7 implies that not only the overall social cost is higher at NEs compared to the social optimum, but the exposure and losses due to attacks are also higher at NEs. Thus, the network security degrades as a result of *free riding* by some players as suggested in [10], [11].

C. Price of anarchy

Inefficiency of NEs is well known in many cases, e.g., [5], [13, Chap. 17-21], and can be easily demonstrated using a simple example of the *Prisoner's Dilemma* [14]. Over the last decade or so, there has been much interest in understanding just how inefficient an NE could be compared to the system optimum and quantifying the suboptimality brought on by selfish nature of players [7].

Two popular ways to measure the inefficiency of NE(s) are POA and POS. The POA (resp. POS) is defined to be the *largest* (resp. *smallest*) ratio between the social cost at an NE and the minimum social cost. The POS can be viewed as the minimum price one needs to pay for *stability* among the players so that no player would have an incentive to deviate from its strategy unilaterally.

Recall that, in our population games, all NEs achieve the same social cost by virtue of Theorem 2 and (9) and, hence, POA and POS are identical. We investigate how large the POA can be in our population games and whether or not there exists a tight upper bound on POA. In particular, we are interested in understanding the relation between the POA and the average degree of nodes.

Theorem 8: Let $\mathbf{s}$ be a population size vector and d_{avg} be the average degree of the populations, i.e., $d_{\text{avg}} = \sum_{d\in\mathcal{D}} d \cdot s_d$. Suppose $\Delta L \cdot \tau_A \leq c_P \leq (\Delta L(c_I + \xi \cdot ded))/(\xi \cdot L_p)$. Then,

$$\frac{C\big(\mathbf{N}(\mathbf{s})\big)}{\tilde{C}\big(\mathbf{y}^\star(\mathbf{s})\big)} \leq 1 + d_{\text{avg}} \cdot \beta_I \cdot p_U^i. \tag{12}$$

The upper bound on POA in (12) is tight in the sense that there are cases where the POA is equal to the bound.

The assumption $c_P \geq \Delta L \cdot \tau_A$ in the theorem is reasonable because it merely requires that the cost of protection is at least the difference in the expected losses sustained only from a single *direct* attack, not including any additional expected losses a player may incur from *indirect* attacks. When this inequality does not hold, all players will invest in protection, leading to a degenerate case. Moreover, since the insurance premium should cover the payout, we can expect $c_I > \tau_A \cdot \xi \cdot L_p$, and there exists c_P satisfying the inequalities in Theorem 8.

IV. Conclusions

We investigated the effects of node degree distribution on network security, by making use of a population game model. Our study revealed several interesting properties and structure of Nash equilibria. In addition, it suggests that the risk or threats seen by nodes from their neighbors tend to diminish as the node degree distribution of neighbors becomes stochastically larger.

One interesting future direction we are currently exploring is how the assortativity of underlying network influences the overall network security. Another open question we are interested in pursuing is when insurance improves the resulting network security at equilibria.

References

[1] T. Alpcan and T. Basar, *Network Security: A Decision and Game-Theoretic Approach*, Cambridge University Press, 2010.
[2] J.C. Bolot and M. Lelarge, "A new perspective on Internet security using insurance," Proc. of IEEE INFOCOM, Phoenix (AZ), Apr. 2008.
[3] P. Bolton and M. Dewatripont, *Contract Theory*, The MIT Press, 2004.
[4] D.S. Callaway, M.E.J. Newman, S.H. Strogatz and D.J. Watts, "Network robustness and fragility: percolation and random graphs," *Physical Review Letters*, 85(25):5468-5471, Dec. 2000.
[5] P. Dubey, "Inefficiency of Nash equilibria," *Mathematics of Operations Research*, 11(1):18, 1986.
[6] L. Jiang, V. Anantharam and J. Walrand, "How bad are selfish investments in network security?", *IEEE/ACM Transactions on Networking*, 19(2):549-560, Apr. 2011.
[7] E. Koutsoupias and C.H. Papadimitriou, "Worst-case equilibria," Proc. of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS), pp. 404-413, 1999.
[8] R.J. La, "Role of network topology in cybersecurity," available at http://www.ece.umd.edu/∼hyongla/PAPERS/La-CDC14-TR.pdf.
[9] J.-J. Laffont and D. Martimort, *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press, 2001.
[10] M. Lelarge and J. Bolot, "A local mean field analysis of security investments in networks," Proc. of the 3rd International Workshop on Economics of Networked Systems (NetEcon), pp. 25-30, Seattle (WA), Aug. 2008.
[11] M. Lelarge and J. Bolot, "Economic incentives to increase security in the Internet: the case for insurance," Proc. of IEEE INFOCOM, Rio de Janeiro (Brazil), Apr. 2009.
[12] A. Melnikov, *Risk Analysis in Finance and Insurance*, 2nd ed., CRC Press, 2011.
[13] N. Nisan, T. Roughgarden, É. Tardos, and V.V. Vazirani, *Algorithmic Game Theory*, Cambridge University Press, 2007.
[14] W. Poundstone, *Prisoner's Dilemma*, Anchor, 1993.
[15] W.H. Sandholm *Population Games and Evolutionary Dynamics*, The MIT Press, 2010.
[16] M. Shaked and J.G. Shanthikumar, *Stochastic Orders*, Springer Series in Statistics, Springer, 2007.
[17] C. Shapiro and H.R. Varian, *Information Rules*, Harvard Business School Press, 1999.
[18] H.R. Varian, "System reliability and free riding," *Economics of Information Security*, 12:1-15, 2004.
[19] J.W. Weibull, *Evolutionary Game Theory*, The MIT Press, 1997.