Blue versus Red: Towards a model of distributed security attacks

Neal Fultz* and Jens Grossklags

School of Information, University of California, Berkeley
102 South Hall, 94720 Berkeley, CA
{nfultz,jensg}@ischool.berkeley.edu

Abstract. This paper analyzes the threat of distributed attacks by developing a two-sided multiplayer model of security in which attackers aim to deny service and defenders strategize to secure their assets. Attackers benefit from the successful compromise of target systems, however, may suffer penalties for increased attack activity. Defenders weigh the likelihood of an attack against the cost of security. We model security decision-making in established (e.g., weakest-link, best-shot) and novel games (e.g., weakest target), and allow defense expenditures in protection and self-insurance technologies.
We find that strategic attackers launch attacks only if defenders do not invest in protective measures. Therefore, the threat of protection can be enough to deter an attacker, but as the number of attackers grows, this equilibrium becomes increasingly unstable.

Key words: Game Theory, Economics of Security, Distributed Denial of Service Attacks, Botnets

1 Introduction

If you encounter an aggressive lion, stare him down. But not a leopard; avoid his gaze at all costs. In both cases, back away slowly; don't run. (Bruce Schneier, 2007 [35])

Commercial and public sector entities have always been targets of directed computer security attacks. Often hackers motivated by peer-recognition, ideological beliefs or the intellectual challenge have been the culprits of security challenges [16]. However, more recently, financially motivated attackers and organized crime entered the cyber security arena substantially multiplying the frequency, ubiquitousness and harm caused by attacks [14, 36, 39]. As a result the

* We thank Alvaro Cárdenas, Nicolas Christin, John Chuang, Roger Dingledine, Paul Laskowski, Doug Tygar, and the anonymous reviewers for their helpful comments to an earlier version of this paper. All remaining errors are our own. This work is supported in part by the National Science Foundation under ITR award ANI-0331659 (100x100) and CCF-0424422 (TRUST - Team for Research in Ubiquitous Secure Technology). Jens Grossklags' work is also funded through a University of California MICRO project grant in collaboration with DoCoMo USA Labs.

arms race between security technologists [3] and perpetrators [22] has increased in pace and utilization of resources. To guide security responses, formal models have been developed to better understand the structure of computer security threats and to design robust mechanisms that are increasingly independent of specific attack vectors [27]. Economic models and empirical studies have become an important contributor to this debate [4]. For example, investment models addressed the fact that not all valuables are worth protecting at all costs [15], partially explaining why defenders lack adequate precautions even if security technologies are cheaply available [2, 5].

The focus of this paper is to complement this work by better understanding attacker motives and behaviors when faced with diverse defense patterns and strategies, and degrees of interdependency. In particular, we want to provide a mathematical framework with enough nuanced structure (e.g., different threat models) to enable stronger and more intuitive statements about characteristics of cyber attack equilibria [12]. We add to the literature on game-theoretic models that have often exclusively focused on the strategic aspects of offensive [34] or defensive [18, 23, 26, 38] actions, respectively.[1] We closely follow and expand on this prior work.

Schechter and Smith [34] draw upon the economics of crime literature to construct a model of attackers in the computer security context [7]. They derive the penalties and probabilities of enforcement that will deter an attacker who acts as an utility optimizer evaluating the risks and rewards of committing an offense [10]. We propose an attack utility function that allows offensive players to allocate resources impacting the frequency of attacks but also potential penalities.

The economy of attackers also exhibits strong interdependencies. For example, security researchers have recorded special cases where worms are coded to attack and replace other worms (e.g., the Netsky email virus removed Mydoom infections), or to strengthen or additionally weaken the defenses of a compromised machine to prevent or enable other malicious code to spread (e.g., by downloading patches or to create backdoors, respectively). We are interested to capture several "naturally" occuring practices in our model. As a first step we study how strategic behavior of defenders changes when they have to respond to a single or multiple attackers. We also explore the competitive effects of a larger population of malicious actors that causes increased attention by defenders, or put differently, overutilizes the resource of weakly protected defenders.

The success of an attack depends not only on the attackers' eagerness to bridge a security perimeter but also on the defenders' cost-benefit calculations [15]. In practice, the arsenal of a defender may include several actions to prevent successful compromises and to limit losses that result from a breach. In Grossklags *et al.* [18] we provide a model that allows such a decoupling of invest-

[1] Several research papers explore the optimal strategies of defenders and attackers in graph-theoretic network inoculation games [6, 28, 29]. We explore economic security incentives in different models capturing public goods characteristics and the trade-off between protection and self-insurance.

ments in the context of computer security. On the one hand, the perimeter can be strengthened with a higher self-protection investment (e.g., implementing or updating a firewall). On the other hand, the amount of losses can be reduced by introducing self-insurance technologies and practices (e.g., backup provisions).[2] It follows that an attacker might perceive the effectiveness of a defense posture differently depending on whether it relies on a preventive or recovery strategy.

The prevalence of widely spread, propagated and correlated threats such as distributed denial of service attacks (DDoS), worms and spam has brought attention to interdependencies existing in computer networks. For an attacker this might create strong economies but sometimes also diseconomies of scale. For example, a single breach of a corporate perimeter may allow an attacker to harvest resources from all machines located within its borders. In other scenarios an attacker may have to shut down every single computer or network connection to achieve an attack goal and thereby incur large costs potentially proportional to network size. More generally, there is an interaction between the structure of the defenders' network, the attack goal and threat model. In Grossklags *et al.* [18] we analyze a set of canonical games that capture some of these interdependencies.

We distinguish between tightly and loosely coupled networks. In a tightly coupled network all defenders will face a loss if the condition of a security breach is fulfilled. This may be a suitable description, for example, of a network perimeter breach that causes the spread of malicious code to all machines, but also applies to independently acting defenders that try to preserve a common secret or resist censorship. In a loosely coupled network consequences may differ for network participants. For example, an attacker might be interested to gain control over a limited set of compromised machines ("zombies" or "bots") to organize them into a logical network ("botnet") to execute a DDoS against third parties [25] or to send unsolicited information to and from the bots (i.e., popup advertisements and spam). At other times an attacker might target a specific set of users for specific reasons (e.g., wealthy users aimed for in spearphishing scams). Other users would stay unharmed and are never targeted.

This idea is captured with a group security contribution function in which defenders' protection efforts are linked with each other. To model tightly coupled networks we draw from prior work that proposed the use of public goods theory in the context of computer security [38]. In particular, we focus on three classical models: average/total effort, weakest link, and best shot [21, 38]. To better understand the sometimes asymmetric incentives in loosely coupled interdependent networks we propose the study of the weakest target game [18]. An attacker is only interested in overwhelming the security of the least protected machines in a network while refraining from directly harming others. In this paper we gener-

[2] We also complement work on market insurance for security and privacy. Cyberinsurance can fulfill several critical functions. For example, audit requirements for cyberinsurance can motivate investments in security, and might contribute to a better understanding of the economic value of the protected resources [24]. Several researchers have investigated the impact of correlation of risks and interdependency of agents in networks on the viability of insurance [8, 9, 32].

alize the weakest target game to allow an attacker to aim for control of exactly $k \leq N$ defender machines.

With our work we hope to provide a more complete framework to understand defenders' and attackers' incentives and expected security actions and outcomes for a variety of decision making situations. In the current paper we are able to discuss which defense actions are plausible given a motivated and strategically acting attacker. We can comment on several important facets of computer security warfare, such as when deterrence will be successful, or when defenders prefer to abstain from any protective action. With our modeling work we expect to provide the foundations for experimental and empirical research but we are also interested to evolve the model so that it captures more facets of fully distributed attacks.

The rest of the paper is organized as follows. In Section 2 we further elaborate on the relationship of our work with related research and introduce our game-theoretic models. We present an analysis of the Nash equilibria for one-shot games with simultaneously acting attackers and defenders in Section 3. In Appendix B we discuss extensions including repeated versions of our games and real-word examples that motivate the use of extensive form games. We conclude in Section 4.

2 Model

In prior work we introduced a *security game* as a game-theoretic model that captures essential characteristics of decision making to protect and self-insure resources within a network [18]. In this paper we expand on this modeling approach by including attackers as active and strategic economic actors.

Security games share the following key assumptions: (i) all defenders in the network share a single purely public protection output, (ii) each defender and attacker selects an effort autonomously (so we do not assume a second layer of organizational decision making), (iii) costs are homogeneous and increase linearly, and (iv) all decisions are made simultaneously. These assumptions are commonly made also in models on decision making of partners in military alliances [33]. We add to these main assumptions that defenders are able to self-insure resources at a homogeneous cost with self-insurance being a purely private good. Further, attackers are able to select a number of targets and an associated effort (i.e., attack frequency).

Following Varian's exposition, who also considers the attacker's utility function, we analyze three canonical contribution functions that determine a global protection level [38]. Different from Varian's work however, here network members have a second action available: They can decide to self-insure themselves from harm. The success of insurance decisions is completely independent of protection choices made by the individual and others. Consequently, the games we consider share qualities of private (on the insurance side) and public (on the protection side) goods. We further add to the research literature by studying two additional games with a more complex determination of protection effectiveness.

2.1 Red: Attacker Incentives

Each of $M \in \mathbb{N}$ attackers has three actions at her disposal. First, she may choose whether to engage in any attacks at all. We denote with m ($0 \leq m \leq M$) the number of players who engage in offensive actions. Second, she decides how many defenders k are targeted ($0 \leq k \leq N$). Third, attackers may choose an attack level, p_a ($0 \leq p_a \leq 1$), which contributes to the success of the attack, with $p_a = 1$ representing a perfect attack. Attackers will receive a benefit L for each not sufficiently protected defender they attack. H_e, the group security contribution function of the defenders, has the decisive impact on whether a targeted defender will be compromised. We will discuss the implementations of H_e in the next section.

Additionally, there is a chance that the attacker may be caught and fined F, $F > 0$. The probability of being caught for *each* attack made, p_c, is independent of whether the attack was successful or not. Therefore, the expected utility of attacker i is:

$$Red_i = E(U_i) = \sum_{1}^{k} p_a L(1 - H_e) - (1 - (1 - p_c)^k)F \tag{1}$$

In the current model, we assume that the likelihood of being penalized is moderated by the number of targeted defenders, k, however, independent of the sophistication of attack, p_a. In practice, the likelihood of being fined may depend on both parameters since defenders may more frequently involve law enforcement or react vigilantly if the attack was more severe. On the other hand, those attackers may be more proficient in covering their tracks limiting the effectiveness of enforcement actions. We defer the analysis of different alternatives for the attacker utility to future work.

2.2 Blue: Defender Incentives

Each of $N \in \mathbb{N}$ defenders receives an endowment W. If she is attacked and compromised successfully by at least one attacker, she faces a loss L.[3] Attacks arrive from each attacker with a probability of p_a ($0 \leq p_a \leq 1$). Players have two security actions at their disposition. Player i chooses an insurance level $0 \leq s_i \leq 1$ and a protection level $0 \leq e_i \leq 1$. Finally, $b \geq 0$ and $c \geq 0$ denote the unit cost of protection and insurance, respectively. The generic utility function for a defender has the following structure:

$$Blue_i = E(U_i) = W - (1 - (1 - p_a)^m)L(1 - s_i)(1 - H(e_i, e_{-i})) - be_i - cs_i \ , \tag{2}$$

where following usual game-theoretic notation, e_{-i} denotes the set of protection levels chosen by players other than i. H is a "security contribution" function

[3] We analyze the case where attacker gain and defender loss are identical (if the defender is not self-insured). In practice, we would frequently expect that there is a disparity between the two *subjective* values [2].

that characterizes the effect of e_i on U_i, subject to the protection levels chosen (contributed) by *all* other players. We require that H be defined for all values over $(0,1)^N$. However, we do not place, for now, any further restrictions on the contribution function (e.g., continuity). From Eqn. (2), the magnitude of a loss depends on three factors: i) how many attackers participate in offensive actions (m) and whether an attack takes place (p_a), ii) whether the individual invested in self-insurance ($1 - s_i$), and iii) the magnitude of the joint protection level ($1 - H(e_i, e_{-i})$). Self-insurance always lowers the loss that an individual incurs when compromised by an attack. Protection probabilistically determines whether an attack is successful. Eqn. (2) therefore yields an expected utility.

2.3 Canonical security contribution functions

In [18] we introduced five security games that we will briefly discuss in the following. In selecting and modeling these games we paid attention to comparability of our security games to prior research (e.g., [21, 33, 38]). The first three specifications for H represent important baseline cases recognized in the public goods literature. The attack consequences in these games are tighly coupled; that is, all defenders will face loss L if the level of the security contribution function is not sufficient to block an attack. To allow us to cover also security dilemmas with loosely coupled attack outcomes we developed two additional games [18]. Please refer to our relevant prior work for more detailed interpretations of all games [18, 19].[4] In this paper we generalize all games by including strategic attackers that target k defenders.

Total effort security game (*tightly coupled*): The global protection level of the network depends on the sum of contributions normalized over the number of all participants. That is, we define $H(e_i, e_{-i}) = \frac{1}{N}\sum_i e_i$, so that Eqn. (2) becomes

$$E(U_i) = W - (1 - (1 - p_a)^m)L(1 - s_i)(1 - \frac{1}{N}\sum_k e_k) - be_i - cs_i \ . \quad (3)$$

Weakest-link security game (*tightly coupled*): The overall protection level depends on the minimum contribution offered over all entities. That is, we have $H(e_i, e_{-i}) = \min(e_i, e_{-i})$, and Eqn. (2) takes the form:

$$E(U_i) = W - (1 - (1 - p_a)^m)L(1 - s_i)(1 - \min(e_i, e_{-i})) - be_i - cs_i \ . \quad (4)$$

Best shot security game (*tightly coupled*): In this game, the overall protection level depends on the maximum contribution offered over all entities. Hence, we have $H(e_i, e_{-i}) = \max(e_i, e_{-i})$, so that Eqn. (2) becomes

$$E(U_i) = W - (1 - (1 - p_a)^m)L(1 - s_i)(1 - \max(e_i, e_{-i})) - be_i - cs_i \ . \quad (5)$$

[4] Varian [38] and Hirshleifer [21] discuss also applications outside of the security context such as maintenance of dikes on an island.

***k*-Weakest-target security game without mitigation (*loosely coupled*):** Here, an attacker will *always* be able to compromise the entities with the k lowest protection levels, but will leave other entities unharmed. This game derives from the security game presented in [11]. Formally, we can describe the game as follows:

$$H(e_i, e_{-i}) = \begin{cases} 0 \text{ if } e_i \leq e_{(k)} \\ 1 \text{ otherwise,} \end{cases} \tag{6}$$

which leads to

$$E(U_i) = \begin{cases} W - pL(1 - s_i) - be_i - cs_i \text{ if } e_i \leq e_{(k)}, \\ W - be_i - cs_i \qquad\qquad\qquad \text{otherwise.} \end{cases} \tag{7}$$

An attacker might be interested in such a strategy if the return on attack effort is relatively low, e.g., when distributing spam. It is also relevant to an attacker with limited skills, a case getting more and more frequent with the availability of automated attack toolboxes [37]; or, when the attacker's goal is to commandeer the largest number of machines using the smallest investment possible [14].

***k*-Weakest-target security game with mitigation (*loosely coupled*):** This game is a variation on the above weakest-target game. The difference is that, the probability that each attack on the weakest protected players is successful is now dependent on each target's security level. Here, an attacker is not necessarily assured of success. In fact, if all individuals invest in full protection, not a single machine will be compromised. This variation allows us to capture scenarios where, for instance, an attacker targets a specific vulnerability, for which an easily deployable countermeasure exists. H_e is defined as:

$$H(e_i, e_{-i}) = \begin{cases} 1 - e_i \text{ if } e_i \leq e_{(k)} \\ 1 \qquad\quad \text{otherwise,} \end{cases} \tag{8}$$

so that

$$E(U_i) = \begin{cases} W - pL(1 - s_i)(1 - e_i) - be_i - cs_i \text{ if } e_i \leq e_{(k)}, \\ W - be_i - cs_i \qquad\qquad\qquad\quad \text{otherwise.} \end{cases} \tag{9}$$

3 Nash equilibrium analysis

3.1 One Attacker, One Defender

In the case of one attacker and one defender, H_e simplifies to $H_e = e$. Red can either choose to attack ($k = 1$) or not ($k = 0$). The utility functions simplify to:[5]

$$Blue = W - p_a L(1 - e)(1 - s) - be - cs \tag{10}$$

[5] The analysis for the weakest target game without mitigation would need to be appropriately changed, but the results are of comparable nature.

$$Red = \begin{cases} p_a L(1-e) - p_c F & \text{if Red attacks } (k=1), \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

We observe that if $p_c F > L$ then the attacker has no incentive to be active ($k = 0$ and $p_a = 0$). As a result the defender will not protect or self-insure his resources ($e = 0$ and $s = 0$).

If protection is more expensive than self-insurance ($b \geq c$) then the defender has no incentive to protect. Consequently, the attacker will always be fully active ($k = 1$ and $p_a = 1$ if $p_c F \leq L$). Self-insurance will be purchased as long as the associated cost is lower than the expected loss ($e = 0$ and $s = 1$, if $L > c$ given that $p_a = 1$), otherwise no investments will be undertaken ($e = 0$ and $s = 0$).

If self-insurance is more costly compared to protection ($b < c$) we can define boundary conditions (e^*, p_a^*) so that the defender and attacker are indifferent about remaining passive or not.

$$e^* = 1 - \frac{p_c F}{L} \tag{12}$$

$$p_a^* = \frac{b}{L} \tag{13}$$

These conditions represent an interior solution ($0 \leq (e^*, p_a^*) \leq 1$) as long as the expected fine for the attacker is not larger than the cost of protection ($p_c F \leq b$), and the loss from a security compromise is at least as large as protection costs ($L \geq b$).

If *only* the first condition delivers a non-permissible value (i.e., $p_c F > L$, but $L \geq b$) then there does not exist a *pure* strategy so that the attacker prefers to be active. That is, when choosing a low attack probability she would evade protection efforts by the defender, however, could not gain enough from the attack to pay for the expected fine. A highly virulent attack would immediately motivate the defender to fully protect. We defer the analysis of mixed strategies for this case to future work.

Now consider the case where the second condition (Eq. 13) does not bind ($L < b$), whether or not Eq. 12 holds. Then the defender will remain passive ($e = 0$ and $s = 0$) and he will enable the attacker to successfully compromise the perimeter ($k = 1$ and $p_a = 1$ if $p_c F \leq L$).

Result 1: When protection is not overpriced relative to losses and self-insurance, the cost-benefit ratio between protection and losses serves as an upper bound on Red's attack probability p_a. Therefore, reducing b would lead to less frequent attacks and a higher expected utility for Blue. Increasing L would also serve to reduce the frequency of attack, but would not increases Blue's expectation. If $p_c F > b$ this prevents the attacker from formulating a pure attack strategy and potentially serves as a deterrent.

3.2 One Attacker, N Defenders

Considering Eqs. 1 and 2 then H_e evaluates to the same value for all players in a tightly coupled network. That is, the vector of inputs is being reduced to a single value. In this case, $Red = p_a kL(1 - H_{(e)}) - (1 - (1 - p_c)^k)F$. Over k, Red is convex and negative until some root r given p_a, which represents how many defenders must be attacked to break even. On the one hand, if $r > N$, then it is never profitable for Red to attack, so $k = 0$ would strictly dominate all other k. If, on the other hand, $r \leq N$, then playing $k = N$ strictly dominates all other k. Given the second case we then investigate the canonical contribution functions. Thus, the internal protection equilibria can be found generally to be:

$$H_e = \frac{1 - (1 - (1 - p_c)^N)F}{p_a NL} \tag{14}$$

$$p_a = \frac{b}{L}(H_e - H_{0,\bar{e}})^{-1} \tag{15}$$

Total Effort: In a total effort game, $H_e = \frac{1}{N}\sum_{i=1}^{N} e_i$. Once again, the second derivative test indicates that Blue is monotone over s and i, so the maximal strategies must be the corner cases:

Full Protection If $Nb = \min(p_a L, Nb, c)$, then Blue plays (1,0).

Full Insurance If $c = \min(p_a L, Nb, c)$, then Blue plays (0,1).

Passivity If $p_a L = \min(p_a L, Nb, c)$, then Blue plays (0,0).

As in the 1-on-1 game, if Blue plays passivity or insurance, Red's best reply is full attack (1,N). If Blue plays protection, Red's best reply is either to not attack at all if $Nb/L < P_c F$, or to play $(p_a, k) = (Nb/L\hat{H}_e, N)$ otherwise.

Result 2: In a multiple defender total effort game, the deterrent effect of protection decreases as N increases. In a scenario where protection is possible, Red's optimal probability of attack grows with N.

Weakest Link: In a weakest link game, $H_e = \min(e)$. The second derivative test indicates that insurance is monotone, but protection may have an internal maxima. Assuming protection is not overpriced, let $\hat{e}_0$ be the empirical minimum of all e; if Blue plays $e_i < \hat{e}_0$, then she has not purchased as much protection as possible. If Blue plays $e_i > \hat{e}_0$, then as a result of the contribution function she receives no additional benefit, but pays the extra cost. Therefore, the pure strategies are $(e_i, s_i) \in \{(0,0), (0,1), (\hat{e}_0, 0), (\hat{e}_0, 1)\}$. Once again, buying both protection and insurance is strictly dominated for nonzero b and c.

Protection If $p_a L > b$ and $\hat{e}_0 > \frac{p_a L - c}{p_a L - b}$, then Blue plays $(\hat{e}_0, 0)$.

Full Insurance If $c = \min(p_a L, p_a L(1-\hat{e}_0) - b\hat{e}_0, c)$, then Blue plays (0,1).

Passivity If $p_a L = \min(p_a L, b, c)$, then Blue plays (0,0).

In this case, Red's best reply to a nonprotection strategy is strictly $p_a = 1$. Red's best reply to Blue's protection strategy depends on $\hat{e}_0$ and whether L is less than c. If $L(1-\hat{e}_0) > c$, Red's best reply is $p_a = 1$, which forces Blue to switch to insurance. Otherwise, Red should play $p_a < b/L$, and Blue switches to passivity.

Result 3: In the case that insurance is overpriced relative to the expected losses with protection, Red is bounded by the cost-benefit ratio just as in the one-on-one game. On the other hand, if insurance is cheap but $\hat{e}_0$ is sufficiently small, Red can actually increase his attack probability and force Blue into an insurance strategy. Therefore, knowledge of $\hat{e}_0$ is extremely important to Blue, just as in [18]. As N increases, protection becomes both less likely for Blue because of coordination issues, and less of a deterrent to Red, whose payoff increases with N.

Best Shot: In a best-shot game, $H_e = \max(e)$. As shown in [18], there is no case in a best shot game with homogeneous defenders in which all defenders choose protection. This is easy to show with an indirect proof: If we assume there is a protection equilibrium for nontrivial parameters, then any single Blue player could profitably deviate by free-riding on his teammates. Because of this, Red's best reply is to always play the maximum attack probability, and Blue chooses the cheaper alternative between passivity and self-insurance. Increasing the number of players has no effect on this equilibrium.

Result 4: Due to the difficulty to coordinate on a protection outcome defenders will prefer to shirk on protection effort and are vulnerable to a motivated attacker (where Red's expected penalties are lower than the potential gains).

k-Weakest Target Game without mitigation: Now consider games for loosely coupled contribution functions.

Let $\hat{e}$ = the k-th smallest e chosen by any defender i. Any Blue player choosing $e > \hat{e}$ would switch to $\hat{e} + \eta$, where $\eta \to 0$. In that case every player choosing $e < \hat{e}$ would choose $\hat{e} + 2\eta$, thus destabilizing any pure protection strategy attempts with a non-strategic attacker [18]. In Appendix A we include the detailed derivations for a mixed strategy equilibrium. Below we summarize the results.

We can derive the probability distribution function of self-protection in a mixed Nash equilibrium:

$$f = \frac{f_{e^*}}{(1 + 2(2k - N) f_{e^*}(e - e*))} \tag{16}$$

$$\text{where } f_{e^*} = \frac{b}{p_a L(N-k)\binom{N-1}{j}} \tag{17}$$

This allows us to compute how often strategy $(e, s) = (0, 1)$ is played:

$$q = .5 + (\sum_{j=0}^{k-1} \binom{N-1}{j} - \frac{c}{p_a L} 2^{N-1}) / \binom{N-1}{k-1}(N-k) \tag{18}$$

Result 5: Because all attacks are successful, Red will always play (1,N). A mixed strategy for defenders exists. The defenders strategy is given by Eqs. 16 - 18.

k-Weakest Target Game with Mitigation: A more nuanced version of the above game allows players a degree of individual protection in a loosely coupled scenario. In this case, a pure protection equilibrium is possible so long as protection is less expensive than insurance. Furthermore, to find additional mixed strategies an analysis quite similar to the above can find a probability distribution of strategies for Blue. Please refer to Appendix A for the general approach to derive the results. The probability distribution function f of self-protection in a mixed Nash equilibrium is:

$$f = \frac{\frac{b}{p_a L} - .5^{N-1} \sum_{j=0}^{k-1} \binom{N-1}{j} + \binom{N-1}{k-1}(N-k).5^{N-2} f_{e^*}(e - e^*)}{(1-e)\binom{N-1}{k-1}(N-k).5^{N-2}[1 + 2(2k-N) f_{e^*}(e - e^*)]}$$

$$\text{where } f_{e^*} \approx [\frac{b}{p_a L} - .5^{N-1} \sum_{j=0}^{k-1} \binom{N-1}{j}] / (1 - e^*)\binom{N-1}{k-1}(N-k).5^{N-2}$$

This distribution is asymptotic at $e = 1$, indicating the benefit of mitigation. Interestingly, the probability of insuring given cheap insurance remains the same as for the unmitigated case (see Eq. 18).

From Red's point of view, k is no longer necessarily increasing after it's second root. Increasing k too high will force Blue to protect. In this case, because red is monotone in p_a, he can first maximize this parameter. He will then choose k such that the cumulative binomial distribution of $(k, N, e^*) < b/L$. Blue then backs down into the mixed strategy, leading to a Nash equilibrium. Red's payout is further reduced by the mitigating factor.

Result 6: In the mitigated weakest target game, Red actually attacks fewer targets more frequently compared to the other games, and Blue plays randomly according to a mixed strategy. Furthermore, as N increases, so does the number of targets that Red attacks.

3.3 M Attackers, N Defenders

Now that the various forms of contribution functions have been analyzed, generalizing from one attacker to M is fairly straightforward. Assuming that Blue does not receive an additional loss from being compromised by one attacker or

many, we find for Blue the new probability of being attacked by substitution, and inverting to find the new ceilings for Red based on the total ceiling of the probability of attack p_A derived earlier. In the case of loosely coupled contribution functions, let p_A be the ceiling on the cumulative distribution of the binomial distribution given f_e and N instead. Then $(1-(1-p_a)^m) = p_A$ and rearranging we find $p_a = 1-(1-p_A)^{1/m}$.

This implies that as m increases, each Red will attack proportionally less in every game where Red is bounded by protection states and the contribution is tightly coupled. As m grows even larger, p_a may shrink small enough to the point that the break even point for Red increases past N, deterring all Red from attacking simultaneously. However, if all the Red quit attacking at once, then it becomes profitable for an individual Red to start attacking again, and there is no Nash. In these cases, it appears that as number of attackers gets large, they begin to suffer from coordination problems just like Blue.

Result 7: For tightly coupled games, we can derive the tipping point as m increases at which bounded equilibria degenerate:

$$(1-(1-p_A)^{1/m})NL > p_cF \tag{19}$$

$$m > \frac{ln(1-p_A)}{ln(1-\frac{p_cF}{NL})} \tag{20}$$

This finding could explain the modern development of botnets. A population of autonomous malware eventually grows too large and forces even unmotivated defenders to protect. Botnets, on the other hand, are characterized by command-and-control communication that can allow attackers to throttle attacks and maintain an equilibrium below this upper bound.

4 Conclusions

There are several key findings from this research:

Nash Equilibria Although the boundaries vary, these games all share common classes of Nash Equilibria (see also Appendix C):

- Unbounded Attack: In the case that either the cost of insurance or the maximum loss is strictly less than the cost of protection, Red attacks with full force, and Blue suffers that cost or insures as appropriate.
- Bounded Attack: If protection is cheap, Red may attack less, dropping the expected loss from an attack below the cost of protection. If attacking at this level is still profitable, Red will do so, and Blue will not protect (assuming a preference of passivity over protection).
- Deterrence: If the fine is so high that attacking at a reduced rate is not profitable, Red will not attack at all, and Blue need not protect or insure.

Non-equilibria states There are several states where pure equilibria do not exist. These include when Blue has a preference for protection over passivity in the bounded cases, or when there are sufficient attackers that attack bounding breaks down.

Attackers Including attackers in the game theoretic model has several important implications. Attackers maximize their utility while minimizing the defense's. Expanding to the multiplayer case, there is an asymmetry between attackers and defenders. Because attackers can attack multiple targets, they can attack less and still be profitable. This forces the defenders into an undesirable subgame of protecting when attackers don't attack or not protecting when they do. Taking into account strategic attackers, full protection equilibria become highly unlikely.

Loosely and Tightly Coupled Contribution Functions The attacker's strategy depends on the nature of the contribution function just as much as the defenders'. In the case of a tightly coupled contribution function, attacking all defenders strictly dominates attacking a subset. On the other hand, this is not necessarily true in a loosely coupled game. Instead, it may be more profitable to target fewer defenders, but attack with more intensity.

Deterrence On the other hand, attackers may be deterred from attacking at all if the expected fine outweighs the expected earnings from an attack. This occurs when the second root of the utility function is greater than N. In other words, there are not enough targets to be profitable. However, this does imply that a government could set enforcement levels and fines such that attackers will be deterred.

Asymmetry The fact that Red can attack many targets leads to a highly asymmetrical game where Red has more ability to control the state of the game than Blue.

Attacker Coordination Bounded attacks becomes less likely as the number of attackers increase. If the attackers are not coordinated, eventually the attackers will over-attack, causing the defenders to protect. Compared to a deterrence equilibrium, this is costly for both the defenders and the attackers.

This implies that future attackers will rely on botnets with command and control communications rather than autonomous agents. It also implies that malware will become increasingly benign, so that defenders are not incentivized to protect against it.

A second way that attackers may solve the coordination problem is through the open market. Phishers have already developed a market economy, of which botnets are a slice [1, 13]. Although these botnets have traditionally been used for spam, they are now also rented for DDoS attacks [40]. This kind of marketplace could have several effects: by leasing time on their bots, attackers get additional utility; by going through a market, it becomes harder to track who really launched an attack, decreasing the chance of being caught; it also significantly reduces the barrier to entry for launching a DDoS attack.

Limitations In developing this model, we have made several assumptions. One major one is the homogeneity of the players. In prior work we have shown that relaxing this assumption can have a significant impact on defenders' strategic behavior [19]. Other assumptions include the perfect attack and perfect defense assumptions. In reality, there is often no such thing as either. As Anderson points out [4], there is an asymmetry in finding exploits that favors the attacker, which this model does not address.

The present analysis relies on game theory and, in particular, Nash equilibrium analysis. We plan to expand the analysis to different behavioral assumptions to narrow the gap between formal analysis and empirical observations in the field and the laboratory [17].[6] Notwithstanding, we expect that the result provided in this paper will be of interest to security practitioners and researchers alike.

We have made several large assumptions about the nature of attackers. First, we have assumed that research and reconnaissance costs are negligible.

Furthermore, we have assumed they are not attacking each other. In reality, rival botnets may be more tempting targets than 'civilians,' and botnet hijacking has been observed 'in the wild' [20]. Malware in the wild has been observed removing other malware, which would certainly reduce damages to a defender, and over time could lead to symbiotic relationships. This also brings to mind defensive hacking and vigilante defenders [30]. There are significant economic and ethical questions when defenders can counterattack. If a vigilante defender compromises a botnet, and damages an infected machine, it may be for the greater good / social optima, but there is a personal risk of legal liability. This is further complicated by the fact that computer security has become highly industrialized [31]. Firms providing security services and research may be in the best position to actually implement vigilante hacking. But simply eliminating attackers would reduce the need for their products.

Another key limitation is the assumption of symmetric utility/loss L. In the case of highly divergent subjective utilities, there are two cases: the defense loss is higher, or the offense gain is higher. If the defense's is higher, we would expect deterrence equilibria to be more common; if the offense's is higher, we would expect bounded attack equilibria to be most common. Similarly, it may not be always the case that an attacker will benefit from a security compromise if the defender is self-insured. For example, installing spyware to gather personal information is of reduced utility if the defender has implemented a credit alert or freeze.

Future Research In the future, we would like to more formally explore the different effects arising from tightly coupled and loosely coupled contribution functions. Heterogeneous defenders and attackers both need to be analyzed, as do endgame strategies for finitely repeated games, and cases of subjective utility. Additionally, hybrid attacker models should be developed to cover vigilante defenders.

[6] See, for example, the application of near rationality to different network games [11].

References

1. C. Abad. The economics of phishing: A survey of the operations of the phishing market. *First Monday*, 10(9), 2005.
2. A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 3(1):26–33, January–February 2005.
3. R. Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. Wiley Computer Publishing, New York, NY, 2 edition, 2001.
4. R. Anderson. Why information security is hard - an economic perspective. In *Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC'01)*, New Orleans, LA, Dec. 2001.
5. AOL/NSCA. Online safety study, Dec. 2005. Available at: `http://www.staysafeonline.info/pdf/safety_study_2005.pdf`.
6. J. Aspnes, K. Chang, and A. Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *Journal of Computer and System Sciences*, 72(6):1077–1093, Sept. 2006.
7. G. Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217, March–April 1968.
8. R. Böhme and G. Kataria. Models and measures for correlation in cyber-insurance. In *Proceedings of the Fifth Annual Workshop on Economics and Information Security (WEIS'06)*, Cambridge, UK, June 2006.
9. J. Bolot and M. Lelarge. A new perspective on internet security using insurance. In *Proceedings of the 27th Conference on Computer Communications (INFOCOM'08)*, pages 1948–1956, Phoenix, AZ, Apr. 2008.
10. S. Cameron. The economics of crime deterrence: A survey of theory and evidence. *Kyklos*, 41(2):301–323, 1988.
11. N. Christin, J. Grossklags, and J. Chuang. Near rationality and competitive equilibria in networked systems. In *Proceedings of ACM SIGCOMM'04 Workshop on Practice and Theory of Incentives in Networked Systems (PINS)*, pages 213–219, Portland, OR, Aug. 2004.
12. R. Cornes and T. Sandler. *The theory of externalities, public goods, and club goods*. Cambridge University Press, Cambridge, UK, 1996. Second edition.
13. T. Cymru. The underground economy: Priceless. *;login: The USENIX Magazine*, 31(6), 2006.
14. J. Franklin, V. Paxson, A. Perrig, and S. Savage. An inquiry into the nature and causes of the wealth of internet miscreants. In *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS'07)*, pages 375–388, Alexandria, VA, October/November 2007.
15. L. Gordon and M. Loeb. The economics of information security investment. *ACM Transactions on Information and System Security*, 5(4):438–4572, Nov. 2002.
16. S. Gordon. The generic virus writer. In *Proceedings of the International Virus Bulletin Conference*, pages 121 – 138, Jersey, Channel Islands, 1994.
17. J. Grossklags, N. Christin, and J. Chuang. Predicted and observed behavior in the weakestlink security game. In *Proceedings of the USENIX Workshop on Usability, Privacy and Security (UPSEC08)*, San Francisco, CA, Apr. 2008.
18. J. Grossklags, N. Christin, and J. Chuang. Secure or insure? A game-theoretic analysis of information security games. In *Proceedings of the 2008 World Wide Web Conference (WWW'08)*, pages 209–218, Beijing, China, Apr. 2008.
19. J. Grossklags, N. Christin, and J. Chuang. Security and insurance management in networks with heterogeneous agents. In *Proceedings of the Ninth ACM Conference on Electronic Commerce (EC'08)*, pages 160–169, Chicago, IL, July 2008.

20. K. J. Higgens. *Dark Reading*, April 2007.
21. J. Hirshleifer. From weakest-link to best-shot: the voluntary provision of public goods. *Public Choice*, 41(3):371–386, Jan. 1983.
22. A. Householder, K. Houle, and C. Dougherty. Computer attack trends challenge internet security. *IEEE Computer*, 35(4):5–7, Apr. 2002.
23. L. Jiang, V. Anantharam, and J. Walrand. Efficiency of selfish investments in network security. In *Proceedings of the 2008 Workshop on the Economics of Networks, Systems, and Computation (NetEcon'08)*, pages 31–36, Seattle, WA, Aug. 2008.
24. J. Kesan, R. Majuca, and W. Yurcik. Three economic arguments for cyberinsurance. In A. Chander, L. Gelman, and M. Radin, editors, *Securing Privacy in the Internet Age*, pages 345–366. Stanford University Press, Stanford, CA, 2008.
25. G. Kessler. Defenses against distributed denial of service attacks. 2000.
26. H. Kunreuther and G. Heal. Interdependent security. *Journal of Risk and Uncertainty*, 26(2–3):231–249, Mar. 2003.
27. C. Landwehr. Formal models for computer security. *ACM Computing Surveys*, 13(3):247–278, Sept. 1981.
28. M. Mavronicolas, V. Papadopoulou, A. Philippou, and P. Spirakis. A network game with attackers and a defender. *Algorithmica*, 51(3):315–341, July 2008.
29. T. Moscibroda, S. Schmid, and R. Wattenhofer. When selfish meets evil: Byzantine players in a virus inoculation game. In *Proceedings of the 25th Annual ACM Symposium on Principles of Distributed Computing (PODC'06)*, pages 35–44, Denver, CO, July 2006.
30. R. Naraine. Kraken botnet infiltration triggers ethics debate. *eWeek.com*, May 2008.
31. B. Potter. Dirty secrets of the security industry, 2007. Presented at Defcon XV (Las Vegas 2007).
32. S. Radosavac, J. Kempf, and U. Kozat. Using insurance to increase internet security. In *Proceedings of the 2008 Workshop on the Economics of Networks, Systems, and Computation (NetEcon'08)*, pages 43–48, Seattle, WA, Aug. 2008.
33. T. Sandler and K. Hartley. Economics of alliances: The lessons for collective action. *Journal of Economic Literature*, XXXIX(3):869–896, Sept. 2001.
34. S. Schechter and M. Smith. How much security is enough to stop a thief? In *Proceedings of the Seventh International Financial Cryptography Conference (FC 2003)*, pages 122–137, Gosier, Guadeloupe, Jan. 2003.
35. B. Schneier. Tactics, targets, and objectives. *Wired.com*, May 2007.
36. StratFore. Situation report: Cyberwarfare and botnets. April 2008.
37. The Honeynet Project. Know your enemy: the tools and methodologies of the script-kiddie, July 2000. Available online at `http://project.honeynet.org/papers/enemy/`.
38. H. Varian. System reliability and free riding. In L. Camp and S. Lewis, editors, *Economics of Information Security (Advances in Information Security, Volume 12)*, pages 1–15. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.
39. N. Weaver and V. Paxson. A worst-case worm. In *Proceedings of the Third Annual Workshop on Economics and Information Security (WEIS'04)*, Mineapolis, MN, May 2004. Available at `http://www.dtc.umn.edu/weis2004/weaver.pdf`.
40. N. Weinberg. Botnet economy runs wild. *Network World*, April 2008.

A Mixed strategy for weakest target game without mitigation

In the following we investigate whether a mixed strategy can be derived. Assume there is a cumulative distribution of protection strategies F. We can use the cumulative distribution of the binomial distribution to represent the chance that a player will be compromised given a fixed e. Then the expected utility of Blue is

$$Blue = p_a L \sum_{j=0}^{k-1} \binom{N-1}{j} F_e^j (1-F_e)^{N-1-j} - be_i - cs_i \tag{21}$$

In Nash equilibria, the first order condition must hold:

$$\begin{aligned} 0 &= p_a L(N-k)\tbinom{N-1}{j} F_e^{k-1}(1-F_e)^{N-1-k}(f) - b \\ \frac{b}{p_a L(N-k)\binom{N-1}{j}} &= F_e^{k-1}(1-F_e)^{N-1-k}(f) \\ \frac{b}{p_a L(N-k)\binom{N-1}{j}} &= exp\{(k-1)lnF_e + (N-1-k)ln(1-F_e)\}(f) \\ f &= \frac{b}{p_a L(N-k)\binom{N-1}{j} exp\{(k-1)lnF_e + (N-1-k)ln(1-F_e)\}} \end{aligned}$$

Then we can expand the exponentiated part about $e^* =$ the median of f using a Taylor expansion. Thus,

$$f = \frac{b}{p_a L(N-k)\binom{N-1}{j}(\frac{1}{2})^{N-2}(1+2(2k-N)f_{e^*}(e-e*))} \tag{22}$$

$$\text{where } f_{e^*} = \frac{b}{p_a L(N-k)\binom{N-1}{j}} \tag{23}$$

$$\text{thus } f = \frac{f_{e^*}}{(1+2(2k-N)f_{e^*}(e-e*))} \tag{24}$$

The approximation of f about e^* is asymptotic as $e \to e*$. Knowing that Blue will never play $e > p_a L/b$ because of dominance, we estimate $e^* = p_a L/b$.

If insurance is not overpriced, then we know $F(0) = q; Blue(0,0) = c$:

$$pl \sum_{j=0}^{k-1} \binom{N-1}{j} q^j (1-q)^{N-1-j} = c \tag{25}$$

Using a Taylor expansion again, we find:

$$(\frac{1}{2})^{N-1}\sum_{j=0}^{k-1}\binom{N-1}{j} - (\frac{1}{2})^{N-1}\binom{N-1}{k-1}(N-k)(q-.5) = c/p_aL \tag{26}$$

$$-\binom{N-1}{k-1}(N-k)(q-.5) = \frac{c}{p_aL}2^{N-1} - \sum_{j=0}^{k-1}\binom{N-1}{j} \tag{27}$$

$$q = .5 + (\sum_{j=0}^{k-1}\binom{N-1}{j} - \frac{c}{p_aL}2^{N-1})/\binom{N-1}{k-1}(N-k) \tag{28}$$

B Extensions

Below we discuss selected extensions of practical relevance for security researchers and practitioners that address limitations of our model. At first, we consider the case of repeated games. Next we analyze situations in which the membership to the class of defender or attacker is not stable. That is, for example, the case if a computer is compromised and acts as another attacker in consecutive rounds.

B.1 Repeated Games

To simplify the analysis of repeated interactions we state selected assumptions: (i) Costs are fixed over time, (ii) Players have an inherent discount factor, δ, (iii) Games are played infinitely, and iv) we do not address the case of self-insurance, as it has no effect on deterring attackers. This analysis is not exhaustive and should serve as an outlook on future results and as a sensitivity analysis for the stage game results.

Total Effort: Consider the Blue trigger strategy "Always protect if the expected losses are higher than the cost, and no one has defected." We set up the game totals to find for which values of δ this strategy holds: $-p_AL(1-\frac{N-1}{N}) - \frac{\delta}{1-\delta} < \frac{b}{1-\delta}$. Solving for p and negating to find Red's new ceiling yields $p_A < \frac{bN}{L(\delta(N+1)-1)}$.

If δ is 1, then the probability of attack has fallen by a factor of N compared to the one-shot game. If the discount factor is 0, then the solution remains the same.

Weakest Link: Setting up a similar inequality, $\frac{1}{1-\delta}(-b) < \frac{1}{1-\delta}(-p_AL)$, and observing that delta is free, because protection is a stage game Nash we can solve for p and negate to find Red's new ceiling: $p_A < \frac{b}{L}$.

In the repeated case of Weakest Link, Red's ceiling is the exact same as in the one shot with an $\hat{e}_0 = 1$. However, the repeated nature of this game is enough to allow Blue to coordinate full protection rather than settling on an expected minimum.

Best Shot: Imagine a strategy where each round a different Blue player would take a turn purchasing full protection, as long as the expected losses were sufficiently high and no one had previously deviated. We can set up the inequality: $\frac{1}{1-\delta^N}(-b) < \frac{1}{1-\delta}(-p_A L)$ to find the new ceiling for the attacker: $p_A < \frac{b(1-\delta)}{L(1-\delta^N)}$.

Unlike the one shot version of the best shot game, the repeated version does bound Red's attack. More significantly, this effect gets stronger rather than weaker as N increases. This is due to the diffusion of costs over several rounds rising from the turn-taking nature of the strategy.

Weakest Target: Consider a similar strategy, where instead of taking turns to protect, k players take turns being an unprotected honeypot and play $\eta \to 0$. Thereby exploiting a similar strategy as the social optimum would prescribe [18]. Setting up the required inequality $-2\eta b - \frac{\delta}{1-\delta} p_A L < \frac{1}{1-\delta^N}(\frac{(1-\delta^{k+1})p_A L - (1-\delta^N)\eta b}{1-\delta})$ and taking the limit we find $\frac{1-\delta^N}{1-\delta^k} > \frac{b}{p_A L}$. Thus, Red's new ceiling is $p_A < \frac{b}{L}\frac{1-\delta^k}{1-\delta^N}$. There is an additional second constraint $k < Nb/L$.

In the case of no mitigation and infinite attacker strength, there is still nothing Blue can do other than remain passive and potentially insure. However, in the mitigation case where $k < N$, we see that this strategy is proportionally strong to the ratio of k and N.

Defender Coordination This shows that coordination issues can be overcome in certain repeated cases as long as the defender has a sufficient discount factor for the trigger strategies. Defenders have a slight advantage because there are several cases where the defenders can coordinate inside the game; attackers cannot.

B.2 Extensive-Form Games

These models still do not take into account any changes in the number of attackers or defenders, which is a major characteristic of automated malicious code, or to put it differently, viruses are viruses because they spread. There are several rules we could use when trying to capture contagious effects. For example:

- Players that do not insure and are compromised become attackers for the rest of the game.
- Players that do insure and are compromised are dropped from the rest of the game.

Sketch of proof Starting with M attackers and N defenders, and working in the limits of both parameters in a loosely coupled game, k players at most are compromised each turn. Of these, βk do not insure. Thus, in round x you would expect $M + \beta k x$ attackers and $N - kx$ defenders. In an infinite number of rounds, there exists some round x such that the number of attackers crosses the point described in Eq. 20, and after that point attackers can no longer coordinate absent some external communication. In this case, there is no Nash and the probability of attack approaches 1, which would activate Blue's protection trigger. Once

this happens, the game remains in that equilibrium until time expires. Thus, we would expect the empirical cumulative distribution of attackers ("infection") over time to be monotone increasing but decelerating up until that point, at which point it remains level.

C Summary of results

Table 1. Attacker Nash Strategies by Network Contribution Function

Type	H_e	One Shot	Repeated
1 player	e	b/L	b/L
Total Effort	$\frac{1}{N}\sum e$	Nb/L	b/L
Weakest Link	$\min(e)$	b/L	b/L
Best Shot	$\max(e)$	1	0
Weakest Target	0 if $e_i < e_{(k)}$	1	1
Weakest Target with Mitigation	e_i if $e_i < e_{(k)}$	b/L	b/L