Security investment (failures) in five economic environments: A comparison of homogeneous and heterogeneous user agents *

Jens Grossklags[a] *Nicolas Christin*[b] *John Chuang*[a]
[a]School of Information
University of California, Berkeley
Berkeley, CA 94720
{jensg,chuang}@ischool.berkeley.edu

[b]CyLab Japan and Information Networking Institute
Carnegie Mellon University
1-3-3-17 Higashikawasaki-cho, Chuo-ku
Kobe, Hyogo 650-0044, Japan
nicolasc@cmu.edu

Working draft: Id: paper.tex 175 2008-03-04 02:30:53Z jensg

Abstract

Security interactions in networked systems, and the associated user choices, due to their complexity, are notoriously difficult to predict, and sometimes even harder to rationalize. We argue that users often underestimate the strong mutual dependence between their security strategies and the economic environment (e.g., threat model) in which these choices are made and evaluated. This misunderstanding weakens the effectiveness of users' security investments.

We study how economic agents invest into security in five different economic environments, that are characteristic of different threat models. We consider generalized models of traditional public goods games (e.g., total effort and weakest link) and two recently proposed games (e.g., weakest target game). Agents may split their contributions between a public good (protection) and a private good (self-insurance). We can examine how incentives may shift between investment in a public good (protection) and a private good (insurance), subject to factors such as network size, type of attack, loss probability, loss magnitude, and cost of technology. We compare Nash equilibrium predictions for homogeneous and heterogeneous user populations. We also provide results for the social optima for different classes of attacks and defenses in the case of homogeneous agents.

*This is a working paper in preparation for archival journal submissions and represents a subset of results from an ongoing project on security economics. Some of the preliminary results on the homogeneous case presented here have been published in [21]. Results on the case of heterogeneous agents are available also in [22]. We hope to further improve this draft with the feedback of the workshop participants.

1 Introduction

The Internet has opened new and attractive channels to publicize and market products, to communicate with friends and colleagues, and to access information from spatially distributed resources. Though it has grown significantly, the network's architecture still reflects the cooperative spirit of its original designers [37]. Unfortunately, today's network users are no longer held together by that same sense of camaraderie and common purpose. For instance, concrete evidence of the tragedy of the commons [23] occurring in peer-to-peer filesharing networks has been documented for a long time [2]. Accordingly, studies of networking protocols and user interaction have been assuming users to be selfish and to act strategically [43].

Selfish users are one thing, but the expansion of the Internet has also attracted individuals and groups with often destructive motivations; these "attackers" intend to improve on their perceived utility by exploiting or creating security weaknesses and harming or inconveniencing other network users [46]. Some malicious entities are motivated by peer recognition, or curiosity, and are often undecided regarding the ethical legitimacy of their behavior [19, 20]. Others have clearly demonstrated financial goals [15]. Problematic behaviors and threats include attacks on the network as a whole, attacks on selected end-points, undesirable forms of interactions such as spam e-mail, and annoyances such as Web pages that are unavailable or defaced. As a result, users cannot rely and trust other network participants [12].

When asked in surveys, network users say they are interested in preventing attacks and mitigating the damages from computer and information security breaches [1]. Researchers and industry have responded by developing numerous security technologies to alleviate many of the aforementioned problems [4, 41], which is expected to help improving individual security practices.

Nevertheless, security breaches are common, widespread and highly damaging. The "I Love You" virus [30], Code Red [32] and Slammer worms [31], to cite the most famous cases, have infected hundreds of thousands of machines and caused, all together, billions of dollars in damages. This high financial impact is explained by recent surveys [6, 8], which shows strong evidence that comprehensive security precautions, be they patching, spyware-removal tools, or even sound backup strategies, are completely amiss from a vast majority of systems surveyed.

In other words, despite a self-professed interest in security, most individuals do not implement any security on their systems, even though security technology is (by and large) readily available. We propose to investigate the root causes of the disconnect between users' actions and their intentions.

In practice, there is a large variety of situations in which users face security threats, and an equally large number of possible responses to threats. However, we postulate in this paper that one can model most security interactions through a handful of "security games," and with a small number of decision parameters upon which each user can act.

More precisely, building upon public goods literature [26, 49], we consider the classical best shot, total effort, and weakest link games, and will analyze them in a security context. We complement these three games with a novel model, called the "weakest target" game, which allows us to describe a whole class of attacks ranging from insider threats to very aggressive worms. Furthermore, while most research on the

economics of security focuses on security investments as a problem with a single variable (e.g., amount of money spent on security), our analysis is the first security study to decouple protection investments (e.g., setting up a firewall) from self-insurance coverage (e.g., archiving data as back up). This decoupling allows us to explain a number of inefficiencies in the observed user behaviors.

This paper is only a first step toward a more comprehensive modeling of user attitudes toward security issues. Indeed, the present study relies on game theory, mostly using Nash equilibrium and social optima concepts. As such, we primarily view this study as a theoretical basis for follow up experimental work using laboratory experiments with human participants. We nevertheless show that the models and results derived here allow to gain considerable insights.

The rest of this paper is organized as follows. We elaborate in Section 2 the relationship of our work with related research, and discuss some important security consequences that results from homogeneous and heterogeneous user agent populations. In Section 3 we introduce our game-theoretic models. We present an analysis of the Nash equilibria in the homogeneous user case in (Section 4) and for a heterogeneous user base in (Section 5). We discuss differences between the individually rational solution and the social optima determined by a social planner in (Section 6) for all of these games, however, we defer the analysis of the heterogenous case for a revised version of this paper. We discuss our findings in Section 7. We conclude in Section 8.

2 Related Work

The economics of information security is a growing research area with a diverse set of participating researchers from various disciplines. Important common anchors are the observations that misaligned incentives and positive and negative externalities play significant roles in the strategies used by each party in the battle between attackers and potential victims [4, 5].

Economics as a tool for security analysis has gained in importance since the economy of attackers has become increasingly rational (e.g., motivated by greed), over the last years [15]. This increasingly rational behavior stands in contrast to that exhibited by the hacker communities of the 1980s and 1990s, who valued reputation, intellectual achievement, and even entertainment above financial incentives [19, 20].

Most of the initial results obtained in security economics research concern the analysis of optimal security investments. For example, Gordon and Loeb [18] as well as Hausken [25] focus on the impact of different security breach functions and degrees of vulnerability on an entity's investment strategy. More specialized models have been proposed to analyze in more depth a subset of important security management problems. For instance, August and Tunca [7] scrutinize optimal system update strategies when patching a system against security vulnerabilities is costly. Rescorla investigates the impact of code quality control on vulnerability of software [36].

From a policy standpoint, Bull et al. [9] observe the state of heterogeneous networks and argue that no single security policy will be applicable to all circumstances. They argue that, for a system to be viable from a security standpoint, individuals need to be empowered to control their own resources and to make

customized security trade-offs. This stands in contrast to the traditional centralized structure where all security decision are made by a central planner (e.g., the IT department). Nevertheless, as Anderson suggests, organizational and structural dependencies have to be considered in individual security decision making [3].

While many models prescribe behavior in individual choice situations, the focus of our work is to model and study strategic interaction with respect to security decisions in networked systems, in an effort to understand the impact of individual choices on a larger group. Such interaction usually involves common as well as conflicting interests. (Pure conflict, in which the interests of the two antagonists are completely opposed, is a special case.) This mutual dependence as well as opposition guarantees for a much richer scenario for analysis [40].

In the context of mutual dependence, Varian [49] introduces the analysis of system reliability within a public good framework. He discusses the best effort, weakest link and total effort games, as originally analyzed by Hirshleifer [26]. The main difference from classical public goods theory is that within the framework of computer reliability "considerations of costs, benefits, and probability of failure become paramount, with income effects being a secondary concern."[49] Varian focuses on two-player games with heterogeneous effort costs and benefits from reliability.[1] He also adds an inquiry into the role of taxes and fines, and differences between simultaneous and sequential moves.

Our work generalizes [49] in several aspects. First, instead of considering security decisions to be determined by a single "security" variable, we identify two key components of a security strategy: self-protection (e.g., patching system vulnerabilities) and self-insurance (e.g., having good backups). More precisely, we allow agents to self-protect and/or self-insure their resources in N-player games. We also contrast the three canonical games discussed by Varian with two more complex "weakest target" games that represent a more complicated incentive structure, which we believe applies to a whole class of security issues.

Outside the information security context, the dual role of self-protection and self-insurance was first recognized by [14]. To provide a more precise definition, self-protection stands for the ability to reduce the probability of a loss – for example, by installing a firewall application which limits the amount of traffic allowed to communicate with one's network. Self-insurance, on the other hand, denotes a reduction in the magnitude of a loss, e.g., by performing regular backups on existing data. Some technologies and practices such as disconnecting a computer from a network do both. Ehrlich and Becker [14] focus in their analysis on the comparison of self-protection and self-insurance to market insurance. They find that, for rare loss events, there is less incentive to self-insure losses than to use market insurance. This is due to their assumption, that the price of self-insurance is independent of the probability of the loss. An additional result is that the demand for self-insurance grows with the base loss of a security threat. As an outcome of their work, they characterize self-insurance and market insurance as substitutes, and self-protection and market insurance as complements. Our analysis complements the work in [14] by extending the concepts of self-protection and self-insurance to the public goods and security context.

[1] A distinction between reliability and security, in terms of consequences, may exist [27]. In this study, we do not follow this distinction and consider reliability as a key component of security.

2.1 Homogeneity versus heterogeneity in system security

Both the homogeneous and heterogeneous cases are relevant to security analysis. Homogeneous agents are characteristic of large populations following the same practices and choices by end-users, for instance, when most security decisions (e.g., patching) are automated, and all users run similar software. The lack of diversity, in particular in the market for operating systems, lends credibility to such scenarios [17], and is cited as a strong motivator for developers of malicious code to exploit the resulting correlated risks or to cheaply repeat attacks.

However, there are strong reasons to compare the homogeneous case with a model that intrpoduces heterogeneous agents into the analysis of security decision making.

Security through diversity. Recent technical proposals aim to achieve higher resilience to attacks by introducing diversity in network and protocol design. For example, Zhuang et al. report of a set of formal analysis tools that introduce heterogeneity in multi-person communication protocols [51]. O'Donnell and Sethu develop and test distributed algorithms optimizing the distribution of distinct software modules to different nodes in a network [33]. Lv et al. study potential improvements in the scalability of P2P networks by exploiting heterogeneity in the user population [29]. Danezis and Anderson find in the presence of heterogeneous preferences that censorship-resistance of a system will be weakened if materials are distributed randomly across all nodes rather than according to users' preferences [13]. Research in IT economics has evaluated the decision making of a firm when faced with the option of increased diversity in its software base. In Chen et al. the decision for increased heterogeneity depends largely on the assumed risk attitudes of the organization [10]. Investments into heterogeneity will change the expectation of losses and attack probabilities, but they also impact the cost of protection and self-insurance.

Chameleonic threats. Increased diversity is not a sufficiently strong protection against correlated security threats anymore. Already in 1995 the first macro viruses started targeting MS office on all compatible systems.[2] Modern cross-platform malware is capable of targeting also different operating systems. For instance, Linux-Bi-A/Win-Bi-A is written in assembler and able to compromise Windows and Linux platforms. Malicious code is also capable of crossing the boundary between desktop and mobile devices. Potentially even more disruptive is malware carrying multiple exploit codes at once. For example, Provos et al. report that Web-based malware often includes exploits that are used 'in tandem' to download, store and then execute a malware binary [35]. These trends render users vulnerable to propagated threats if owners of different IT systems perceive protection as too costly or ineffective.

Heterogeneous investments patterns. Different organizations follow distinct patterns of IT investment. Parts of organizations often depend on legacy systems including weakly protected systems, or "boat anchors"

[2]The macro virus (Winword-Concept) targeted Microsoft Word on Apple and Microsoft systems. For more details see: `http://web.textfiles.com/virus/macro003.txt`.

with limited value to an organization [50]. Such legacy systems can allow skilled attackers to intrude a network. More generally, organizations and end users justify security investments with different assumptions about potential losses and probabilities of being attacked. This often depends on different knowledge about threats and means of protection and insurance [1]. This diversity is reflected in users' choices and security practices [6, 8]. Similarly, security decisions can follow different security paradigms often reflected in different organizational structures, for instance remote replication vs. offsite tape storage.

In this paper, we formally explore both assumptions about the composition of agent populations, by studying individuals' incentives in non-cooperative games. In particular, we focus on the impact of homogeneous and heterogeneous agents on system security in different network structures.

3 Five canonical security games

A security game is a game-theoretic model that captures essential characteristics of decision making to protect and self-insure resources within a network of agents. In this section, we summarize the security games we analyze. We provide the equations for the heterogeneous user case and omit presenting the simplified formulas for the homogeneous case that are available in [21].

As discussed earlier, we model security as a hybrid between public and private goods. On the one hand, as was previously observed by Varian [49], the success of security (or reliability) decision making frequently depends on a joint protection level determined by all participants of a network. The computation of the protection level will often take the form of a public goods contribution function. Because network protection is a public good, it may allow, for certain types of contribution functions, individuals to free-ride on others' efforts. At the same time, some individuals may also suffer from inadequate protection efforts by other members if those have a decisive impact on the overall protection level.

In addition to self-protection, network participants 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.

All security games we introduce share the following key assumptions: (i) all entities in the network share a single purely public protection output, and (ii) a single individual decides on protection efforts for each entity – we do not assume a second layer of organizational decision making.

Compared to the homogeneous case [21], protection costs per unit are not necessarily identical for each entity, and, while in the formal analysis that follows we make the assumption that all decisions are made simultaneously, we later discuss the impact of relaxing the synchronization assumption.

We develop security games from a basic model with the following payoff structure. Each of $N \in \mathbb{N}$ players receives an individual endowment M_i. If she is attacked and compromised successfully she faces a loss L_i. Attacks arrive with a probability of p_i ($0 \leq p_i \leq 1$), which albeit exogenous, is also dependent on the player under consideration; p_i remains constant over time. 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,

$b_i \geq 0$ and $c_i \geq 0$ denote the unit cost of protection and insurance, respectively. The generic utility function of Player i is defined as:

$$U_i = M_i - p_i L_i (1 - s_i)(1 - H(e_i, e_{-i})) - b_i e_i - c_i s_i \ , \tag{1}$$

where, following common game-theoretic notation, e_{-i} denotes the set of protection levels chosen by players other than i. H is a contribution function of e_i, which is required to 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. (1), the magnitude of a loss depends on three factors: i) whether an attack takes place (p_i), 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. (1) therefore yields an expected utility.

We rely on five games in the following discussion. In selecting and modeling these games we paid attention to comparability of our security games to prior research (e.g., [26, 38, 49]). The first three specifications for H represent important baseline cases recognized in the public goods literature. To allow us to cover most security dilemmas, we add two games, which we originally introduced only in the context of homogeneous agents [21].

Total effort security game: 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. (1) becomes

$$U_i = M_i - p_i L_i (1 - s_i)(1 - \frac{1}{N} \sum_k e_k) - b_i e_i - c_i s_i \ . \tag{2}$$

Economists identified the sum of efforts (or total effort) contribution function long before the remaining cases included in this paper [26]. We consider a slight variation of this game to normalize it to the desired parameter range.

As a practical example of a total effort game in practice, consider parallelized file transfers, as in the BitTorrent peer-to-peer service. It may be the case that an attacker wants to slow down transfer of a given piece of information; but the transfer speed itself is a function of the aggregate effort of the machines participating in the transfer. Note that, the attacker in that case is merely trying to slow down a transfer, and is not concerned with completely removing the piece of information from the network: censorship actually results in a different, "best shot" game, as we discuss later.

Weakest-link security game: 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. (1) takes the form:

$$U_i = M_i - p_i L_i (1 - s_i)(1 - \min(e_i, e_{-i})) - b_i e_i - c_i s_i \ . \tag{3}$$

The weakest-link game is the most often recognized public goods problem in computer security. Once the perimeter of an organization is breached it is often possible for attackers to leverage this advantage. This initial compromise can be the result of a weak password, an inconsistent security policy, or some malicious code infiltrating a single client computer. Another example is that of a two-way communication (e.g., TCP flow), where the security of the communication is determined by the least secure of the communication parties. For instance, a TCP flow between a host with a perfectly secure TCP/IP stack and a host with an insecure TCP/IP stack can be easily compromised.

Best shot security game: 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. (1) becomes

$$U_i = M_i - p_i L_i (1 - s_i)(1 - \max(e_i, e_{-i})) - be_i - cs_i \; . \tag{4}$$

Among information systems, networks with built-in redundancy, such as peer-to-peer, sensor networks, or even Internet backbone routes, share resilience qualities with the best shot security game; for instance, to completely take down communications between two (presumably highly connected and highly secure) backbone nodes on the Internet, one has to shut down all possible routes between these two nodes. Censorship-resistant networks are another example of best shot games. A piece of information will remain available to the public domain as long as a single node serving that piece of information can remain unharmed [13].

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

$$H(e_i, e_{-i}) = \begin{cases} 0 & \text{if } e_i = \min(e_i, e_{-i}), \\ 1 & \text{otherwise,} \end{cases} \tag{5}$$

which leads to

$$U_i = \begin{cases} M_i - p_i L_i (1 - s_i) - b_i e_i - c_i s_i & \text{if } e_i = \min(e_i, e_{-i}), \\ M_i - b_i e_i - c_i s_i & \text{otherwise.} \end{cases} \tag{6}$$

The weakest target game differs from the weakest link. There is still a decisive security level that sets the benchmark for all individuals. It is determined by the individual(s) with the lowest chosen effort level. However, in this game all entities with a protection effort strictly larger than the minimum will remain unharmed.

In information security, this game captures the situation in which an attacker is interested in securing access to an arbitrary set of entities with the lowest possible effort. Accordingly, she will select the machines with the lowest security level. An attacker might be interested in such a strategy if the return on attack effort is relatively low, for example, if the attacker uses a compromised machine to distribute spam. Such a strategy is also relevant to an attacker with limited skills, a case getting more and more frequent with the availability of automated attack toolboxes [47]; or, when the attacker's goal is to commandeer the largest number of

machines using the smallest investment possible [15]. Likewise, this game can be useful in modeling insider attacks – a disgruntled employee may for instance very easily determine how to maximize the amount of damage to her corporate network while minimizing her effort.

Weakest target security game (with mitigation): This game is a variation on the above weakest target game. The difference is that, the probability that the attack on the weakest protected player(s) is successful is now dependent on the security level $\min e_i$ chosen. That is,

$$H(e_i, e_{-i}) = \begin{cases} 1 - e_i & \text{if } e_i = \min(e_i, e_{-i}), \\ 1 & \text{otherwise,} \end{cases} \tag{7}$$

so that

$$U_i = \begin{cases} M - p_i L_i (1 - s_i)(1 - e_i) - b_i e_i - c_i s_i & \text{if } e_i = \min(e_i, e_{-i}), \\ M - b_i e_i - c_i s_i & \text{otherwise.} \end{cases} \tag{8}$$

This game represents a nuanced version of the weakest target game. Here, an 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.

Limitations: As suggested by Hirshleifer [26], practical scenarios may involve social composition functions combining two or more of these five games. Revisiting our earlier examples, protecting a communication flow between two hosts may be a "weakest-link" game, until a certain level of host security is reached at both hosts. At that point, the attacker may start to target the routes between the hosts rather than the hosts themselves, and it becomes a "best-shot" game. Other realistic environments may be better characterized by slight variations on a given game (e.g., "the total of the three best shots"). We nevertheless believe that the five games described above may capture a large number of practical cases, as our argument made in earlier work [21] is actually strengthened by being able to compare the representative user case with the scenario with diverging user preferences.

4 Nash equilibrium analysis with homogeneous agents

We next determine the equilibrium outcomes where each individual chooses protection effort and self-insurance investments unilaterally, in an effort to maximize her own utility. We first consider the case of homogeneous agents. In (Section 5) we will identify differences to the heterogeneous case. In Section 6, we then compare results Nash equilibria to the protection efforts and self-insurance levels chosen if coordinated by a social planner. However, currently we have only completed the social optimality analysis for the representative user case.

We assume homogeneous users to share the same values for cost of protection and self-insurance. Individuals also face the same threats with identical consequences if compromised.

4.1 Total effort

Let us focus on player i, and consider e_k for $k \neq i$ as exogenous. Then, U_i is a function of two variables, e_i and s_i. From Eqn. (2), U_i is twice differentiable in e_i and s_i, with $\partial^2 U_i/\partial s_i^2 = 0$ and $\partial^2 U_i/\partial e_i^2 = 0$. Hence, according to the second derivative test, only $(e_i, s_i) \in \{(0,0),(0,1),(1,0),(1,1)\}$ can be an extremum – that is, possible Nash equilibria are limited to these four values (or to strategies yielding a payoff constant regardless of e_i and/or s_i). As long as at least one of b or c is strictly positive, $(e_i, s_i) = (1,1)$ is always dominated by either $(e_i, s_i) = (1,0)$ or $(e_i, s_i) = (0,1)$ and cannot define a Nash equilibrium. Let us analyze the three other cases:

- $(e_i, s_i) = (0,0)$. Replacing in Eqn. (2), we get

$$U_i = M - pL\left(1 - \frac{1}{N}\sum_{k\neq i} e_k\right) . \tag{9}$$

- $(e_i, s_i) = (0,1)$. Replacing in Eqn. (2), we get

$$U_i = M - c\,. \tag{10}$$

- $(e_i, s_i) = (1,0)$. Replacing in Eqn. (2), we get

$$U_i = M - pL\left(1 - \frac{1}{N} - \frac{1}{N}\sum_{k\neq i} e_k\right) - b\,. \tag{11}$$

Result 1: *After investigating Eqs. (9–11) we can identify three Nash equilibrium strategies.*

- *Full protection eq.:* If $pL > bN$ and $c > b + pL\frac{N-1}{N}$, meaning that protection is cheap, potential losses are high, and insurance is extremely overpriced, then the (only) Nash equilibrium is defined by everybody protecting but not insuring, that is, $(e_i, s_i) = (1,0)$.

- *Full self-insurance eq.:* In the other cases where $pL > bN$, $(e_i, s_i) = (0,1)$ is a Nash equilibrium. Also, if $c < pL < bN$ (expected losses above insurance costs), then $(e_i, s_i) = (0,1)$, is a Nash equilibrium.

- *Passivity eq.:* If $pL < bN$ and $pL < c$, then the expected losses are small enough so that complete passivity, defined by $(e_i, s_i) = (0,0)$ for all players, is a Nash equilibrium.

Proof. First, consider the case when $pL > bN$. This means that U_i as given by Eqn. (11) is higher than that given by Eqn. (9), which means that only $(e_i, s_i) = (0,1)$ and $(e_i, s_i) = (1,0)$ are potential utility-maximizing strategies. If, for a player i, the initial protection levels picked by other players satisfy $\sum_{k\neq i} e_k > N - 1 - N\frac{c-b}{pL}$, then U_i given by Eqn. (11) dominates that given by Eqn. (10), and player i

accordingly chooses $(e_i, s_i) = (1, 0)$. (Note this is only possible if $b < c$.) All players have the same rationale, and we end up with everybody picking $(e_i, s_i) = (1, 0)$, which is a Nash equilibrium.

On the other hand, if $\sum_{k \neq i} e_k < N - 1 - N\frac{c-b}{pL}$, then player i chooses $(e_i, s_i) = (0, 1)$. All players adopt the same strategy, so that we end up with an insurance-only configuration, where everybody gets a utility $M - c$. If $pL\frac{N-1}{N} + b > c$, that is, $pL > \frac{(c-b)N}{N-1}$, this is a Nash equilibrium with $\forall i (e_i, s_i) = (0, 1)$. (This is always the case when $b > c$.) On the other hand, with $pL < \frac{(c-b)N}{N-1}$, people have an incentive to switch to the strategy $(e_i, s_i) = (1, 0)$, and to remain in that strategy.

Next, consider the case when $pL < bN$. The utility given by Eqn. (9) is greater than that given by Eqn. (11). Then, for each player i, if the initial protection levels picked by other players satisfy $\sum_{k \neq i} e_k > N(pL - c)$, then Eqn. (9) gives a higher utility than Eqn. (10).[3] Hence, player i picks $(e_i, s_i) = (0, 0)$ as a strategy. If the game is synchronized, all players have the same rationale and we end up in a situation where everybody selects $(e_i, s_i) = (0, 0)$. If $pL < c$ then all players are content staying at $(e_i, s_i) = (0, 0)$, otherwise, they switch to $(e_i, s_i) = (0, 1)$, and stay in that strategy. □

Increasing number of players N: As the number of players increases, protection equilibria become more and more unlikely to occur. Indeed, in a total effort scenario, "revenues" yielded by a player's investment in security have to be shared with all of the other participants, making it an increasingly uninteresting strategy for the player as the network grows.

4.2 Weakest-link

Let $e_0 = \min_i(e_i)$. From Eqn. (3), we have $U_i = M - pL(1 - s_i)(1 - e_0) - be_i - cs_i$, so that for all i,

$$U_i \leq M - pL(1 - s_i)(1 - e_0) - be_0 - cs_i \, ,$$

which is reached for $e_i = e_0$. So, in a Nash equilibrium, everybody picks the same $e_i = e_0$. It follows that Nash equilibria are of the form $(e_0, 0)$ or $(0, 1)$. For strategy $(e_i, s_i) = (e_0, 0)$, we have

$$U_i = M - pL + (pL - b)e_0 \, , \tag{12}$$

while for strategy $(e_i, s_i) = (0, 1)$, we have

$$U_i = M - c \, . \tag{13}$$

Which strategy is a Nash equilibrium depends therefore on the relative values of pL, b and c, and the sign of $(pL - c) - (pL - b)e_0$.

[3] The border case $\sum_{k \neq i} e_k = N(pL - c)$ yields a situation where the initial constellation of parameters form an unstable Nash equilibrium, where no one has any incentive to change their strategy.

Result 2: *In the weakest link security game, we can identify three types of Nash equilibrium strategies. However, there exist multiple pure protection equilibria.*

Denote by $\hat{e}_0$ the minimum of the protection levels initially chosen by all players. We have

- *Multiple protection equilibria:* If $pL > b$ and either 1) $pL < c$ or 2) $pL \geq c$ and $\hat{e}_0 > (pL-c)/(pL-b)$ then $(e_i, s_i) = (\hat{e}_0, 0)$ for all i is a Nash equilibrium. Everybody picks the same minimal security level, but no one has any incentive to lower it further down. This equilibrium can only exist for $b \leq c$, and may be inefficient, as it could be in the best interest of all parties to converge to $e_i = 1$, as we discuss later in Section 6.

- *Full self-insurance eq.:* If $pL > c$ and either 1) $pL < b$ or 2) $pL \geq b$ and $\hat{e}_0 < (pL - c)/(pL - b)$, then $(e_i, s_i) = (0, 1)$ for all i is a Nash equilibrium: essentially, if the system is not initially secured well enough (by having all parties above a fixed level), players prefer to self-insure.

- *Passivity eq.:* If $pL < b$ and $pL < c$, then $(e_i, s_i) = (0, 0)$ is the only Nash equilibrium – both insurance and protection are too expensive.

Notice that if $\hat{e}_0 = (pL - c)/(pL - b)$, then both full self-insurance ($(e_i, s_i) = (0, 1)$ for all i) and protection ($(e_i, s_i) = (\hat{e}_0, 0)$ for all i) form a Nash equilibrium. In particular, if $b = c$ ($pL > b$ and $pL > c$), and $\hat{e_0} = 1$, full protection $(e_i, s_i) = (1, 0)$ and full self-insurance $(e_i, s_i) = (0, 1)$ are Nash strategies.

Increasing number of players N: The weakest link security game, much like the tacit coordination game of [48] has highly volatile protection equilibria when the number of players increase. In fact, any protection equilibrium has to contend with the strategic certainty of a self-insurance equilibrium. To view this, consider the cumulative distribution function $F(e_i)$ over the protection strategies e_i of a given player i. From what precedes, with pure strategies, in the Pareto-optimum, $F(1) = 1$ and $F(e_i) = 0$ for $e_i < 1$. Assuming all N players use the same c.d.f. F, then the c.d.f. of $e_0 = \min_i\{e_i\}$ is given by $F_{\min}(e_0) = 1 - (1 - F(e_0))^N$ [48]. So, $F_{\min}(1) = 1$ and $F_{\min}(e_0) = 0$ for $e_0 < 1$ as well. Now, assume there is an arbitrarily small probability $\varepsilon > 0$ that one player will defect, that is $F(0) = \varepsilon$. Then, $F_{\min}(0)$ converges quickly to 1 as N grows large. That is, it only takes the slightest rumor that one player may defect for the whole game to collapse to the $(e_i, s_i) = (0, 1)$ equilibrium.

4.3 Best shot

Let $e^* = \max_i(e_i)$. Eqn. (4) gives

$$U_i = M - pL(1 - s_i)(1 - e^*) - be_i - cs_i \, .$$

Clearly, $(e_i, s_i) = (1, 1)$ is suboptimal, so that three strategies may yield the highest payoff to user i.

- Selecting $(e_i, s_i) = (0, 0)$ yields $U_i = M - pL(1 - e^*)$.

- Selecting $(e_i, s_i) = (1, 0)$ yields $U_i = M - b$.
- Selecting $(e_i, s_i) = (0, 1)$ yields $U_i = M - c$.

Result 3: *From the above relationships, we can identify the following pure Nash equilibrium strategies.*

- *Full self-insurance eq.:* If $b < c$ we find that the self-insurance equilibrium ($\forall i, (e_i, s_i) = (0, 1)$) is the only possible Nash equilibrium.
- *Passivity eq.:* If $pL < b$ and $pL < c$ agents prefer to abstain from security actions ($\forall i, (e_i, s_i) = (0, 0)$).

In particular, there is no protection equilibrium in this game. For one protection equilibrium to exist, we would need $b < c$ and $pL > b$. But even assuming that this is the case, as long as the game is synchronized, players endlessly oscillate between securing as much as possible ($e_i = 1$) and free-riding ($e_i = 0$). This is due to the fact that as soon as one player secures, all others have an incentive to free-ride. Conversely, if everybody free-rides, all players have an incentive to deviate and secure as much as possible.

Increasing number of players N: In the absence of coordination between players, the outcome of this game is globally independent of the number of players N, as there is no protection equilibrium, and the insurance equilibrium is independent of the number of players. However, the game may be stabilized by using player coordination (e.g., side payments) for low values of N, something harder to do as N grows.

4.4 Weakest-target (without mitigation)

Fix the strategy point and let $\varepsilon < \frac{pL}{2b}$. Let e_0 be the minimum effort level of any player. Then no player selects a higher effort than $e_0 + \varepsilon$ because it dominates all higher effort levels. However, any player at e_0 would prefer to switch to $e_0 + 2\varepsilon$. Then the change in her payoff is greater than $pL - 2\frac{pL}{2b}b = 0$. Because this deviation is profitable this strategy point is not an equilibrium.

Result 4: *In the weakest-target game with an attacker of infinite strength we find that pure Nash equilibria for non trivial values of b, p, L and c do not exist.*

Mixed strategy equilibria. While no pure Nash equilibria exist, let us explore the existence of a mixed strategy equilibrium. We use the shorthand notation $e_i = e$, $s_i = s$ here, and consider mixed strategies for choosing e. There are two cases to consider.

Case $c > pL$: If $c > pL$ then dominance arguments immediately lead to $s = 0$ meaning that nobody buys any self-insurance.

An equilibrium strategy may be parameterized by e. For a given player, the utility function U becomes a function of a single variable e. Let $f(e)$ be the probability distribution function of effort in the weakest-target game and let $F(e)$ be the cumulative distribution function of effort. Assuming only one player is at the

minimum protection level, shall an attack occur, the probability of being the victim is then $(1 - F(e))^{N-1}$. (All N players choose protection levels greater than e.)

Then the utility is given by

$$U = M - pL(1 - F(e))^{N-1} - be \, . \tag{14}$$

In a Nash equilibrium, the first-order condition $dU/de = 0$ must hold, so that:

$$(N-1)pLf(e)[1 - F(e)]^{N-2} - b = 0$$

If we substitute $G = (1 - F(e))$ and $g = -f$ we can write $G^{N-2}dG/de = -b/p(N-1)L$, which, by integration yields

$$\int_{G(e)}^{G(0)} G^{N-2} dG = \int_e^0 \frac{-b}{p(N-1)L} d\hat{e} \, ,$$

that is

$$G^{N-1}\Big|_{G(e)}^{G(0)} = \frac{-b}{pL} e \, . \tag{15}$$

With $G(0) = 1$,

$$G(e) = \left(1 - \frac{b}{pL} e\right)^{\frac{1}{N-1}} \, .$$

Differentiating, we get

$$g(e) = -\frac{1}{N-1} \frac{b}{pL} \left(1 - \frac{b}{pL} e\right)^{-\frac{N-2}{N-1}} \, ,$$

and, replacing $g = -f$ we find,

$$f(e) = \frac{1}{N-1} \frac{b}{pL} \left(1 - \frac{b}{pL} e\right)^{-\frac{N-2}{N-1}} \, , \tag{16}$$

as the probability distribution function of self-protection in a mixed Nash equilibrium.

Case $c \leq pL$: Now let us consider a game with insurance under the more reasonable assumption $c \leq pL$; that is, insurance is not overpriced compared to expected losses. Dominance arguments indicate that a Nash strategy must be of the form $(e, s) \in \{(e, 0), e \geq 0\} \cup \{(0, 1)\}$.

Let q be the probability that a player chooses strategy $(e, s) = (0, 1)$. That is, $F(0) = q$. Because insurance is independent of protection, we can reuse Eqn. (15) with the new boundary $G(0) = 1 - q$:

$$G(e) = \left((1 - q)^{N-1} - \frac{b}{pL} e\right)^{\frac{1}{N-1}} \tag{17}$$

However, since we are now including self-insurance, a second condition must hold. The payoff for strategy $(e, s) = (0, 1)$ must equal the payoff for all other strategies.

Specifically, we may compare payoffs for strategies $(e, s) = (\varepsilon, 0)$ and $(e, s) = (0, 1)$ which gives, by continuity as $\varepsilon \to 0$,

$$pL(1-q)^{N-1} = c\,. \tag{18}$$

Together Eqs. (17) and (18) yield:

$$F(e) = 1 - G(e) = 1 - \left(\frac{c - be}{pL}\right)^{\frac{1}{N-1}}\,,$$

which, differentiating, gives

$$f(e) = \frac{1}{N-1}\frac{b}{pL}\left(\frac{c - be}{pL}\right)^{\frac{1}{N-1}-1}\,. \tag{19}$$

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

$$q = F(0) = 1 - \left(\frac{c}{pL}\right)^{\frac{1}{N-1}}\,. \tag{20}$$

Result 5: *In the weakest-target game with an attacker of infinite strength, a mixed Nash equilibrium strategy exists. The individual's strategy is given by Eqs. (19) and (20).*

Increasing number of players N: From Eqn. (20), we can directly infer that an increase in the number of participating players decreases the probability that a full self-insurance strategy is chosen. When N grows large, q tends to zero, which means that players increasingly prefer to gamble in order to find a protection level that leaves them unharmed.

4.5 Weakest target (with mitigation)

Let us assume that there exists a Nash equilibrium where $0 < K < N$ players who satisfy $e_i = e_0 = \min(e_i, e_{-i})$, while $(N - K > 0)$ players satisfy $e_i > e_0$. We can show that such an equilibrium does not exist and that players rather congregate at the highest protection level if certain conditions are met. By computing the partial derivatives $\partial U_i/\partial s_i$ and $\partial U_i/\partial e_i$, and discriminating among values for e_i and s_i, we get the following results. (We refer the reader to Appendix A for the complete proof.)

Result 6: In contrast to the infinite strength weakest-target game we find that a pure Nash equilibrium may exist.

- *Full protection eq.:* If $b \leq c$ we find that the full protection equilibrium $(\forall i, (e_i, s_i) = (1, 0))$ is the only possible pure Nash equilibrium.
- *For $b > c$ we can show that no pure Nash equilibrium exists.*
- *There are no pure self-insurance equilibria.*

Mixed strategy equilibrium To complement this analysis we also present the mixed strategy equilibrium. The derivation is similar to the one given by Eqs. (14–20), however, with an additional substitution step. This gives the resulting distribution,

$$F(e) = 1 - \left(\frac{c - be}{pL(1-e)}\right)^{\frac{1}{N-1}} , \tag{21}$$

so that

$$f(e) = \frac{1}{N-1}\left(\frac{(b-c)pL}{pL^2(1-e)^2}\right)\left(\frac{c-be}{pL(1-e)}\right)^{-\frac{N-2}{N-1}} .$$

Interestingly, the probability of playing $(e, s) = (0, 1)$ remains

$$q = F(0) = 1 - \left(\frac{c}{pL}\right)^{\frac{1}{N-1}} \tag{22}$$

Note that if $c < b$ there is a zero probability that $e = 1$ will be chosen by any player. The upper bound for protection effort is given by $e_{\max} = c/b$.

Result 7: *In the weakest-target game with an attacker of finite strength we find that a mixed Nash equilibrium strategy exists. The relevant equations are given in Eqs. (21–22).*

5 Nash equilibrium analysis with heterogeneous agents

In this section, our focus is to understand how the inclusion of heterogeneous actors influences predictions compared to a model with representative agents. In Section 2, we have discussed arguments for and against homogeneity in security models. In the modeling of economic phenomena, added complexity (e.g., adding agents with more diverse tastes) does not always change strategic predictions substantially. On the other hand, we expect that heterogeneity impacts the actions of agents in security games in different ways, for example by: 1) Negotiating the trade-off between protection and insurance, 2) Highlighting certain strategies and focal points due to the inherent differences in the agent population, 3) (De-)stabilizing equilibrium predictions derived in the homogeneous case. We expect several conclusions from the homogeneous case to remain relevant. But as Hartley [24] argued "representative agents models conceal heterogeneity whether it is important or not." This analysis aims at pinpointing key differences and discuss their implications.

5.1 Total effort

The total effort game yields considerably different results depending on the number of players involved.

Two-player game Let us first start the discussion for the simple case $N = 2$. From the game description given by Eqn. (2), we get $U_1(e_1, s_1) = M_1 - p_1 L_1(1 - s_1)(1 - (e_1 + e_2)/2) - b_1 e_1 - c_1 s_1$ for Player 1. The second partial derivative test indicates that there is no local extremum, so that the only possible maxima of U_1

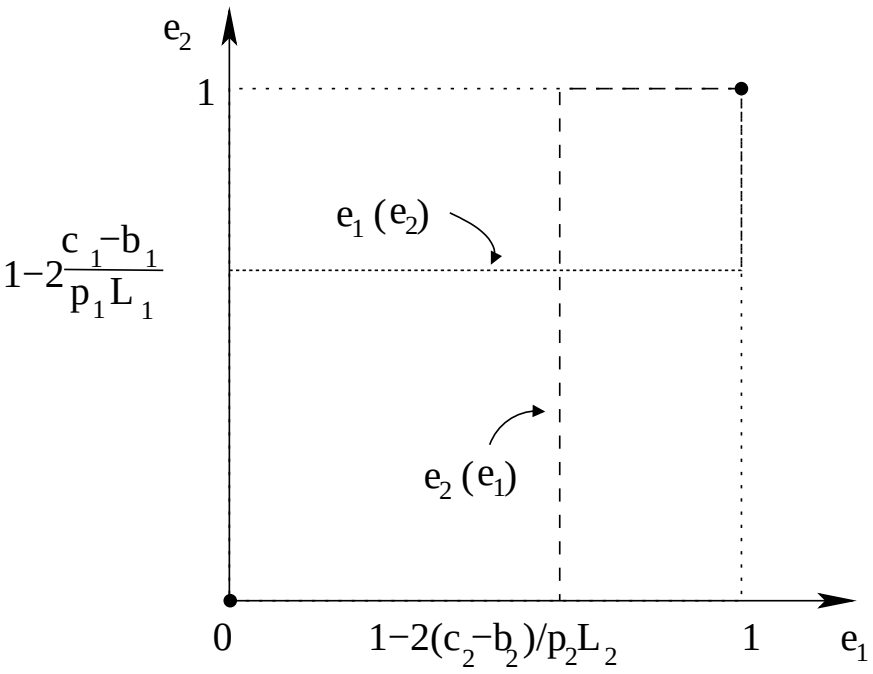

Figure 1: **Reaction functions for a two-player total effort game**. Bold lines and dots indicate potential Nash equilibria.

are given by $U_1(0,0) = M_1 - p_1 L_1(1 - e_2/2)$, $U_1(1,0) = M_1 - p_1 L_1(1/2 - e_2/2) - b_1$, $U_1(0,1) = M_1 - c_1$, or $U_1(1,1) = M_1 - b_1 - c_1$. With $b_1 > 0$, we immediately see that $U_1(0,1) > U_1(1,1)$, which tells us that fully insuring and protecting at the same time is a strictly dominated strategy for Player 1. The passivity strategy $(e_i, s_i) = (0,0)$ dominates the "protect-only" strategy $(e_i, s_i) = (1,0)$ when $b_1 > p_1 L_1/2$.

Assuming $b_1 \leq p_1 L_1/2$, the "protect-only" $(1,0)$ strategy dominates the "insure-only" $(0,1)$ strategy for Player 1 if and only if (all quantities being assumed to be defined):

$$e_2 > 1 - 2\frac{c_1 - b_1}{p_1 L_1} \, . \tag{23}$$

A similar rationale yields the corresponding conditions for Player 2, leading to the reaction functions $e_1(e_2)$ and $e_2(e_1)$ plotted in Figure 1. By definition, Nash equilibria are characterized by fixed points $e_1(e_2) = e_2(e_1)$. From the above analysis summarized in Figure 1, this occurs for two values: when both agents fully protect and when both agents abstain from investing in protection. We note that both fixed points are stable, meaning that, if they are reached, minimal deviations in the strategy of one player are unlikely to perturb the actions of the other player.

Result 8: *The two-player total effort security game with heterogeneous agents presents the following equilibria:*

- *Full protection eq.:* If $b_1 \leq p_1 L_1/2$, $b_2 \leq p_2 L_2/2$ (protection costs are modest for both players), and the initial values $e_1(0)$ and $e_2(0)$ satisfy either $e_1(0) > 1 - 2(c_2 - b_2)/(p_2 L_2)$ or $e_2(0) > 1 - 2(c_1 - b_1)/(p_1 L_1)$ (at least one player is initially fairly secure, or at least one player faces very high insurance costs) then the (only) Nash equilibrium is defined by both players protecting but not insuring, that is, $(e_i, s_i) = (1, 0)$.

- *Multiple eq. without protection:* If the conditions above do not hold, then we have an insecure equilibria. Both players converge to $e_1 = 0$ and $e_2 = 0$. Their respective investments in insurance depend on whether their insurance premium is smaller than their potential losses: a player will fully insure if and only if $c_i < p_i L_i$, and will be passive otherwise.

A particularly interesting feature of the two-player version of the game is that expensive insurance or protection costs at *either* of the players directly condition which equilibrium can be reached. For instance, if one of the players has to pay a very high insurance premium in front of its protection costs, she will elect to protect, likely leading the other player to protect as well. Conversely, if either of the players faces a high protection premium ($b_i > p_i L_i/2$), the game will likely converge to an equilibrium without protection efforts. As we discuss later, this property can be used by some form of intervention to have the game converge to a desirable equilibrium.

More generally, in this game, each of the two players generally tracks what the other is doing. When moves are made perfectly simultaneously, this may result in oscillations between insecure and secure configurations. The only exception to this tracking behavior occurs when one player faces high security costs and a low insurance premium, while the other faces the opposite situation (low security costs, very high insurance premium). In such a case, the game converges to the first player insuring, and the second player protecting. In short, extreme parameter values allow to remove network effects in this game.

N-player game (N large) In the more general case $N \geq 2$, we first notice that, for a security strategy to be meaningful, we need to have $b_i < p_i L_i/N$. This means that, as the number of player increases, individual protection costs have to become very small, or expected losses have to considerably increase. Failing that, insurance or passivity is always a better option.

Second, from Eqn. (2), we obtain that Eqn. (23) is generalized to

$$\frac{1}{N-1} \sum_{j \neq i} e_j > 1 - \frac{N}{N-1} \frac{c_i - b_i}{p_i L_i} \,, \tag{24}$$

as a condition for player i to select a protection-only strategy as opposed to an insurance-only strategy. Eqn. (24) tells us that, for large values of N, changes in a single player's protection strategy are unlikely to have much of an effect on the other players' strategies. Indeed, each player reacts to changes in the average protection level over the $(N-1)$ other players.

This observation brings the question of exactly how robust the N-player game is to a change in the strategy played by a given individual. Are "domino effects" possible, where changes in a single player's

strategy, albeit with a minimal effect on all other players, lead another player to switch strategies, and eventually to large groups changing their plays?

To help us answer this question, let us consider $N > 2$, and $K \leq N$ arbitrary players that are initially (at time 0) unprotected. For instance, assume without loss of generality that Players $1, \ldots, K$ are initially unprotected, and that

$$\frac{c_2 - b_2}{p_2 L_2} \geq \frac{c_3 - b_3}{p_3 L_3} \geq \ldots \geq \frac{c_K - b_K}{p_K L_K} .$$

Further assume that at a later time $t > 0$, Player 1 switches her strategy to full protection, that is, $e_1(t) = 1$. Assuming all players may have an incentive to protect (i.e., for all i, $b_i < p_i L_i / N$), Player 2 would also switch to full protection only if

$$\frac{1}{N-1} \sum_{j \neq 2} e_j(t) > 1 - \frac{1}{N-1} \frac{c_2 - b_2}{p_2 L_2}$$

that is, only if

$$\frac{1}{N-1} \sum_{j \neq 2} e_j(0) + \frac{1}{N-1} > 1 - \frac{1}{N-1} \frac{c_2 - b_2}{p_2 L_2} ,$$

which reduces to

$$\frac{1}{N-1} \sum_{j > K} e_j(0) + \frac{1}{N-1} > 1 - \frac{1}{N-1} \frac{c_2 - b_2}{p_2 L_2} . \tag{25}$$

Player 2's switch causes Player 3 to switch too only if

$$\frac{1}{N-1} \sum_{j \neq 3} e_j(t) > 1 - \frac{1}{N-1} \frac{c_3 - b_3}{p_3 L_3} ,$$

that is,

$$\frac{1}{N-1} \sum_{j > K} e_j(0) + \frac{2}{N-1} > 1 - \frac{1}{N-1} \frac{c_3 - b_3}{p_3 L_3} . \tag{26}$$

From Eqs. (25) and (26) we get

$$\frac{c_2 - b_2}{p_2 L_2} - \frac{c_3 - b_3}{p_3 L_3} < 1 .$$

Iterating over the K players that are initially not protecting, we get:

$$\max_{2 \leq i \leq K} \frac{c_i - b_i}{p_i L_i} - \min_{2 \leq i \leq K} \frac{c_i - b_i}{p_i L_i} < K - 1 .$$

We can follow an identical derivation for the case where the K players switch from a protection strategy to a non-protection strategy. We then obtain the following necessary condition for "domino effects" to occur over K players, that is a switch in Player 1's strategy causing a switch in the strategy of K players:

$$\left| \max_{2 \leq i \leq K} \frac{c_i - b_i}{p_i L_i} - \min_{2 \leq i \leq K} \frac{c_i - b_i}{p_i L_i} \right| < K - 1 . \tag{27}$$

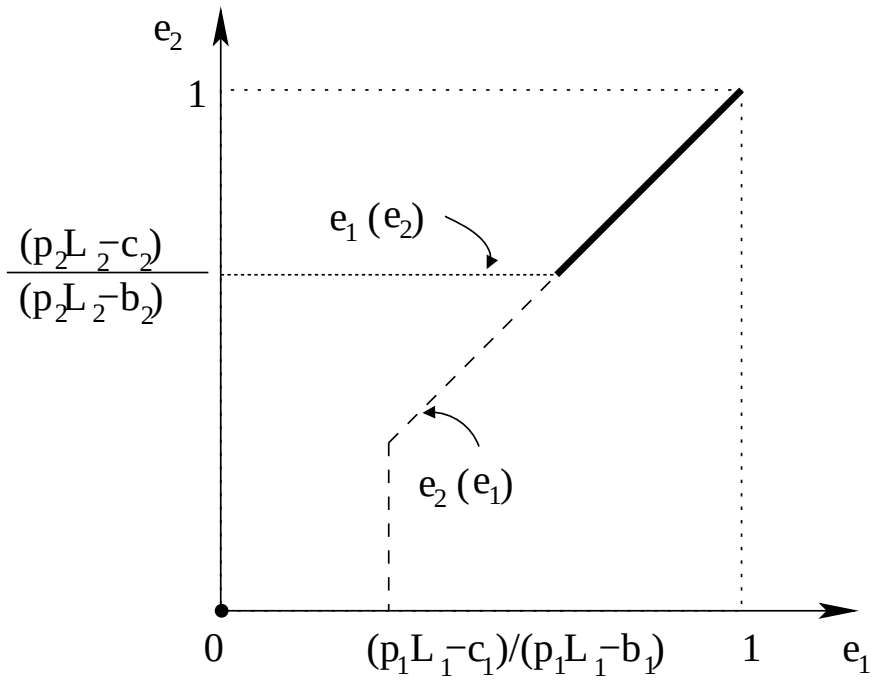

Figure 2: **Reaction functions for a two-player weakest-link game**. Bold lines and dots indicate potential Nash equilibria.

Result 9: *We have derived a stability measure of the heterogeneity of a total effort security game with N agents (Eqn. (27)). The more heterogeneous the players are, the more unlikely Eqn. (27) is to hold for large values of K. In other words, the more heterogeneous a system is, the more likely it is to be resilient to perturbations due to a single individual changing strategies.*

5.2 Weakest-link

Here again, we start by considering a two-player game. Computing partial derivatives in e_i and s_i from Eqn. (3), we observe that each player chooses either $(e_i, s_i) = (0, 1)$ (insurance strategy) or $(e_i, s_i) = (\min_{j\neq i} e_j, 0)$ (protection strategy, where in the two-player version of the game $\min_{j\neq i} e_j$ is naturally equal to the protection value chosen by the other player) in order to maximize their utility function.

Looking at the payoffs that can be obtained in both cases leads us to the reaction functions of both players, which we plot in Figure 2. In the figure, we see that a fixed-point is attained when $e_1 = e_2 = 0$ (insurance-only equilibria) and when both e_1 and e_2 are greater than $\max\{(p_1L_1-c_1)/(p_1L_1-b_1), (p_2L_2-c_2)/(p_2L_2-b_2)\}$.

Result 10: *Generalizing to N players, we obtain the following distinction for the weakest link security game:*

- *Full protection eq.:* If, for all i, $p_iL_i > b_i$, and either 1) $p_iL_i < c_i$, or 2) $p_iL_i \geq c_i$ and $\hat{e}(0)$, the minimum of the security levels initially chosen by all players, satisfies

$$\hat{e}(0) > \max_{1 \leq i \leq N} \{(p_iL_i - c_i)/(p_iL_i - b_i)\} ,$$

  then we have a Nash equilibrium where everyone picks $(\hat{e}(0), 0)$.

- *Multiple eq. without protection:* All players select $e_i = 0$ if the conditions above do not hold. The value of insurance they select depends on their respective valuations. Players for whom insurance is too expensive ($p_iL_i < c_i$) do not insure, with $s_i = 0$, while others choose full insurance, that is $s_i = 1$.

The likelihood of reaching a full protection equilibrium is conditioned by the player which has the largest difference between protection and insurance costs relative to its expected losses. In particular, it only takes one player with an insurance premium smaller than its protection cost ($b_i > c_i$) to make the full protection equilibrium unreachable. Hence, when N grows large, we expect protection equilibria to become more and more infrequently observed.

5.3 Best shot

Looking at the variations of the payload function U_i given in Eqn. (4) as a function of e_i and s_i tells us there are three possibilities for maximizing U_i: a passivity strategy $(0, 0)$, a secure-only strategy $(1, 0)$ and an insure-only strategy $(0, 1)$.

We get $U_i(0, 0) = M_i - p_iL_i(1 - \max\{e_{-i}\})$, $U_i(1, 0) = M_i - b_i$, and $U_i(0, 1) = M_i - c_i$. We immediately notice that $b_i > c_i$ leads Player i to never invest in protection: either the player is passive, or she insures. If, on the other hand $b_i \leq c_i$, then player i chooses a protection strategy over a passivity strategy if and only if (b_i assumed greater than 0) we have $\max\{e_{-i}\} < 1 - b_i/p_iL_i$. We plot the reaction functions, in a two-player case, in Figure 3.

Result 11: *For the two-player best shot security game we can identify the following equilibria:*

- *Protection eq.:* In contrast to the homogeneous case a protection equilibrium does exist. The Nash equilibrium is a free-riding equilibrium where one player protects, and the other does not.

- *Multiple eq. without protection:* If $b_i > c_i$ for all player i individuals will choose to self-insure or remain passive.

In the homogeneous version of the game, we note that these Nash equilibria are not reached in a synchronized game with N players, as players would constantly oscillate between free-riding and protecting. With heterogeneous players, however, it is possible to reach a Nash equilibrium. Indeed, if the initial protection levels chosen satisfy $\max\{e_{-i}(0)\} > 1 - b_i/p_iL_i$ for all players *but one*, this last player will be the only one to secure, while everybody else will defect. Note that there should be only one player choosing

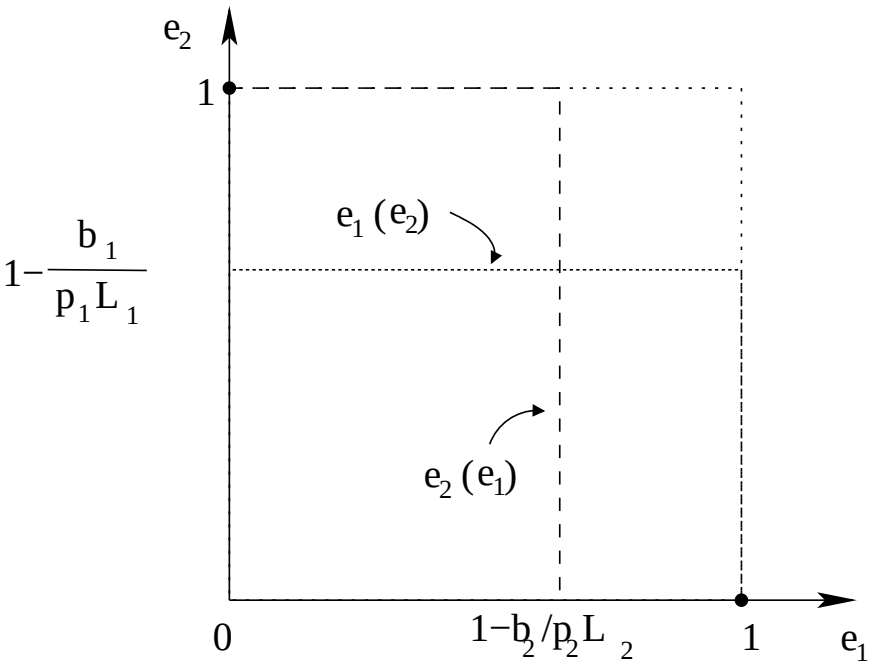

Figure 3: **Reaction functions for a two-player best shot game**. Bold dots indicate potential Nash equilibria. Protection costs are assumed here to be smaller than insurance costs for both players.

to secure for a Nash equilibrium to be reached – as soon as at least two players decide to protect, each will defect in the next round hoping to free-ride on the other protecting players. In other words, if there exists a unique i for which the initial constellation of protection levels satisfies

$$\max\{e_{-i}(0)\} < 1 - b_i/p_i L_i \ , \tag{28}$$

then a Nash equilibrium where all players free-ride on player i is reached as long as $b_i < c_i$. This situation could happen when only one player faces disproportionate losses compared to other players, or her security costs are very small.

Result 12: *When protection levels are initially randomly set, protection equilibria in the best shot game are increasingly unlikely to happen as the number of players N grows.*

Assume that the initial protection levels, $e_i(0)$ for $1 \leq i \leq N$ are set independently and at random, that is, that they can be expressed as a random variable with cumulative distribution function F. Then for any Player k, the probability that $e_k(0) < 1 - b_i/p_i L_i$ is simply $F(1 - b_i/p_i L_i)$. It follows that Eqn. (28) is satisfied for Player i with probability $F(1 - b_i/p_i L_i)^{N-1}$.

Next, we want Eqn. (28) to be violated for all players other than i. Eqn. (28) is defeated for a given Player k with probability $1 - F(1 - b_k/p_k L_k)^{N-1}$. Consequently, it is defeated for all Players $j \neq i$ with probability $\prod_{j \neq i}(1 - F(1 - b_j/p_j L_j)^{N-1})$.

It follows that the probability ρ_i that Eqn. (28) is satisfied *only* for Player i is given by

$$\rho_i = F\left(1 - \tfrac{b_i}{p_i L_i}\right)^{N-1} \prod_{j \neq i} \left(1 - F\left(1 - \tfrac{b_j}{p_j L_j}\right)^{N-1}\right) .$$

Then, the probability that a protection equilibrium can be reached is given by $\sum_i \rho_i$, since the ρ_i's characterize mutually exclusive events. To simplify notations, let $x_i = F\left(1 - \frac{b_i}{p_i L_i}\right)$. Rearranging terms gives

$$\sum_i \rho_i = \sum_i \prod_{j \neq i} \left(1 - x_j^{N-1}\right) - N \prod_j \left(1 - x_j^{N-1}\right) .$$

Let $k = \arg\max_i \left\{\prod_{j \neq i} \left(1 - x_j^{N-1}\right)\right\}$. Then we have

$$\sum_i \rho_i \leq N \prod_{j \neq k} \left(1 - x_j^{N-1}\right) - N \prod_j \left(1 - x_j^{N-1}\right) ,$$

which gives us, after rearranging

$$\sum_i \rho_i \leq N x_k^{N-1} \prod_{j \neq k} \left(1 - x_j^{N-1}\right) ,$$

which tends to zero as N increases, as soon as $x_k = F(1 - b_k/p_k L_k) < 1$.

This is notably the case if we assume a function F strictly monotonous increasing on $[0, 1]$, and positive security costs ($b_i > 0$) for all players.

5.4 Weakest target

As in the homogeneous case, Nash equilibria for the weakest target game are quite different depending on whether or not we are considering that mitigation is possible.

Without mitigation. In the weakest target game without mitigation, we find that, in the homogeneous case where $b_i = b$, $c_i = c$, $p_i = p$ and $L_i = L$, there are no pure strategy Nash equilibrium. The proof can be extended to the heterogeneous case, as we discuss next.

Let us assume that the minimum protection level over all players is set to $\hat{e} < 1$. Then, we can group players in two categories: those who play $e_i = \hat{e}$, and those who set $e_i > \hat{e}$. By straightforward dominance arguments coming from the description of the payoffs in Eqn. (6), players who select $e_i > \hat{e}$ select $e_i = \hat{e} + \varepsilon$, where $\varepsilon > 0$ is infinitesimally small, and $s_i = 0$. Let

$$\varepsilon < \min_i \left\{ \frac{p_i L_i}{2 b_i}(1 - s_i) + \frac{c_i s_i}{2 b_i} \right\} .$$

Players who play $e_i = \hat{e}$ would actually prefer to switch to $\hat{e} + 2\varepsilon$. Indeed, the switch in strategies allows a payoff gain of

$$U_i(\hat{e} + 2\varepsilon, 0) - U_i(\hat{e}, s_i) = -2 b_i \varepsilon + p_i L_i (1 - s_i) + c_i s_i > 0 .$$

Hence, this strategy point is not a Nash equilibrium. It follows that the only possible equilibrium point would have to satisfy $e_i = 1$ for all e_i. However, in that case, all players are attacked, which ruins their security investments. All players therefore have an incentive to instead select $e_i = \hat{e} = 0$, which, per the above discussion, cannot characterize a Nash equilibrium.

Result 13: *In the weakest-target game without mitigation we find that pure Nash equilibria for non trivial values of b_i, p_i, L_i and c_i do not exist.*

With mitigation. In the weakest target game with mitigation, we find that, with homogeneous agents, a full protection Nash equilibrium exists as long as protection costs are smaller than insurance costs. An exactly identical proof can be conducted in the heterogeneous case to show that a full protection equilibrium is reached if $b_i < c_i$ for all i.

On the other hand, it only takes one of the players to face high security costs to make this equilibrium collapse. Indeed, if there exists k such that $b_k > c_k$, then Player k will always prefer a full-insurance strategy ($(e_k, s_k) = (0, 1)$) over a full-protection strategy ($(e_k, s_k) = (1, 0)$). This will immediately lead other players to try to save on security costs by picking $e_i = \varepsilon > 0$ as small as possible. We then observe an escalation as in the unmitigated version discussed above. Hence, heterogeneity actually threatens the (precarious) stability of the only possible Nash equilibrium.

Result 14: In contrast to the weakest-target game without mitigation we find that a pure Nash equilibrium may exist.

- *Full protection eq.:* If $b_i \leq c_i$ for all agents we find that the full protection equilibrium ($\forall i, (e_i, s_i) = (1, 0)$) is the only possible pure Nash equilibrium.
- If $b_i > c_i$ for *any* agent we can show that no pure Nash equilibrium exists.
- There are no pure self-insurance equilibria.

6 Identification of social optima with homogeneous agents

Organizations and public policy actors frequently attempt to identify policies that provide the highest utility for the largest number of people. This idea has been operationalized with the social optimum analysis. It states that a system has reached the optimum when the sum of all players' utilities is maximized. That is, the social optimum is defined by the set of strategies that maximize $\sum_i U_i$. Consider N players, and denote by $\Phi(e_1, s_1, \ldots, e_N, s_N)$ the aggregate utility, $\Phi(e_1, s_1, \ldots, e_N, s_N) = \sum_i U_i(e_i, s_i)$. The social optimum maximizes $\Phi(s_i, e_i)$ over all possible $(s_i, e_i) \in [0, 1]^{2N}$. Because enforcing a social optimum may at times be conflicting with the optimal strategy for a given (set of) individual(s), to enforce a social optimum in practice, we may need to assume the existence of a "social planner" who essentially decides, unopposed, the strategy each player has to implement.

Below we discuss key-differences of the optimal strategy derived by a social planner with the individually-optimal strategy for the case of homogeneous agents. We are currently working on the comparison of the heterogeneous case for potential presentation at the workshop.

6.1 Total effort game

Summing the utility given by Eqn. (2) over i, and performing the substitution $E = \sum_i e_i$ and $S = \sum_i s_i$, we get

$$\Phi = N(M - pL) + (pL - b)E + (pL - c)S - \frac{pL}{N}ES \; .$$

Φ is continuous and twice differentiable in E and S, with

$$\begin{cases} \frac{\partial \Phi}{\partial E} &= pL - b - \frac{pL}{N}S \; , \\ \frac{\partial \Phi}{\partial S} &= pL - c - \frac{pL}{N}E \; . \end{cases}$$

Furthermore, $\partial^2\Phi/\partial E\partial S = -pL/N < 0$, while $\partial^2\Phi/\partial E^2 = \partial^2\Phi/\partial S^2 = 0$. The second derivative test tells us that the only possible extrema of Φ are reached for the boundary values of E and S, that is $(E, S) \in \{0, N\}^2$. In other words, the only possible social optima are 1) passivity (for all i, $(e_i, s_i) = (0, 0)$), 2) full protection (for all i, $(e_i, s_i) = (1, 0)$), or 3) full insurance (for all i, $(e_i, s_i) = (0, 1)$). As long as one of b or c is strictly positive, a social planner will never advise agents to invest into protection and self-insurance at the same time.

By comparing the values of Φ in all three cases, we find that if $b < pL$ and $b < c$ then all agents are required to exercise maximum protection effort $(e_i, s_i) = (1, 0)$. With $c < pL$ and $c < b$ all agents will self-insure at the maximum possible $(e_i, s_i) = (0, 1)$. A social planner will not encourage players to invest in security measures if they are too expensive ($c > pL$ and $b > pL$).

Result 15: *In the total effort security game we observe that in the Nash equilibrium there is almost always too little protection effort exerted compared to the social optimum. In fact, for a wide range of parameter settings no protection equilibria exist while the social optimum prescribes protection at a very low threshold.*

- *Protection:* Except for very unbalanced parameter settings (i.e., $pL > bN$ and $c > b + pL\frac{N-1}{N}$) agents refrained from full protection. Now full protection by all agents is a viable alternative.

- *Self-insurance:* Full self-insurance now has to compete with full protection effort under a wider range of parameters.

- *Passivity:* Agents remain passive if self-insurance is too expensive ($c > pL$). However, we find a substantial difference with respect to protection behavior. Agents would selfishly refrain from protection efforts if $pL < bN$ since they would only be guaranteed the N-th part of their investments as returns. Now the social planner can ensure that all agents protect equally so that it is beneficial to protect up until $b < pL$.

6.2 Weakest link game

In the weakest link game agents are required to protect at a common effort level to be socially efficient. We compute Φ by summing Eqn. (3) over i, and can express Φ as a function of e_i, s_i and $e_0 = \min_i(e_i)$:

$$\Phi = NM - NpL(1-e_0) - b\sum_i e_i + (pL(1-e_0) - c)\sum_i s_i \; .$$

In particular, for all i, we obtain

$$\partial\Phi/\partial s_i = pL(1-e_0) - c \; .$$

Studying the sign of $\partial\Phi/\partial s_i$ as a function of e_0 allows us to determine the social optimum.

First assume, that $\frac{\partial\Phi}{\partial s_i} > 0$. This requires $pL > c$, and is equivalent to $e_0 < 1 - \frac{c}{pL}$. $s_i = 1$ gives the extremum, which is of the form $\Phi = NM - b\sum_i e_i - cN$ (independent of e_0). We then need to pick $e_i = e_0 = 0$ for all i, to get

$$\Phi = N(M-c) \; , \tag{29}$$

as a possible social optimum.

Now, assume that $\frac{\partial\Phi}{\partial s_i} \leq 0$, which is equivalent to having $e_0 \geq 1 - c/pL$. Then, to get the extremum, one should pick $s_i = 0$, which leads to $\Phi = NM - pL(1-e_0)N - b\sum_i e_i$. Φ is maximized for $e_i = e_0$ for all i, leading to $\Phi = N(M - pL + (pL-b)e_0)$. There are two cases to distinguish:

- If $pL > b$, then one should choose $e_0 = 1$, and get

$$\Phi = N(M-b) \; . \tag{30}$$

- If, on the other hand, $pL \leq b$ then one should choose $e_0 = \max(1 - \frac{c}{pL}, 0)$. That is, if $pL \leq c$, then we get $e_0 = 0$, and

$$\Phi = N(M - pL) \; , \tag{31}$$

  but if $pL > c$, then the social optimum could only be given by $e_0 = 1 - \frac{c}{pL}$, to satisfy our initial condition, and be of the form

$$\Phi = N\left(M - b - c + \frac{bc}{pL}\right) \; . \tag{32}$$

However it is clear that as long as the coefficients are non-trivial, that last equation does not characterize a social optimum.

In summing, if $b < c$ and $b < pL$ the social planner requires all agents to protect with maximum effort $(e_i, s_i) = (1, 0)$. If $c < b$ and $c < pL$ the social planner requires all agents to self-insure $(e_i, s_i) = (0, 1)$. Finally, the Nash equilibrium and social optimum coincide when security costs are high. Agents do not invest in protection or self-insurance if $b > pL$ or $c > pL$.

Result 16: *The availability of self-insurance lowers the risk of below-optimal security in the Nash equilibrium since agents have an alternative to the unstable Pareto-optimal protection equilibrium. From the analysis of the weakest link game with many agents we know that deviation from the Pareto-optimal highest protection level is very likely. A social planner can overcome these coordination problems.*

- *Protection:* The *Pareto-optimal* Nash equilibrium coincides with socially optimal protection. However, the protection level would likely be lower in the Nash case due to coordination problems.
- *Self-insurance:* The self-insurance equilibria are equivalent for the Nash and social optimum analysis.
- *Passivity:* A social planner cannot expand the range of parameter values at which it would be socially beneficial to protect or self-insure while passivity would be prescribed in the Nash equilibrium.

6.3 Best shot game

We compute the social optimum by summing U_i given in Eqn. (4) over i, yielding that Φ can be expressed as a function of e_i, s_i, and $e^* = \max_i(e_i)$, as

$$\Phi = NM - NpL(1-e^*) - b\sum_i e_i + (pL(1-e^*) - c)\sum_i s_i \ .$$

It is immediate that, to maximize Φ, one should pick $e_i = 0$ for all i, except for one participant j, where $e_j = e^* \geq 0$. We then get

$$\Phi = NM - NpL(1-e^*) - be^* + (pL(1-e^*) - c)\sum_i s_i \ ,$$

so that $\partial\Phi/\partial s_i = pL(1-e^*) - c$, which tells us under which conditions on e^* (and consequently on b, c, and pL) self-insurance is desirable.

We distinguish between two cases:

- If $e^* < 1 - \frac{c}{pL}$, then $\frac{\partial\Phi}{\partial s_i} > 0$, and so, one should pick $s_i = 1$ for all i. Then, we get

  $$\Phi = NM - Nc - be^* \ ,$$

  which is maximized for $e^* = 0$ (satisfying our condition $e^* < 1 - \frac{c}{pL}$) and

  $$\Phi = N(M-c) \ . \tag{33}$$

- If $e^* \geq 1 - \frac{c}{pL}$, then $\frac{\partial\Phi}{\partial s_i} \leq 0$, and so, one should pick $s_i = 0$ for all i.[4] Then, we get

  $$\Phi = NM - NpL + (NpL - b)e^* \ .$$

[4] In the equality case, the value of Φ is independent of that of s_i so that we can pick $s_i = 0$.

If $b < NpL$, then Φ increases in e^*, so and is maximized for $e^* = 1$, with

$$\Phi = NM - b \,. \tag{34}$$

If $b < NpL$, then one should pick e^* as small as possible, that is, given our initial condition, $e^* = 1 - \frac{c}{pL}$, and we get

$$\Phi = N(M - c) - \frac{bc}{pL} \,. \tag{35}$$

With $b > 0$, $c > 0$, and $pL > 0$, it is clear that Eqn. (35) is not a social optimum, since it is dominated by Eqn. (33).

To summarize, we find that if $b/c < N$ (i.e., protection is not at a prohibitive cost compared to insurance and/or there is a reasonably large number of players), the social optimum is to have one player protect as much as possible, the others not protect at all, and no one insures. In practice, this may describe a situation where all participants are safely protected behind an extremely secure firewall. If, on the other hand $b/c > N$, which means there are either few players, insurance is very cheap compared to protection, then the best strategy is to simply insure all players as much as possible.

Result 17: *In the best shot security Nash outcome there is almost always too little effort exerted compared to the social optimum. Exceptions are few points in which full self-insurance remains desirable for the social planner and all agents remain passive.*

- *Protection:* Surprisingly, while protection is not even a Nash strategy we find that a social planner would elect an individual to exercise full protection effort.
- *Self-insurance:* Full self-insurance by every player is only desirable if protection costs are large. Therefore, for most cases the strategy of a social planner will not coincide with the only Nash equilibrium strategy.
- *Passivity:* In the Nash equilibrium agents are also too inactive. Passivity is highly undesirable from a social planner's perspective. Only if $NpL < b$ no agent will be selected to exercise maximum protection effort (while self-insurance might remain an option).

It is important to note that the social optimum variation that requires full protection by one individual results in the whole population being unharmed, since one highly secure individual is enough to thwart all attacks. Therefore, it is easy to see that protection is extremely desirable from a planners perspective. Out of the three classical public goods games with homogeneous agents the best shot game can benefit the most from a guiding hand.

6.4 Weakest target security game (without mitigation)

We compute the social optimum by using Eqn. (8), assuming that $1 \leq K \leq N$ players pick $e_0 = \min_i(e_i)$. Further assume, without loss of generality, that these players' indices are ranked from 1 to K. From Eqn. (6),

we get

$$\Phi = NM - b\sum_i e_i - c\sum_i s_i - pL\sum_{i=1}^{K}(1-s_i)\ .$$

For all i, we have

$$\frac{\partial \Phi}{\partial e_i} = -b\ ,$$

so, with $b > 0$, the social minimum is reached for a set of e_i's as small as possible. Further, consider the function Φ_{ins} where players $1, \ldots, K$ all pick an insurance level $s_0 = 1$ (all other values are chosen the same as in Φ). Then we get

$$\Phi_{\text{ins}} - \Phi = -c\sum_{i=1}^{K}(1-s_i) - pL\sum_{i=1}^{K} -(1-s_i)\ ,$$

that is, rearranging

$$\Phi_{\text{ins}} - \Phi = (pL-c)\sum_{i=1}^{K}(1-s_i)\ .$$

With $K \geq 1$ and a non-prohibitive insurance cost $c < pL$, we get

$$\Phi_{\text{ins}} \geq \Phi\ ,$$

for any other constellation of s_i. From what precedes, the K players picking the lowest security level e_0 should choose $e_0 = 0$, $s_i = 1$. The $N - K$ players above e_0 should pick as low as possible a positive security level, say $\varepsilon > 0$. Replacing in Φ, we get

$$\Phi = NM - b(N-K)\varepsilon - cK - c\sum_{i>K} s_i\ ,$$

which implies that for $i > K$, the social optimum satisfies $s_i = 0$, so that,

$$\Phi = NM - bN + K(b\varepsilon - c)\ ,$$

from which we conclude that $K = 1$ maximizes Φ.

To summarize, we find that in the weakest target game without mitigation a social planner would direct a single player to exacerbate no protection effort.

Essentially, this player serves as a direct target for a potential attacker. However, as long as $c < pL$ the player would be directed to maximize self-insurance $(e_i, s_i) = (0, 1)$. If insurance is too expensive ($c > pL$) then the social planner would prefer to leave the player uninsured $(e_i, s_i) = (0, 0)$. This strategy is independent of the cost of protection. The remaining $N - 1$ players have to select their protection effort as $e_i = \varepsilon > 0$ (as small as possible). These players will not be attacked, and therefore will set their self-insurance to the possible minimum $(\varepsilon, 0)$. Passivity by all players is never an option in the social optimum.

Result 18: *A social planner can easily devise a strategy to overcome the coordination problems observed in the Nash analysis for the weakest target game with mitigation. We found that no pure Nash strategy exists and, therefore, had to rely on the increased rationality requirement for entities to play a mixed strategy.*[5] *The average payoff for each player in the social optimum is considerably higher compared to the mixed Nash equilibrium.*

Understandably, without side-payments the node with the lowest protection effort is worse off compared to his peers. However, the social planner could choose to devise a so-called "honeypot" system with the sole goal of attracting the attacker while only suffering a marginal loss. A honeypot is a computer system (or another device) that is explicitly designed to attract and to be compromised by attackers. It serves usually a double purpose. First, it will detract attention from more valuable targets on the same network. Second, if carefully monitored it allows gathering of information about attacker strategies and behaviors, e.g., early warnings about new attack and exploitation trends [34].

An interesting aspect of the social optimum solution is the question how the individual is selected (if a honeypot system cannot be devised). Obviously, a social planner might be able to direct an individual to serve as a target (in particular, if $c < pL$). However, if insurance costs are large being a target requires an almost certain sacrifice (dependent on the value of p). In anthropology and economics there are several theories that relate to an individuals willingness to serve as a sacrificial lamb. Most prominently, altruism and heroism come to mind. Simon also introduced the concept of docility. This theory refers to an individual's willingness to be taught or to defer to the superior knowledge of others [45].

6.5 Weakest target security game (with mitigation)

We adopt the same strategy for finding Φ's maximum as in the unmitigated case – that is, summing Eqn. (6) over i, and then studying the variations of Φ over K, s_i and e_0.

We have here:

$$\Phi = NM - b\sum_i e_i - c\sum_i s_i - pL(1-e_0)\sum_{i=1}^{K}(1-s_i)\ .$$

Clearly the social optimum will satisfy $s_i = 0$ for $i > K$. Call Φ_0 the total utility obtained when $s_i = s_0$ for $i \leq K$, and compute the difference between Φ_0 and Φ:

$$\Phi_0 - \Phi = \sum_{i=1}^{K}(s_0 - s_i)(pL(1-e_0) - c)\ .$$

If $e_0 < 1 - \frac{c}{pL}$, then $\Phi_0 \geq \Phi$ if $s_0 \geq s_i$ for all i. In other word, an extremum is reached for $s_0 = 1$. The same derivation as in the infinite strength case yields: $e_0 = 0$, $s_0 = 1$, $K = 1$ and

$$\Phi = NM - c - b(N-1)\varepsilon\ ,$$

[5] Economists are generally cautious regarding the assumption that individuals can detect and adequately respond to mixed strategy play by opponents [42].

for an arbitrarily small $\varepsilon > 0$. Conversely if we choose $e_0 > 1 - \frac{c}{pL}$ then we have to choose $s_0 \leq s_i$, so that $s_0 = 0$, and an extremum is reached for $e_0 = 1$, $K = N$. We get

$$\Phi = NM - Nb\,.$$

Which of these two extrema is the larger one? Given that ε can be arbitrarily small, it basically depends on whether $Nb < c$ or not, that is, the social optimum depends on the relative cost of insurance versus protection.

From this derivation, the first observation is that the social planner might prescribe the same strategy as in the case of the weakest target game without mitigation. However, now the planner has a second alternative. Since an attacker will not be able to compromise players if they are fully protected we find that $(e_i, s_i) = (1, 0)$ for all N players is a feasible strategy. The tipping point between the two strategies is at $Nb < c$. If this condition holds the social planner would elect to protect all machines in favor of offering one node as honeypot and investing in its self-insurance. Note that again we find that if protection and self-insurance are extremely costly the planner will elect to sacrifice one entity without insurance. Passivity is not a preferable option.

Result 19: *Compared to the weakest target game without mitigation the social planner is better off if protection is cheap. Otherwise the planner has to sacrifice a node with or without self-insurance. Interestingly, while compared to the pure Nash equilibrium outcome the social planner can increase the overall utility in the network we find that security expenditures are lowered. In the Nash equilibrium agents were willing to fully protect against threats as long as* $(b \leq c)$.

The last observation also holds for the mixed strategy case in both weakest target games (with or without mitigation). That is, agents exert ***more*** *effort in the Nash equilibrium (except when* $Nb < c$ *for the game with mitigation).*

7 Discussion of results

The results we obtained, and notably the disconnect between social optima and Nash equilibria we observed, lead to a number of remarks that may prove relevant to organizational strategy. However, we want to preface this discussion by pointing out that our analysis is a first comparison of different security games with two security options under common, but restrictive assumptions.

Most notably, we assume agents to be risk-neutral providers of the public protection good. In our game formulation we also simplified cost of protection (and insurance) to be linear. Including different risk preferences, as well as uncertainty and limited information about important parameters of the game would be important steps towards a sensitivity analysis of our results. Shogren found, for example, that risk-averse agents will increase their contributions if information about other agents actions is suppressed [44]. Others, e.g., [39], have obtained more nuanced results. We defer a more extensive analysis of such phenomena

to future work, but believe that the main trends and differentiating features between security games we observed remain largely unchanged.

Security scenario identification: We find that security predictions vary widely between the five different games. Similarly, policies set by a social planner do not only yield different contribution levels but may also switch the recommended security action from protection to self-insurance and vice versa. Chief Security Officers' tasks involve a careful assessment of threat models the company is faced with.

We want to emphasize that an integral part of the threat model should be an assessment of the organizational structure including system resources and employees. Similarly important is a detailed consideration whether resources are protected independently or by an overarching system policy. For example, replication, redundancy and failover systems (that automatically switch to a standby database, server or network if the primary system fails or is temporarily shut down for servicing) should most likely not be treated as independent resources.

Managers should consider how the organizational structure of resources matches potentially existing policies. For example, we can see that a policy that requires full protection by every individual is sub-optimal if the most likely threat and organizational structure fits the description of a best shot game. Contributions resources are squandered and are likely to deteriorate. Not to mention that employees may simply ignore the policy over time. See, for example, recent survey results that highlight that 35% of white-collar employees admit to violations of security policies [28].

Security scenario selection: A security professional might be faced with a unidentifiable organization and system-policy structure. However, we want to highlight that our research allows a more careful choice between security options if managers can redesign organizations and policies. For example, the choice between a system-wide firewall and intrusion detection system versus an individual alternative has important implications on how incentives drive security-relevant behavior over time. Individual systems will better preserve incentives, however, might have negative cost implications. The same choice applies between the availability of backup tools and protective measures.

Leveraging strategic uncertainty: The example of the weakest-target game shows the importance of the degree of dependency between agents. We show that in larger organizations a much lower average level of self-insurance investments will be achieved because the strategic dependence between actors is reduced. However, in turn more agents will elect to protect their resources ($e_i > 0$ for more players). In contrast, agents in small groups will respond to the increasing strategic uncertainty caused by the increased interdependency by self-insuring their resources more often.

Introducing a social planner into the weakest-target game completely removes strategic uncertainty and leads to both reduced self-insurance and protection investments. This apparent paradox emphasizes that higher security investments do not necessarily translate in higher security – but instead that *how* the investments are made are crucial to the returns.

8 Conclusions

We model security decision-making by homogeneous and heterogeneous agents in a selection of five games. Some of these games have historical foundations in public good theory (weakest-link, best-shot, and total effort) whereas others were proposed recently (weakest target, with or without mitigation). Agents have two security actions at their disposal. They can contribute to a network-side protection pool or invest in a private good to limit losses.

At first we are considering homogeneous populations of users, where all participants have the same utility function. In practice, the homogeneity assumption is reasonable in a number of important cases, particularly when dealing with very large systems where a large majority of the population have the same aspirations. For instance, most Internet home users are expected to have vastly similar expectations and identical technological resources at their disposal; likewise, modern distributed systems, e.g., peer-to-peer or sensor networks generally treat their larger user base as equals.

However, the fact that the Internet is increasingly used as a common vector between different businesses, and even as a bridge between completely different user bases – for instance, acting as a bridge between mobile phone networks, home users, and e-commerce retailers, emphasizes the need for considering heterogeneous agents, even though the games considered may become far less tractable.

We find several key differences between the case of representative and diverse agents. For example, we found that in the total effort game stability increases with more pronounced heterogeneity in the agent population. The existence of a protection equilibrium in the weakest link game is threatened if only one agent prefers to self-insure or to remain passive. In the best shot game heterogeneous agents can overcome coordination problems more easily, so that a protection equilibrium is now possible, even though reaching this equilibrium grows increasingly unlikely with a larger number of agents participating in the network. Surprisingly, predictions for pure Nash equilibria of the weakest target games remain unchanged. However, mixed strategies do now have to take consideration of the heterogeneity of agents. We defer the computation of mixed strategies to future work.

When comparing the Nash and social optima for homogeneous agents we find that the effects of central planning compared to laissez-faire considerably differ according to the game considered. While in a number of traditional cases borrowed from the public good literature, we observe that a central planner may increase the average protection level of the network, we also note that strategic decisions are highly impacted by the level of inter-dependency between the actions of different players.

In particular, we found that the common wisdom that having a central planner who decides upon security implementation always yields higher protection contributions by individual players does not hold. Indeed, it may at times be much more advantageous from an economic standpoint to invest in self-insurance instead of protecting systems, or to select a few, unprotected, sacrificial lambs in order to divert the attention of potential attackers. This is particularly the case in situations which exhibit a "strategic uncertainty" due to a very strong correlation between the actions of different agents, for instance, in our weakest target game where the least secure player is always the one attacked.

8.1 Future research directions

First, we wish to extend our analysis to more formally explain the impact of limited information on agents strategies. In particular, in computer security and distributed networks the assumption of full information is useful as a first approximation but requires further validation. Similarly, we plan to analyze the robustness of our model by studying the influence of other simplifying assumption (e.g., linear cost parameters). Furthermore, we intend to evaluate strategy changes if moves are conducted sequentially rather than simultaneously. In the context of decision making on the Internet Friedman et al. also distinguish between synchronous and asynchronous moves [16].

Second, we are currently developing a set of laboratory experiments to conduct user studies and attempt to measure the differences between perfectly rational behavior and actual strategies played. Our preliminary investigations in the field notably evidence that players often experiment with different strategies to try to gain a better understanding of the game they are playing.

We are determined to incorporate our findings in updated models of system security. In prior work we challenged the assumption that all players are perfectly rational. In [11] we assumed agents to also accept strategies that are near rational and studied how system convergence prediction change.

Our research agenda of formal analysis combined with laboratory experiments is aimed to increase the understanding of individual and organizational security decision making. However, we are also interested in the design of meaningful security policies and aim at developing actionable guidelines for IT managers and other practitioners.

9 Acknowledgments

Paul Laskowski greatly improved this manuscript with his tremendously helpful feedback. We further appreciate the detailed evaluations of the anonymous reviewers at WWW08, EC'08, and WEIS 2008. This work is supported in part by the National Science Foundation under ITR award ANI-0331659. Jens Grossklags' research is also partially funded by TRUST (Team for Research in Ubiquitous Secure Technology), under support from the NSF (award CCF-0424422) and the following organizations: BT, Cisco, ESCHER, HP, IBM, iCAST, Intel, Microsoft, ORNL, Pirelli, Qualcomm, Sun, Symantec, Telecom Italia, United Technologies, and AFOSR (#FA 9550-06-1-0244).

References

[1] A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 3(1):26–33, January–February 2005.

[2] E. Adar and B. Huberman. Free riding on Gnutella. *First Monday*, 5(10), October 2000.

[3] R. Anderson. Why cryptosystems fail. In *Proceedings of ACM CCS'93*, pages 215–227, Fairfax, VA, November 1993.

[4] R. Anderson. Why information security is hard - an economic perspective. In *17th Applications Security conference (ACSAC)*, New Orleans, LA, December 2001.

[5] R. Anderson and T. Moore. The economics of information security. *Science*, 314(5799):610–613, October 1998.

[6] AOL/NSCA. Online safety study, October 2004. Available at: `http://www.security.iia.net.au/downloads/safety_study_v04.pdf`.

[7] T. August and T. Tunca. Network software security and user incentives. *Management Science*, 52(11):1703–1720, November 2006.

[8] Bruskin Research. Nearly one in four computer users have lost content to blackouts, viruses and hackers according to new national survey, 2001. Condensed results available at: `http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=iom&script=410&layout=-6&item_id=163653`.

[9] J.A. Bull, L. Gong, and K. Sollins. Towards security in an open systems federation. In *Proceedings of the Second European Symposium on Research in Computer Security (ESORICS), Springer LNCS No. 648*, pages 3–20, Toulouse, France, November 1992.

[10] P. Chen, G. Kataria, and R. Krishnan. On software diversification, correlated failures and risk management, April 2006. Available at SSRN: `http://ssrn.com/abstract=906481`.

[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, August 2004.

[12] D. Clark, J. Wroclawski, K. Sollins, and R. Braden. Tussle in cyberspace: defining tomorrow's Internet. In *Proceedings of ACM SIGCOMM'02*, pages 347–356, Pittsburgh, PA, August 2002.

[13] G. Danezis and R. Anderson. The economics of resisting censorship. *IEEE Security & Privacy*, 3(1):45–50, January–February 2005.

[14] I. Ehrlich and G.S. Becker. Market insurance, self-insurance, and self-protection. *Journal of Political Economy*, 80(4):623–648, July 1972.

[15] 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 14th ACM CCS*, Alexandria, VA, October/November 2007. Available from CMU at `http://www.cs.cmu.edu/~jfrankli/acmccs07/ccs07_franklin_eCrime.pdf`.

[16] E. Friedman, M. Shor, S. Shenker, and B. Sopher. An experiment on learning with limited information: nonconvergence, experimentation cascades, and the advantage of being slow. *Games and Economic Behavior*, 47(2):325–352, May 2004.

[17] D. Geer, C. Pfleeger, B. Schneier, J. Quarterman, P. Metzger, R. Bace, and P. Gutmann. Cyberinsecurity: The cost of monopoly. how the dominance of microsoft's products poses a risk to society, 2003. Available from Computer & Communications Industry Association at `http://www.ccianet.org/papers/cyberinsecurity.pdf`.

[18] L.A. Gordon and M. Loeb. The economics of information security investment. *ACM Transactions on Information and System Security*, 5(4):438–4572, November 2002.

[19] S. Gordon. The generic virus writer. In *Proceedings of the International Virus Bulletin Conference*, pages 121 – 138, Jersey, Channel Islands, 1994.

[20] S. Gordon. Virus writers - the end of the innocence? In *10th Annual Virus Bulletin Conference (VB2000)*, Orlando, FL, September 2000. Available from IBM Research at `http://www.research.ibm.com/antivirus/SciPapers/VB2000SG.htm`.

[21] 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 (WWW08)*, pages 209–218, Beijing, China, April 2008.

[22] 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)*, Chicago, IL, July 2008.

[23] G. Hardin. The tragedy of the commons. *Science*, 162(3859):1243–1248, December 1968.

[24] J. Hartley. Retrospectives: The origins of the representative agent. *The Journal of Economic Perspectives,*, 10(2):169–177, Spring 1996.

[25] K. Hausken. Returns to information security investment: The effect of alternative information security breach functions on optimal investment and sensitivity to vulnerability. *Information Systems Frontiers*, 8(5):338–349, December 2006.

[26] J. Hirshleifer. From weakest-link to best-shot: the voluntary provision of public goods. *Public Choice*, 41(3):371–386, January 1983.

[27] P. Honeyman, G.A. Schwartz, and A. van Assche. Interdependence of reliability and security. In *Workshop on Information Systems and Economics (WISE 2007)*, Pittsburgh, PA, June 2007.

[28] Information Systems Audit and Control Association. Telephone survey conducted by MARC Research, October 2007. Find information at `http://biz.yahoo.com/bw/071031/20071031005079.html?.v=1`.

[29] Q. Lv, S. Ratnasamy, and S. Shenker. Can heterogeneity make gnutella scalable? In *Proceedings of the First International Workshop on Peer-to-Peer Systems (IPTPS)*, Cambridge, MA, March 2002.

[30] S. Malphrus. The "I Love You" computer virus and the financial services industry, May 2000. Testimony before the Subcommittee on Financial Institutions of the Committee on Banking, Housing, and Urban Affairs, U.S. Senate. `http://www.federalreserve.gov/BoardDocs/testimony/2000/20000518.htm`.

[31] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver. Inside the Slammer worm. *IEEE Security and Privacy*, 1(4):33–39, July 2003.

[32] D. Moore, C. Shannon, and J. Brown. Code-Red: a case study on the spread and victims of an internet worm. In *Proceedings of 2nd ACM/USENIX Internet Measurement Worskop*, pages 273–284, Marseille, France, November 2002.

[33] A. O'Donnell and H. Sethu. On achieving software diversity for improved network security using distributed coloring algorithms. In *Proceedings of the 11th ACM Conference on Computer and Communications Security (CCS)*, pages 121–131, Washington DC, USA, October 2004.

[34] N. Provos. A virtual honeypot framework. In *Proceedings of the 13th USENIX Security Symposium*, pages 1–14, San Diego, CA, August 2004.

[35] N. Provos, D. McNamee, P. Mavrommatis, K. Wang, and N. Modadugu. The ghost in the browser: Analysis of web-based malware. In *Proceedings of the First USENIX Workshop on Hot Topics in Understanding Botnets (HotBots'07)*, cambridge, MA, April 2007.

[36] E. Rescorla. Security holes... who cares? In *Proceedings of the 12th USENIX Security Symposium*, pages 75–90, Washington, DC, August 2003.

[37] J. Saltzer, D. Reed, and D. Clark. End-to-end arguments in system design. *ACM Transactions on Computer Systems*, 2(4):277–288, November 1984.

[38] T. Sandler and K. Hartley. Economics of alliances: The lessons for collective action. *Journal of Economic Literature*, XXXIX(3):869–896, September 2001.

[39] T. Sandler, F. Sterbenz, and J. Posnett. Free riding and uncertainty. *Economic Review*, 31(8):1605–1617, December 1987.

[40] T.C. Schelling. *The Strategy of Conflict*. Oxford University Press, Oxford, UK, 1965.

[41] B. Schneier. *Applied Cryptography: Protocols, Algorithms, and Source Code in C (2nd edition)*. Wiley Computer Publishing, New York, NY, 2 edition, 1995.

[42] J. Shachat and J.T. Swarthout. Do we detect and exploit mixed strategy play by opponents? *Mathematical Methods of Operations Research*, 59(3):359–373, July 2004.

[43] S. Shenker. Making greed work in networks: A game-theoretic analysis of switch service disciplines. *IEEE/ACM Transactions on Networking*, 3(6):819–831, December 1995.

[44] J.F. Shogren. On increased risk and the voluntary provision of public goods. *Social Choice and Welfare*, 7(3):221–229, September 1990.

[45] H. Simon. Altruism and economics. *American Economic Review*, 83(2):156–161, May 1993.

[46] E. Skoudis. *Malware: Fighting malicious code*. Prentice Hall, Upper Saddle River, NJ, 2004.

[47] 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/`.

[48] J.B. Van Huyck, R.C. Battallio, and R.O. Beil. Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, 80(1):234–248, 1990.

[49] H.R. Varian. System reliability and free riding. In L.J. 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.

[50] N. Weaver and V. Paxson. A worst-case worm. In *Proceedings (online) of the Third Annual Workshop on Economics and Information Security (WEIS'04)*. Available at `http://www.dtc.umn.edu/weis2004/weaver.pdf`.

[51] L. Zhuang, J. D. Tygar, and R. Dhamija. Injecting heterogeneity through protocol randomization. *International Journal of Network Security*, 4(1):45–58, 2007.

A Nash equilbrium in the weakest target game with mitigation

Assume the existence of a Nash equilibrium where $0 < K < N$ "type I" players satisfy $e_i = \min(e_i, e_{-i}) = e_0$, while $(N - K > 0)$ players satisfy $e_i > e_0$ ("type II" players). Type-II players converge to $s_i = 0$ and $e_i = e_0 + \varepsilon$ with $\varepsilon > 0$ infinitesimally small, since they are not going to be attacked, and therefore satisfy

$$U_i = M - be_0 - b\varepsilon \ .$$

Type-I players, for their part, satisfy

$$U_i = M - pL(1 - s_i)(1 - e_0) - be_0 - cs_i \ .$$

Assume for now that potential losses L are "large enough" compared to insurance and security prices, namely $pL > c$, and $pL > b$. We have

$$\frac{\partial U_i}{\partial e_0} = pL(1 - s_i) - b \ ,$$

and

$$\frac{\partial U_i}{\partial s_i} = pL(1 - e_0) - c \ .$$

We have three possibilities:

- $s_i = 1 - b/pL$. Then, regardless of the value chosen for e_0, we have U_i constant and equal to

$$U_i = M - b - c + \frac{bc}{pL} \ .$$

- $s_i < 1 - b/pL$. Then, U_i increases in e_0, which leads to picking $e_0 = 1$. Likewise, U_i decreases in s_i, which leads to picking $s_i = 0$. We get

$$U_i = M - b \ .$$

- $s_i > 1 - b/pL$. Then, U_i decreases in e_0, which leads to picking $e_0 = 1$. Likewise, U_i decreases in s_i, which leads to picking $s_i = 0$. We obtain

$$U_i = M - c \ .$$

Because $c - \frac{bc}{pL} = bc\left(\frac{1}{b} - \frac{1}{pL}\right) = bc\frac{pL-b}{pLb} > 0$, we know the best type-I players can do is either $U_i = M - c$, or $U_i = M - b$, depending on the respective values of b and c. Again, we have to distinguish between three cases:

- $b < c$ (insurance is more expensive than protection). Then, type-I players pick $e_0 = 1$, $s_i = 0$, and $U_i = M - b$. This is a stable Nash equilibrium, and because $e_0 = 1$, all players are type-I players (i.e., everybody gets attacked, but doesn't really care because the attacker is not powerful enough to knock down the security protections).

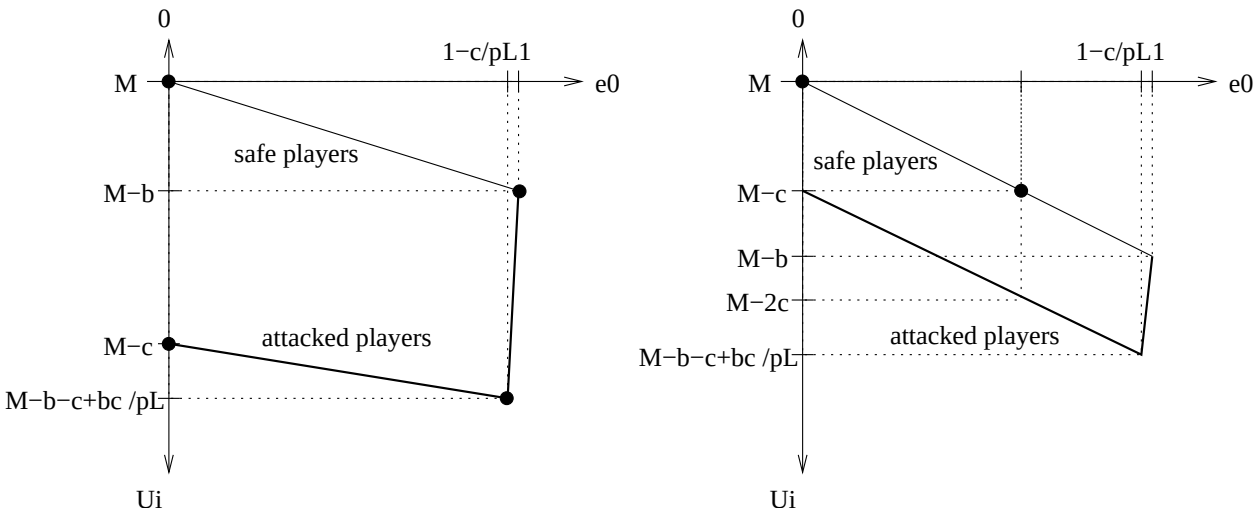

Figure 4: **Utility functions and Nash equilibria in the weakest target game.** The existence of Nash equilibria depends on the relative values of b and c.

- $b > c$ (protection is more expensive than insurance). Then, type-I players pick $e_0 = 0$, $s_i = 1$, and $U_i = M - c$. Type-II players are at $U_i = M - bh$ for any $h > 0$. This means that type-I players have an incentive to become type-II players by slightly increasing their protection, and discarding their insurance. Then, everybody switches back to being attacked but now with $e_0 = h$ and $s_i = 0$, which, since $\frac{\partial U_i}{\partial s_i}\Big|_{e_0=h} > 0$, immediately draws people to revert to $s_i = 1$. The utility U_i is then

$$U_i = M - bh - c\,,$$

which leads all players to again increase their security level by h, and shutting down insurance by setting $s_i = 0$. (The associated gain in utility is $c - bh$, which is much greater than the bh saved by reverting back to $e_0 = 0$ while keeping insurance.)

This escalation continues until everybody reaches $e_0 = 1 - \frac{c}{pL}$. Then, because $\frac{\partial U_i}{\partial s_i}\Big|_{e_0=1-\frac{c}{pL}} = 0$, people do not have an incentive to increase $s_i = 0$ at all. The utility of everybody, at that stage is

$$U_i = M - c - b + \frac{bc}{pL}\,.$$

Increasing e_0 infinitesimally leads to not be attacked (a saving of c), so people do that. Eventually, people keep upping their security levels until approaching $e_0 = c/b$, $s_i = 1$. (See Figure 4.) At that stage, all type-I (attacked) players have two equivalent choices: either set $e_0 = c/b$, $s_i = 0$ (not attacked), or decreasing e_0 to zero, with $s_i = 1$. The utility is indeed $M - c$ in both cases. If everybody adopts the same protection strategy, however, then everybody becomes a type-I player, and the utility goes down to $M - 2c$. Then, everybody has an incentive to have $e_0 = 0$, $s_i = 1$, but once

this is done, everybody has an incentive to defect, and set again $e_0 > 0$, $s_i = 0$. In short, there is a perpetual cycle - no Nash equilibrium in that configuration.

- $b = c$. Here again $e_0 = 1$, $s_i = 0$ is the only stable Nash equilibrium for the same reasons as stated above.

In summing, when the attacker is not omnipotent, people should have a strong incentive to pick high levels of protection and to not insure. The situation is markedly different from the previous case, where the power of the attacker essentially drove people to only insure, without protecting.

If protection is expensive, however, there is no stable Nash equilibrium.