Predicted and Observed User Behavior in the Weakest-Link Security Game

Jens Grossklags
UC Berkeley/School of Information
jensg@ischool.berkeley.edu

Nicolas Christin
CMU/CyLab Japan
nicolasc@cmu.edu

John Chuang
UC Berkeley/School of Information
chuang@ischool.berkeley.edu

Abstract

We aim to advance the understanding of individual security decision-making, by combining formal and behavioral analysis. We sketch a game-theoretic model of security decision-making that generalizes the "weakest link" game, and describe a controlled laboratory experiment to reveal differences between predicted and observed user behavior. Results of a pilot study yield possible explanations for behaviors observed in the wild: users show some willingness to experiment with parameters, rarely converge to a fixed behavior, and face difficulties isolating the impact of individual parameters.

1 Introduction

Collective awareness of the need for information security has considerably risen in the past few years. Yet, user behavior toward security in networked information systems remains obfuscated by complexity, and may seem hard to rationalize [1]. This observation is perhaps not surprising, considering that users' security decisions appear often in contradiction with their own stated attitudes and desires. Indeed, when asked in surveys, computer users say they are interested in preventing attacks and mitigating the damages from computer and information security breaches [1]. On the other hand, studies (e.g., [3, 9]) evidence the low levels of privacy and security precautions taken by a vast majority of users.

Academic research has isolated the misalignment of user incentives as a root cause for the observed dichotomy between behavior and stated attitudes and low overall system security. Specifically, users often fail to internalize the impact of their own decisions on 1) the motivation of other users to secure; and on 2) the level of security that can be achieved in an interconnected network [2, 6].

This paper proposes to explore the relationship between economic and psychological-behavioral incentives for improved or declining system security. We aim to enhance the understanding of the often puzzling and frustrating security and privacy decision-making of individuals and groups to advance institutional and policy responses.

As a first step, we focus here on a weakest-link security scenario. For instance, consider a single unpatched system being infected by a worm, which, as a result, allows an attacker to assail (e.g., deny service to) the rest of the network without facing additional defenses. The incentives other users have to protect their own systems against security breaches are weakened, since irrespective of their individual decisions, they still suffer the consequences induced by the existence of the infected host.

The weakest-link scenario has been formalized as an economic game [7, 8] and applied to system security and reliability [10]. We generalize the model by allowing users to not only protect their resources (e.g., by installing a firewall), but also to self-insure against the consequences of security breaches (e.g., by backing up valuable data) [6].

Applying game-theoretical analysis, such as computation of Nash equilibria, requires several assumptions regarding user behavior. All users are supposed to have all information about potential actions and associated payoffs available, to infallibly select actions that are profit-maximizing, and to perfectly take into consideration other users' optimizing decisions [4]. In the context of security games, these assumptions can be challenged, and individuals frequently deviate from perfect rationality, and do not always select purely selfish actions.

In an effort to better understand the origin of these departures from optimality, we have started to systematically study the actual behavior of individuals in the weakest-link security game. We contrast the behavior predicted by game-theoretic analysis with the behavior we observe in a a pilot study of a controlled laboratory experiment. Participants have to select protection and recovery strategies in an environment with non-deterministic attacks and a lack of information about parameters of the game.

2 Model and theoretical predictions

Formal description. The game is played amongst N players, denoted by $i \in (1, \ldots, N)$ who are all potential victims of a security threat. The attacker is exogenous to the game. The game consists of multiple rounds. In each round, security attacks occur probabilistically, according to a fixed, exogenous probability p, which determines the baseline rate of security attacks. For instance, $p = 0.5$ means that the attacker attempts to compromise the network half of the time, and is idle the rest of the time. All attempted compromises are not necessarily successful, however; the success or failure of an attempted attack depends on the defensive measures put in place by the players, as we discuss below. Successful security compromises will result in *all* players incurring losses. Indeed, in a weakest-link game, compromising one player (the so-called weakest-link) means compromising the whole network. However, the extent of each player's losses, L_i, is dependent on the choices made by that player.

More precisely, each player i receives an endowment $M_i > 0$ in each round. The endowment can be utilized for two security actions: self-protection ($0 \leq e_i \leq 1$) and self-insurance ($0 \leq s_i \leq 1$) with linear associated (positive) effort costs b_i and c_i, respectively. Self-protection acts to probabilistically block attacks. Self-insurance deterministically lowers the penalty occurred during an attack that is not blocked. So, each player can face three states in each round: 1) no attack occurs, 2) an attack takes place but it is blocked due to self-protection (and self-protection of all other players), and 3) an attack happens and it is not blocked. Self-protection influences the likelihood of occurrence of states 2 and 3. Self-insurance lowers the hurt in state 3. Formally, we express the payoff π_i to a player i as:

$$\pi_i = M_i - pL_i(1-s_i)(1-\min[e_i, e_{-i}]) - b_i e_i - c_i s_i \,, \tag{1}$$

where e_{-i} denotes the set of protection levels picked by players other than i.

In practice, many security mechanisms combine features of both protection and insurance as we have defined them. For example, a spam filter might not automatically block all dubious messages but redirect some of them into a special folder. Reviewing these messages separately from the main mailbox will often result in lower cognitive cost to the individual. Hence, the spam filter lowers both the probability of spam and the magnitude of the associated cost if spam passes the filter. Even though protection and insurance are often intertwined in practice, for the purposes of this research study we chose to clearly separate both actions in order to be able to dissociate their effects.

Nash equilibrium. Considering the payoff function in Eq. 1, if we assume symmetric costs homogeneous across users, that is, if for all i, $b_i = c_i = b = c$, and $L_i = L$, the weakest-link security

game has two Nash equilibria that are not Pareto-ranked: Either all individuals protect with a certain effort but neglect insurance (*protection equilibrium*), or everybody fully insures and decides not to protect (*insurance equilibrium*). Both Nash equilibria yield identical payoffs. We sketch the proof in Appendix A, where we also show that these results extend to the asymmetric case where individual players face different b_i, c_i and L_i.

3 Focus of experimental observation

We next contrast the analytic findings outlined above with actual player behaviors evidenced through preliminary laboratory experiments. At a high level, we want to focus on equilibrium selection and learning behavior of players. We aim in particular to obtain answers to the following set of questions.

Will the game converge to a Nash equilibrium outcome? Prior experimentation centered on non-continuous and non-probabilistic versions of the weakest-link game with or without limited information about important parameters of the game [8]. Data shows that experiments usually converge and individuals are able to tacitly coordinate on a Nash equilibrium outcome. Disagreements between players usually disappear quickly with all players focusing on one Nash strategy.

Does the self-insurance equilibrium dominate other outcomes? Because the protection equilibrium is sensitive to defection by even a single player, we expect this equilibrium to be observed less frequently, in particular as the group size increases [6]. From a behavioral perspective, however, we would expect individuals to at first attempt to protect their resources.

Is experimentation a prominent part of players' strategies? We expect that the limited information environment of the game stimulates players to experiment. Similar to the findings of [5], we suggest that players in security games will systematically probe the action space. In contrast to [8], participants in our experiments are unaware of the type of security situation they are facing, i.e., they do not know that it is a weakest-link game, creating a further incentive for experimentation.

4 Experimental observations

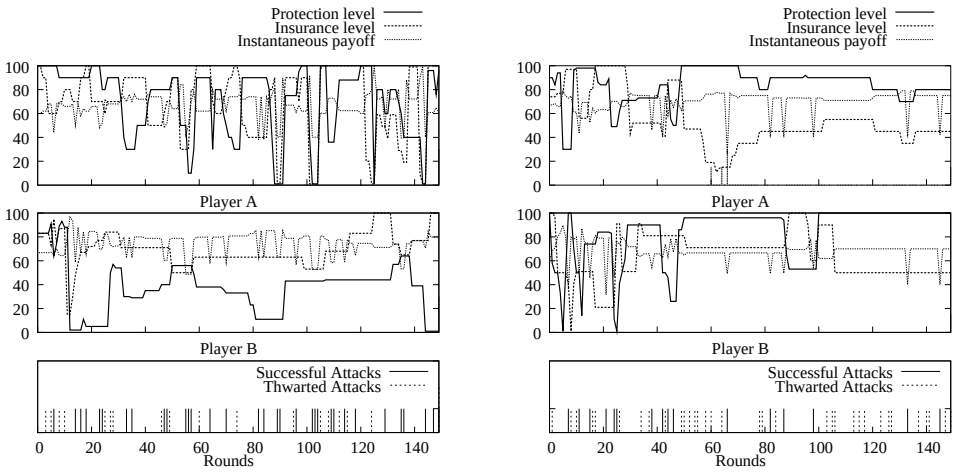
Setup. We recruit participants from a student subject pool at UC Berkeley, and have them participate in the experiment in a special computer laboratory, isolated from each other by separation walls. After reading and signing consent forms, they receive instructions for the experiment that set the context, explain the two main user actions (self-protection and self-insurance) and introduce the user interface.

The user interface provides individuals with two slider-type input devices that allow them to modify their security settings. Feedback is given both numerically and in graphical panels to help users recognize patterns and trends in the data.

The experiment proceeds continuously, without pause between payoff rounds. The length of a round is 5 seconds. The whole experiment lasts 150 rounds. The average attack probability is $p = 0.33$, i.e., attacks occur about every 3 rounds. Protection and insurance costs are symmetric ($b = c$).

To capture the low-information feature of security decisions, participants do not receive any specific information about the structure of the model or its parametrization. We do not inform them about the number of players in their group, other players' actions, or payoffs. Participants, however, do receive feedback on the attack state of the past round.

Results. We describe the outcomes of two 2-player games, and one 3-player game. Fig. 1 displays the data for two 2-player games. All four players exper-



(a) First game (b) Second game

Figure 1: **Experimental data for two-player games.**

iment with parameter settings throughout the whole duration of the session.

In the first game (Fig. 1(a)) both players follow different approaches in search of rewarding strategies. Player A's strategy selection resembles a heartbeat pattern with values kept for less than 5 periods. Protection and insurance levels are often modified in unison. Player B, on the other hand, usually keeps one parameter fixed while modifying the other. Furthermore, Player B keeps settings constant for longer periods on average. This first 2-player game does not converge to a Nash equilibrium.

The players in the second 2-player game (Fig. 1(b)) follow a different experimentation and convergence pattern. After an initial phase of exploration with relatively sudden, and sometimes extreme, changes in both parameters, the two players settle on a more moderate pattern of change. Both players converge to settings with high protection efforts. Surprisingly, even though few attacks beyond round 50 are successful, both players keep up a relatively high insurance effort.

We also provide data for a 3-player game (Fig. 2). Most remarkable is the strategy play by Player B, who quickly settles on a low protection and high insurance strategy with little experimentation. At round 65 we observe a short-termed complete reversal of this strategy for experimentation, during which the subject suffers from one security compromise that might be the cause for the quick return to the prior strategy. Player C experiments thoroughly with parameter settings that pit protection and security against each other. For most of the game beyond round 50 player C plays close to the individually rational strategy to insure and not protect. Player B selects a lower insurance level but approximately follows the same strategy from round 30 on. Surprisingly, player A never adapts, even though at least one player selects low protection settings from round 30 until the end of the game.

5 Conclusions

We consider the pilot experiments presented here as preparation for a larger study that compares the economic determinants of different organizational structures and attack patterns analytically and experimentally. We also plan to conduct experiments with different modes of intervention to improve convergence to a desirable equilibrium.

The initial results we report here suggest that the weakest-link security game has several properties that distinguish it from the classical weakest-link

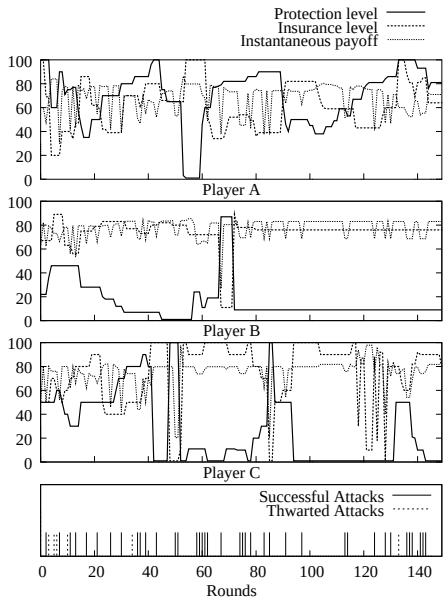


Figure 2: **Three-player game.**

game (with non-probabilistic payoffs and without self-insurance). First, individuals experiment frequently and often thoroughly. Second, convergence to a Nash equilibrium is not achieved within a few periods. In fact, in the data gathered so far, we do not observe convergence to any of the predicted equilibria at all. Given that each game lasted 150 rounds (i.e., 12.5 mins), this result is surprising.

We find initial evidence that the individual approach to experimentation has a distinct impact on whether a player will find an individually rational strategy and the game will converge to a Nash equilibrium. Our results further evidence that some players hesitate to try strategies that require them to decouple the protection and self-insurance parameters from each other.

We contribute to a better understanding of the psychology of security decision-making, by providing economic models that capture important aspects of organizational structure and add a so far overlooked aspect of decision complexity, i.e., the difference between protection and self-insurance [6]. Our experiments aim to uncover how well individuals can follow economic incentives, and where complexity impedes the realization of good security outcomes.

Acknowledgments

We thank the anonymous reviewers and Paul Laskowski for their valuable comments and editorial guidance, and Neal Fultz for his contributions to the software used in the experiments. This work is supported in part by the National Science Foundation under awards ANI-0331659 and CCF-0424422, and by the Air Force Office for Scientific Research under award #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] R. Anderson. Why information security is hard - an economic perspective. In *Proc. ACSAC'01*, New Orleans, LA, Dec. 2001.

[3] AOL/NSCA. Online safety study, December 2005. Available at: `http://www.staysafeonline.info/pdf/safety_study_2005.pdf`.

[4] N. Christin, J. Grossklags, and J. Chuang. Near rationality and competitive equilibria in networked systems. In *Proc. ACM SIGCOMM'04 PINS Workshop*, pages 213–219, Portland, OR, August 2004.

[5] 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.

[6] J. Grossklags, N. Christin, and J. Chuang. Secure or insure? A game-theoretic analysis of information security games. In *Proc. WWW'08*, Beijing, China, April 2008.

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

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

[9] S. Spiekermann, J. Grossklags, and B. Berendt. E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior. In *Proc. ACM EC'01*, pages 38–47, Tampa, FL, October 2001.

[10] 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.

A Appendix: Game-theoretical analysis

A Nash equilibrium is a "best response" equilibrium, where each player i picks the pair (e_i, s_i) which maximizes his/her own payoff, given the set of values (e_{-i}, s_{-i}) chosen by all other players. Nash equilibria are expected to be observed under the assumption that all players are perfectly rational, and know all strategies available to all players, as well as the associated payoffs.

Let us assume that all players have identical parameters (i.e., for all i, $M_i = M$, $L_i = L$, $b_i = b$, and $c_i = c$). If we denote by $\hat{e_0}$ the minimum of the protection levels initially chosen by all players, a study of the variations of π_i as a function of e_i and s_i yields that three types of equilibria exist [6].

First, a protection equilibrium $(e_i, s_i) = (\hat{e}_0, 0)$ occurs, when $pL > b$ and either 1) $pL < c$ or 2) $pL \geq c$ and $\hat{e}_0 > (pL-c)/(pL-b)$. That is, everybody picks the same minimal security level, and no one has any incentive to lower it further down. This equilibrium can only exist for low protection costs ($b \leq c$), and may be inefficient, as it could be in the best interest of all parties to converge to $e_i = 1$, to have a higher chance of deflecting incoming attacks [6]. This protection equilibrium depends on the cooperation of all players, and is therefore very unstable. It requires only a remote possibility that any of the $n-1$ players will not select the full-protection equilibrium for the remaining player to defect. In the second type of equilibrium, all players will self-insure themselves completely ($e_i = 0$ and $s_i = 1\ \forall i$), when $pL > c$ and either 1) $pL < b$ or 2) $pL \geq b$ and $\hat{e}_0 < (pL - c)/(pL - b)$. Essentially, if the system is not initially secured well enough (by having all parties above a fixed level), players prefer to self-insure. The effectiveness of this security measure does not depend on the cooperation of other players. A third, trivial equilibrium, is a passivity equilibrium, where all players choose $(e_i, s_i) = (0, 0)$, when $pL < b$ and $pL < c$.

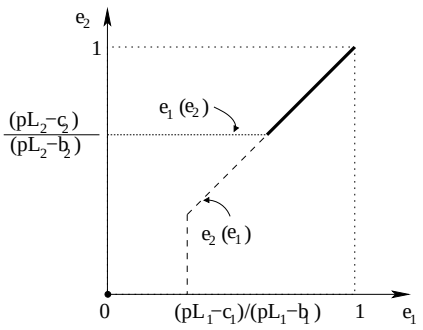


Figure 3: **Reaction functions for a two-player weakest-link game**. Bold lines and dots indicate potential Nash equilibria.

We can extend the presentation to an asymmetric case, where different players have different valuations L_i, c_i, and b_i. For simplicity, consider first a two-player game. By definition, Nash equilibria are characterized by the reaction functions $e_2(e_1)$ and $e_1(e_2)$ reaching a fixed point, that is $e_2(e_1) = e_1(e_2)$. Indeed, the effects of self-insurance on the payoffs received is independent of the other player's actions, and is therefore not a factor here. In Fig. 3, we see that a fixed-point is attained when $e_1 = e_2 = 0$ (self-insurance-only equilibria, as discussed before), and when both e_1 and e_2 are greater than $\max\{(pL_1 - c_1)/(pL_1 - b_1), (pL_2 - c_2)/(pL_2 - b_2)\}$.

Generalizing to N players, we obtain the following distinction: if, for all i, $pL_i > b_i$, and either 1) $pL_i < c_i$, or 2) $pL_i \geq c_i$ and $\hat{e_0}$, the minimum initial protection level, is greater than $\max_{1 \leq i \leq N}\{(pL_i - c_i)/(pL_i - b_i)\}$, then we have a Nash equilibrium where everyone picks $(\hat{e_0}, 0)$. Otherwise, all players select $e_i = 0$. The value of self-insurance they select depends on their respective valuations. Players for whom insurance is too expensive ($pL_i < c_i$) do not insure, with $s_i = 0$, while others choose full self-insurance, that is $s_i = 1$. This result extends observations made by Varian [10] in the absence of self-insurance strategies.