SAND2019-8062C

# Can social influence be exploited to compromise security: An online experimental evaluation

Soumajyoti Sarkar, Paulo Shakarian
Arizona State University, Tempe, USA
Email: {ssarka18, shak}@asu.edu

Mika Armenta, Danielle Sanchez, Kiran Lakkaraju
Sandia National Laboratory, New Mexico, USA
Email: {mlarmen, dnsanc, klakkar}@sandia.gov

***Abstract*—While social media enables users and organizations to obtain useful information about technology like software and security feature usage, it can also allow an adversary to exploit users by obtaining information from them or influencing them towards injurious decisions. Prior research indicates that security technology choices are subject to social influence and that these decisions are often influenced by the peer decisions and number of peers in a user's network. In this study we investigated whether peer influence dictates users' decisions by manipulating social signals from peers in an online, controlled experiment. Human participants recruited from Amazon Mechanical Turk played a multi-round game in which they selected a security technology from among six of differing utilities. We observe that at the end of the game, a strategy to expose users to high quantity of peer signals reflecting suboptimal choices, in the later stages of the game successfully influences users to deviate from the optimal security technology. This strategy influences almost 1.5 times the number of users with respect to the strategy where users receive constant low quantity of similar peer signals in all rounds of the game.**

## I. Introduction

Social influence is key to technology adoption, and research on the role of persuasion in security technology adoption indicates that various social influence factors impact a user's decision when making decisions to purchase or use a given technology [1]. However, these studies have primarily investigated the role of benign social influence and not how it can be harnessed to harm users, e.g. by cyber-adversaries. Specifically, social influence has primarily been studied in the context of it having a net positive impact on society [2], especially when considering the utility of the decisions made through influence.

Given the slew of recent events in which cyber warriors exploit social media with malicious intent, researchers and policy-makers are reconsidering the role of social influence as a tool for change. Consider the example of the experiment where an American security firm created fake Facebook accounts of a fictitious user in order to entice users to befriend her and inappropriately share information [3]. The study showed that transitive trust influenced users to make connections with her and in some cases even share sensitive geo-location information. It is especially alarming that users did not verify the account. These studies prompt the questions: do users place too much trust on peers and too little on their own knowledge when adopting security technologies, and what are the mechanisms that enable cyber-adversaries to influence users into adopting technologies that might be less secure or could be used for hacking into their systems?

To address these questions, we conducted an online, controlled experiment in which human participants played a multi-round game where they selected one security technology from among six of differing utilities. We modulated the number and temporal pattern of concurrent social signals that participants received from others who were also ostensibly playing. An example of the Linear Cascade, one of the six patterns of peer influence, is shown in Figure 1. We investigate three main questions: (1) Can social signals cause users to deviate from the optimal security technology? (2) Does social influence encourage option exploration when users are already aware of the most optimal technology? (3) Does the role of social influence factor more than other cognitive aspects that might impact the choices made by the users?

We observe that while early exposure to higher social signals leads to suboptimal decision-making, the effect disappears following the user's exploration period., and that delayed exposure to more social signals leads to suboptimal decision-making at the end of the game. Additionally, we find that social signals are more predictive of adoption choices than other factors such as the number of options explored by user or the number of alternating switches made. Our research opens new avenues for considering influence as a tool for exploiting security technology usage among users in social media.

## II. Methods

We ran an online, controlled decision-making game hosted by the Controlled Large Online Social Experimentation (CLOSE) platform and developed at Sandia National Laboratories [12], in which participants took on the role of a security officer at a bank. Participants were told that they and several of their peers at different banks were being asked to invest (or make a selection) in a cyber-defense provider for 18 rounds, once for each round. All participants could view brief descriptions of provider capabilities – e.g. one of them being "Secure.com utilizes algorithmic computer threat detection to keep systems safe. It prides itself on its efficiency and success rate in warding against attacks." Participants were able to choose from 6 different providers - among which only one was optimal. The optimal choice prevented 7 attacks and the remaining 5 providers each prevented 6. Thus, all suboptimal technologies had the same utility. This information

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

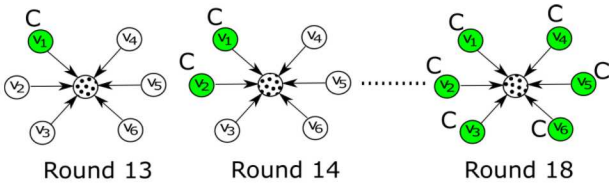



Fig. 1: Illustration of the *linear cascade* diffusion. The suboptimal technology $C$ chosen by us for influencing the user (in dots) cascades through the peers of the user over the last 6 rounds. Colored nodes denote the activated peers w.r.t. $C$ (manually preprogrammed by us) at each round. Note that although at Rounds starting at 13 and ending at 18, there are subjects (uncolored) among peers who have not adopted $C$, their decisions (technologies adopted which may not be $C$) at those rounds are visible to the subject in consideration (in dots). However, which users among the peers have been preprogrammed manually is by default unknown to the target subject. This linear cascade pattern is represented by $V = \{1, 2, 3, 4, 5, 6\}$, where the indices in the list denote the sequence of 6 rounds in an ordered manner.

about the optimal and suboptimal providers is not available to the participants in every group at the start of the game. An example of the screen is shown in Appendix[1]. We separated participants into 5 groups based on pattern of social signal exposure which will be described in the Design subsection following this. For each group, we controlled the number of signals which were the technology purchase decisions of others ostensibly playing the game at the same time.

The entire game was partitioned into two phases. For the first 12 rounds, no other information but a short excerpt about the six potential providers was given. After the participants made one selection (among the 6 security technologies available) for a given month, they saw the number of attacks their provider had prevented in the corresponding period. For every attack they prevented, participants received \$0.02. Thus, they were incentivized to avoid more attacks and earn more money. However, since the participants have to explore the technologies to first acquire the knowledge of the technology utilities, the first 12 rounds allow for individual decision making and exploration in the absence of any external knowledge about the technologies or peers. In the second phase of the experiment which started at Round 13, we introduced social influence by allowing participants to see their peers' decisions after every round, and by varying the temporal pattern by which concentric decisions among the peers. We attempt to avoid network effects by using pre-programmed bots (these are the peers that the users see in their screen) and holding the network structure constant. All participants have 6 peers, and participants can view their choices. An example of the network structure is shown in Figure 1, where a participant receives social signals from its six neighbors about a technology, $C$, among the 5 suboptimal security technology providers. The concentric decision from peers is always a suboptimal technology provider. We call this $C$ the *influence decision* for the respective user. The motivation behind this deliberate selection of suboptimal $C$ as the peer choice is to investigate whether participants would be tempted to select the suboptimal choice in the presence of social signals and when they already have acquired knowledge about the technologies from the first 12 rounds. We describe the signal patterns and conditions unique to each study in the following $Design$ section.

**Design.** We recruited a total of 357 participants to play the cyber-defense provider game. Participants were paid \$2 with the opportunity to earn up to \$4.52 because they received a bonus of \$0.02 for every attack prevented. Thus, the participants have a motivation to prevent more attacks in order to earn more money. In a between-subjects design, participants were randomly assigned to five groups: No Message (NM), Uniform Message (UM), Linear Cascade (LC), Early Cascade (EC), or Delayed Cascade (DC). For all groups, decisions in the first 12 rounds are made without any peer signals. Let $V(t)$ ($t$ denotes a round among the last 6 rounds) denote the number of peers of a user who at time $t$ were programmed to select a chosen suboptimal technology by us.

For the last six rounds participants in all groups except the No Message group, receive signals in the following sequence which we denote as the *patterns of influence*: (1) *No Message (NM)*: V = $\{0, 0, 0, 0, 0, 0\}$, (2) *Uniform Message (UM)* : V = $\{1, 1, 1, 1, 1, 1\}$, (3) *Linear Cascade (LC)*: V = $\{1, 2, 3, 4, 5, 6\}$, (4) *Early Cascade (EC)*: V = $\{4, 5, 6, 6, 6\}$ and, (5) *Delayed Cascade (DC)*: V = $\{1, 1, 1, 4, 5, 6\}$. Figure 1 shows the $V(t)$ = $\{1, 2, 3, 4, 5, 6, \}$ pattern of influence. (Figure 2 in the Appendix shows the signal patterns for the groups). In all conditions, users can switch back to any choice in the next round after having selected an option in the current round. The NM and UM groups are control groups where we control the peer signals in a way such that they do not have any variations in the pattern over the rounds and the suboptimal security technology $C$ chosen by us does not cascade through majority ($>= 4$ out of 6) peers at any round. On the other hand the LC, EC, DC groups are our treatments groups of interest where the technology $C$ cascades through majority peers in at least one out of the 6 rounds. Note that while the pattern remains the same, the suboptimal technology chosen by peers is randomized between and within participants, which allows us to eliminate effects of a specific technology or its description. An example of the Linear Cascade (LC) pattern is shown in Figure 1, where the subject (marked in dots) is able to receive social signals from its 6 neighbors. We emphasize that all the peers of the subjects are bots and do not share any topology between themselves (the treatments are therefore i.i.d., a treatment on one user does not produce rippling effect on other users), thereby we prevent the effects of network on the individual behavior changes.

## III. Analysis

### A. Distribution of attacks prevented

For all the research questions we investigate, the outcomes of interest are the decisions made by participants in the last

[1]Due to space limitation Appendix is uploaded online: Link

six rounds, in the presence of social signals from peers. Table I shows the distribution of attacks prevented by subjects in each group. We observe that, on average subjects in the EC and LC groups prevent fewer attacks compared to others. Based on a survey analysis, we found that none of the traits like computer anxiety, computer confidence, computer liking, intuition or neuroticism were correlated to the number of attacks prevented [4]–[6].

| Group | # participants | Average number of attacks prevented |
|---|---|---|
| **No Message** | 55 | 105.2 |
| **Uniform Message** | 71 | 106.28 |
| **Linear Cascade** | 79 | 103.81 |
| **Delayed Cascade** | 81 | 104.8 |
| **Early Cascade** | 71 | 103.83 |

TABLE I: Average number of attacks prevented by subjects in each group. The lower attacks suggest participants deviated more from the optimal decision responding to social influence.

*RQ1. Will participants deviate from the optimal security technology and move towards their peer suboptimal choice in the presence of social signals?*
Under this research question, we investigate two components. First, we measure the proportional of individuals in the last round (Round 18) who do not opt for the optimal decision and the proportion who settle on their respective *influence decision*. Next, we obtain these metrics for each subject in each group at the round when the influence is reflected in majority of the peers of that user. That is to say, the first round when 4 out of 6 peers of a user adopt the influence decision ($> 50\%$ peers) - this happens at round 13 for EC and at round 16 for LC and DC groups.

| **At the last round (Round 18)** | **NM** | **UM** | **LC** | **DC** | **EC** |
|---|---|---|---|---|---|
| Proportional of individuals not on optimal | 45.61 | 30.55 | 48.10 | **49.41** | 43.83 |
| Proportional of individuals on influence decision | 8.77 | 18.05 | 20.25 | **22.5** | 19.17 |

TABLE II: Comparison of the five groups w.r.t. the decisions made in the last round.

For the first component, Table II shows the proportion of users in each group for the two metrics discussed above - the results show that the DC group participants deviated most from the optimal with 49.41% of users settling on suboptimal choices in the last round. Among these, around 22.5% of DC participants switched to their influence decision (which is different for each participant) which is also the maximum among all the groups. In fact, while these results shed light on the retention power of the influence patterns - while the DC group's power of retention could be attributed to late exposure to larger quantity of peer signals, the EC group fails to retain many of the users after the initial rounds. In fact when compared to the UM group users (30.55%, the DC pattern of influence successfully influence almost 1.5 times the number of users to deviate from optimal decision (49.41%).

| **First round where majority peers reflect influence decision** | **LC (Round 16)** | **DC (Round 16)** | **EC (Round 13)** |
|---|---|---|---|
| Proportional of individuals not on optimal | 46.83 | 54.11 | **57.49** |
| Proportional of individuals on influence decision | 16.45 | 24.70 | **30.13** |

TABLE III: Comparison of the decisions made by the 3 treatment groups when majority peers reflect influence decision.

For the second component, Table III shows that on the contrary, the EC pattern of influence is able to draw more participants at the round where the participant clearly observes its influence decision as the one that majority of its peers ($geq4$ out of 6) choose. This suggests that while early subjugation to exposures demonstrates a better proxy for social influence, the exploration time following this early exposures motivates users to move away from this decision and so we see a substantial drop in the values for the last round for EC (compared to DC and LC) from table II.

*RQ2. Does the presence of social signals influence users to explore and revisit different security technologies?*
To further measure variations in decisions, we try to analyze the effect of the pattern of influence on decision explorations by users. The goal is to understand whether the introduction of peer signals prompts users to explore more options even in the presence of already acquired knowledge of the security technologies. We measure this through *entropy* as a means to analyze tendency towards fluctuations in decision making by users [7]. Formally, given a list of decisions $X$ made by each subject over the last 6 rounds, we define entropy as $H(X) = -\sum_{i=1}^{n} P(X_i) \log_b P(X_i)$ where $b$ is the base of the logarithm and $n$ denotes the number of possible decisions, which is 6 in our case. We use $b = 6$ to normalize the entropy values to be between 0 and 1 – we note that this is an artifact of the experiment design as there are 6 decisions types. Intuitively, given decisions made by two subjects, the participant with higher entropy value has changed its decisions more frequently compared to a participant with lower entropy value. Fig. 2 shows the entropy of the decisions for the second phase of the experiment. We performed the Kolmogorov-Smirnov (KS) tests between pairwise distributions and we found that with respect to the UM and NM participants (control), the distributions showed statistically significant differences with the LC, EC and DC participants ($p < 0.05$ for all these pairwise tests) considering the different technologies that were explored. Particularly, we find that for the groups NM and UM, the entropy distributions peak near 0, which explains the fact that users do not explore much under the control setup whereas the peak for EC group is evident at around 0.3 suggesting that more users explore options in this group. We conclude from this basic analysis that the introduction of peer influence in the form of

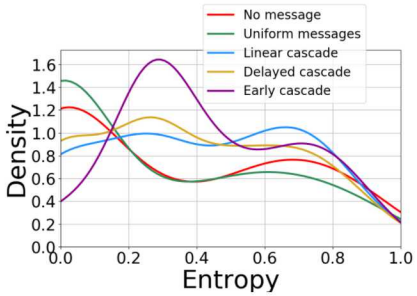



Fig. 2: Entropy distributions for the second phase (Rounds 13 to 18).

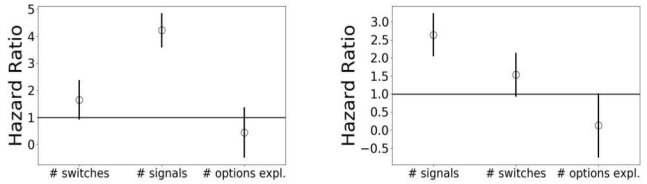
treatment patterns of influence does indeed prompt users to explore more, but the more subtle question here is whether the exploration differs among the LC, EC and DC groups. To this end, we find that from the same pairwise KS tests, there is no statistically significant difference between these 3 groups when considered in pairs. However, while most of these distributions are multi-modal (having multiple peaks), the LC group tends to have more users having higher entropy shown by the observation that one of its modes lie near entropy value of 0.6.

*RQ3. Do peer signals factor more than other cognitive aspects that might impact users like the number of switches made by the user so far?*

We use Cox proportional hazards model, which is the standard technique for assessing contagion in economics, marketing, and sociology [11]. This tool measures the hazard or likelihood of adoption of an individual at time $t$ as a function of individual characteristics and social influence: $\lambda(t, X_{ti})$ = $\lambda_{0t} exp(X_{ti}\beta)$ where $\lambda$ represents the hazard of adoption for a subject after the $t^{th}$ round ($t \in Rounds[13, 18]$), $\lambda_{0t}$ represents the baseline hazard of adoption and $X_{ti}$ represents the static set of covariates for subject $i$ after round $t$ - namely the number of signals reflecting *influence decision*, the number of decision switches made by the participant at $t$ (we define switch at round $t$ as being made by a user if decision at $t$ differs from $t-1$) and the number of technology options (among the 6 possible) explored by the participant at $t$. Figure 3 displays the results for the DC and EC group participants and it shows that when the 3 factors are considered together for the hazard of adoption, the number of signals is the only significant factor playing into the adoption of the security features - the 95% CI lies above 1 (which denotes significance) for both the DC and EC participants. However, we find that none of the factors were significant for the LC group participants - these together suggest that the pattern of influence in the EC and DC groups were more effective than the implicit aspects like the number of decisions the user explored or the number of switches it made.

## IV. Related Work and Conclusions

The literature documents several experimental results on the adoption of behaviors including network structure – such as the study conducted in [8], [9] and decision making [10]. Under these studies, it is observed that individual adoption




(a) (b)

Fig. 3: The hazard ratio of adoption considering the 3 covariates on the X axis - the number of switches (when user makes different decisions on successive rounds) made, the number of peer signals, the number of options explored. Shown are the 95% confidence intervals from the time varying Cox proportional hazards model. The effect of the covariates on the likelihood of adoption is significant if the 95% confidence interval does not contain 1. (a) Delayed cascade (b) Early cascade.

is much more likely when participants received social reinforcement from multiple neighbors in the social network as opposed to a single exposure. These studies focus on the effect of network structure on the dynamics of behavioral diffusion. Contrary to this, we quantify influence using only the number of signals temporally sent to a user irrespective of how the signals diffused to its neighbors prior to its own adoption. Our focus here is on using social influence as a strategic tool for exploitation and from our experiments, one of the conclusions that emerges is that the pattern of influence can indeed be manipulated to steer users towards suboptimal technologies.

**Acknowledgments**. Some of the authors are supported through the ARO grant W911NF-15-1-0282.

## References

[1] Das, Sauvik, et al. "The role of social influence in security feature adoption." Proceedings of the $18^{th}$ ACM conference on computer supported cooperative work social computing. ACM, 2015.

[2] Das, Sauvik, et al. "The effect of social influence on security sensitivity." 10th Symposium On Usable Privacy and Security (SOUPS 2014). 2014.

[3] Shakarian, Paulo, Jana Shakarian, and Andrew Ruef. Introduction to cyber-warfare: A multidisciplinary approach. Springer, 2013.

[4] Nickell, Gary S., and John N. Pinto. "The computer attitude scale." Computers in human behavior 2.4 (1986): 301-306.

[5] Scott, Susanne G., and Reginald A. Bruce. "Decision-making style: The development and assessment of a new measure." Educational and psychological measurement 55.5 (1995): 818-831.

[6] Modic, David, Ross Anderson, and Jussi Palomäki. "We will make you like our research: The development of a susceptibility-to-persuasion scale." PloS one 13.3 (2018): e0194119.

[7] Wilson, Alan Geoffrey. "The use of the concept of entropy in system modelling." Journal of the Operational Research Society 1970.

[8] Centola D. The spread of behavior in an online social network experiment. science. 2010.

[9] Bao P, Shen HW, Chen W, Cheng XQ. Cumulative effect in information diffusion: empirical study on a microblogging network. PloS one.

[10] Grujić J, Eke B, Cabrales A, Cuesta JA, Sánchez A. Three is a crowd in iterated prisoner's dilemmas: experimental evidence on reciprocal behavior. Scientific reports.

[11] Aral, Sinan, and Dylan Walker. "Creating social contagion through viral product design: A randomized trial of peer influence in networks." Management science 57.9 (2011).

[12] Lakkaraju K, Medina B, Rogers AN, Trumbo DM, Speed A, McClain JT. The Controlled, Large Online Social Experimentation Platform (CLOSE). In Springer SBP 2015.