POSTER: Equitable Security: Optimizing Distribution of Nudges and Resources

Elissa M. Redmiles, John P. Dickerson, Krishna P. Gummadi, and Michelle L. Mazurek
eredmiles@cs.umd.edu

ABSTRACT

Security behaviors can help users avoid incidents, but can also increase costs, both to users – in time and mental effort – and to platforms – in user engagement and engineering resources. As such, we should consider when it is most efficient and effective to encourage security behaviors. Recent work has shown that users attempt to make security decisions based on cost benefit tradeoffs (boundedly, rationally). Yet, sometimes security nudges (e.g., create unique passwords for every website) encourage users toward irrational behavior: creating strong, unique passwords even for those sites that contain no personal data. In this work-in-progress, we present a mechanism design (a framework) that can be used to optimize the distribution of security nudges and requirements among users with different levels of risk or different levels of investment in a given system. Further, we introduce a new paradigm: the distribution of resources (e.g., ubikeys) that can lower the cost of security behaviors to those users with the most need (the highest time cost from 2FA or lowest Internet skill). Future work will involve simulations showing the value of optimizing distribution of nudges and resources using this framework, and evaluating such an approach in a live test.

ACM Reference Format:
Elissa M. Redmiles, John P. Dickerson, Krishna P. Gummadi, and Michelle L. Mazurek, eredmiles@cs.umd.edu . 2018. POSTER: Equitable Security: Optimizing Distribution of Nudges and Resources. In *2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18), October 15–19, 2018, Toronto, ON, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3243734.3278507

1 INTRODUCTION

Digital security requirements or nudges are often established with little consideration of unique differences among end users. The same is true in other domains related to risk: insurance companies and doctors recommend that obese patients exercise copiously without accounting for the cost – in time or gym fees – of that recommendation to the patient, nor how that effort may correlate – or not – with the patient's interest in their own health. Similarly in security, online websites require or forcefully recommend that users engage in security mechanisms such as long, strong, and complex passwords or enabling two-factor authentication without accounting for the effort and time cost of those requirements in relationship to the user's investment or valuation of their online account on that system – and the risk that account may or may not pose to other accounts; or the effort they are already investing in a multitude of similar systems.

Even those entities that do not stand to gain from recommending security such as the U.S. National Institute of Standards and Technology [1] and Teen Vogue [3] recommend that everyone adopt two-factor (or multi-factor) authentication for the sites that offer it, again with little mention of variable costs and difficulty to users with differing skill levels, numbers of online accounts, and etc.

Behavioral mechanism design traditionally enables the balancing of user utility (e.g., what value a user generates from an account or from protecting it) with firm utilities (e.g., the value an online site gets from having a user enable a security behavior or use their system) within constraints (e.g., cyberinsurance or governmental policies). While such approaches have been used to solve a diverse set of problems, to our knowledge behavioral mechanism design has never been applied to end-user security. In this work-in-progress, we define a general behavioral mechanism for balancing user and firm utility in systems with inherent risk and protective behaviors that must be adopted by users with the goal of understanding (1) how firms or systems communicate with users about the value of protective behavior in order to encourage adoption, (2) how government policies should be set to ensure that firms behave *fairly* toward users, where fairness is defined as reducing the risk variance between different users until a minimum level of safety is met and reducing the effort variance between users with different resources (e.g., not marginalizing groups of users) and (3) how resources can be distributed among users to minimize inequities between people with different Internet skill or ability.

2 MECHANISM

In our prior work [4] we constructed an online experimental system in which crowdworkers made a security choice – enabling two-factor authentication (2FA) or not – given an explicit set of risks (a percent chance that their study account would be hacked and they would not be compensated, and a percent protection from hacking they would receive from enabling 2FA). We measured the cost of the security behavior to the crowdworker in terms of the time it took them to log in and sign up; since crowdworkers earn money from completing micro tasks, seconds or minutes wasted in our game lead to direct wage losses. Using these measurements, we model users' security decisions as a function of costs (C), risks (R), and user tendencies and attributes (U) and find that (1) we were able to model security decisions with high accuracy (R^2=0.61) and (2) this model of behavior is robust across users of different demographics, skill, and security tendencies (i.e., password strengths). Finding that overall, user behavior relates to: a) costs (e.g., time it takes to login to a system or enable 2FA) and b) prior behaviors, but c) can be adjusted through messages communicating risk and efficacy.

CCS '18, October 15–19, 2018, Toronto, ON, Canada

ACM ISBN 978-1-4503-5693-0/18/10.
https://doi.org/10.1145/3243734.3278507

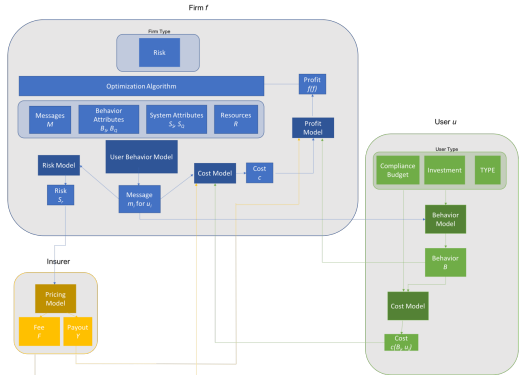

Figure 1: Overall mechanism design.

Here, we generalize this approach to design a mechanism (framework) for mathematically selecting the values of different system features can be used to maximize utility for both users and online services (Figure 1).

2.1 General Mechanism

People (users) use systems that offer them some value (e.g., buy into insurance systems that lower their healthcare costs, store money in bank accounts that offer them some interest). The world has some inherent risks (W_r) that will cause loses for these users. System owners (firms) sustain losses when the users sustain losses (e.g., firms sustain losses proportional to user loss; user loss is dependent on world risk, the user's type, and the user's system-relevant behavior). Firms attempt to reduce user risk (and thus the firm's losses) by making protective behaviors available to the user, the firm can invest to make these protective behaviors more or less valuable, and can also communicate true (or false) information to the user in order to get them to enact protective behaviors. These protective behaviors reduce the user's risk and thus the firm's risks and costs. The behaviors also, however, cost the user (in effort, time, or even money) and may cost different users different amounts (e.g., cost of protective behavior is dependent on user type). The behaviors may in some cases also have a cost to the firm (e.g., the price to send a SMS message to each user that enables two factor authentication).

In sum, firm's build their systems to protect users from world risks inherently, but can only do so up to a certain point. To gain additional protection, users must adopt protective behaviors. In our formulation, we account for the costs and benefits to *users* from adopting a protective behavior *and* we account for the costs and benefits to a firm that hosts some digital system from the behaviors chosen by users of that system. In our mechanism we consider systems in which there are a set of n possible system users. Each user, $u_i, i \in n$, gains utility from using the system (e.g., they benefit from using the system). Falling victim to the "world risks" lead to losses for both the users and the firms, thus protective behaviors offer some utility to both the users and the firms (but these behaviors also come at some cost to users, and sometimes also to firms).

The System Parameters. We define the system as having a set of private parameters. The system has an overall quality, S_q that influences the cost of protective behaviors for the user and risk S_s that is equal across users – this risk is lower than the risk of the world (improvements are made by the firm investing resources, but no system can be 0% risky). There is also a set of protective behavior(s) that users can enable within the system B to improve reduce their personal risk. These behaviors have some quality B_q (e.g., how much they cost the user and the firm) and protection level B_s (e.g., how risk reduction the behavior offers). Finally, the system also has a parameter that can be varied per user, in which they can allocate some resources to certain users (R) to improve B_q (reduce user cost). Thus, the system's private parameters consist of: S_q, S_s, B_q, B_s, R, where the parameters can be increased by some addition of resources by the firm (up to a threshold).

The Firm's Utility Function. The firm generates some utility (and in some cases, some cost) from users using the system and users adopting protective behaviors within the system. That utility is dependent on both the users and their behaviors and drives some real monetary value for the firm: $f^s : (B, U)^n = \mathbb{R}$. This utility can be computed as $f^s(B_i, u_i)_{i=1\ldots n} = \sum_{i=1}^{n} g(B_i, u_i) - c(B_i, u_i)$ where g is some gain function and c is some cost function.

The User's Parameters. Users have a private TYPE (e.g., their capability to do a protective behavior, their risk tolerance, their investment in the system), where u_i has some $t_i \in TYPE$. User's individual risk, u_r, depend on the system, W_r and the user's TYPE.

The User's Utility Function. The utility a user gains from a system depends on their type, the protective behavior they select, and the resources invested in them by the system. That is, $f^u : (TYPE, B, R) = g(B_i, t_i, R_i) - c(B_i, t_i, R_i)$.

The Firm's Levers. The firm has some levers that it can adjust to alter user protective behavior: it can adjust the true parameters of the system and/or it can communicate persuasive information to the user. Persuasive information takes the form of some revealed message m. This message may provide the true values of the system parameters (for example, B_s) or fictitious values of those parameters. m can be tailored to drive some user protective behavioral response.

User Behavior Adjustment. User behavior adjusts dependent on the users' parameters, the message they are shown, and the resources invested in them, subject to the constraint that behavior will not be adjusted if the cost of enabling a behavior is above the user's overall cost budget, which is part of their parameters and depends on the amount of cost they have sustained from behaviors across all of their systems of the same kind. That is, user behavior adjusts based on the following equation: $if(\sum_{d=0}^{e} budget) < \sum_{t=0}^{e} cost(B_i, U_i) : m_i \times t_i \times r_i$, where $d = 0$ is the time at which the user starts using the system and $d = e$ is the time at which the user stops using the system and $budget \in TYPE$.

2.2 Application to Digital Security

We can apply this mechanism to end-user digital security. When users use a website they are at some risk of being hacked (W_r). The website itself has some level of security (S_s) that reduces W_r and some level of overall quality S_q (e.g., how long it takes for a user to login). There is a set of protective (i.e., security) behavior(s) that users can enable within the system B in the system (e.g., turning on two factor authentication (2FA)). The quality B_q of these behaviors is (e.g., how long it takes for a 2FA text to get sent) as well as the overall system quality S_q influences the cost of the behavior to the user. The protection value (B_s) is e.g., how much using 2FA reduces the risk of hacking, influences the value of the behavior to both the user and the firm.

System private parameters can be adjusted by greater engineering investment (to increase S_s, S_q, B_q, and potentially, B_s) or, for R, monetary investment (e.g., in support staff to help people learn the security behavior, provision of easier to use but more expensive tools like Ubikeys).

The User's Parameters. The user's TYPE includes their capability to do a security behavior, their typical security behavior on similar websites, their risk tolerance (e.g., how much risk of hacking they are comfortable with), and their investment in the system (e.g., if this is a bank account, how much money they have stored in the account).

The User's Utility and Goal. Where the user's goal is to maximize their utility, $max(f^u)$ over the space of behaviors that they can possibly adopt.

The Firm's Utility, Goal, and Levers. The firm sustains losses from users getting hacked (as they have to pay back users and they lose good PR), and may have costs from the level of risks of the system users (cyber-insurance plans are often priced based on estimated risk of the firm, which is based in part on user risk of hacking) and thus the firm gains utility from users adopting security behaviors. However, the firm also sustains costs from adoption of security behaviors as these behaviors may decrease user engagement or have direct costs (e.g., cost to send 2FA text message). The firm's goal is to maximize their gains over the set of users and behaviors by adjusting the m shown to the user and the true values of the systems parameters.

2.3 Constrained Mechanism: Introducing Policies

In the mechanism we have defined thus far, there are no restrictions on what the firm can communicate in m (that is, how much the firm can lie to users in order to encourage them to behave in a particular way), no requirements on fairness for equity in risk between different users of the same system, and no requirements around equity of effort or cost of behaviors to different users of the same system.

In many cases, we may wish to impose such restrictions, however. False advertising laws restrict the "lies" firms can tell consumers, and maximum tolerable risk is often defined for a multitude of publicly used systems (e.g., power plants [2]). Similarly, then, we could imagine imposing constraints such that online websites cannot put their users at more than a H% risk of being hacked.

To achieve H% may require user behavior. In this case, we may wish to impose additional policies. We may wish to impose additional policies, a fairness constraints. For example, a risk fairness constraint: all people in the system should have as equal as possible risk of a negative outcome (e.g., a user of one race should not have a greater risk of being hacked than one of another race), given the same protective behaviors. Or, effort fairness: resources should be assigned and behaviors should be created such that user variance in cost (effort) of behavior required to reach H is minimized. In this context, a policymaker wishes to set optimal policies and maximize compliance.

3 FUTURE WORK

Using data collected from prior empirical experiments (see Redmiles et al.), we can model how varying the message shown – e.g., showing different values of B_s or S_s – can alter participant behavior. In these empirical experiments, u_i creates an account in the system (bank.cs). During this process some information is collected about the user's TYPE, t_i, through observation. Users are provided a set investment in the system I. The system's private parameter S_s is set as H, the percent chance that a users' account will get hacked and they will lose their investment, I; S_q is a function of the engineering of the system: users have some average login speed. The system has a binary set of security behaviors B: enable or do not enable two-factor authentication (2FA) that users can turn on or off. B_{2FAon} reduces the risk of hacking by some protection percent ($B_s = P$), costs the user some amount of time, dependent on their type and the quality of the behavior and the system, $c(B_{2FAon}, t_i) = B_q \times t_i \times S_q$, and costs the firm some amount of money (e.g., \$0.09 cents per 2FA text message sent). For both the user and the firm, $u(B_{2FAon}, t_i) = t_i \times S_q \times S_s \times B_q \times B_s \times R$. The system attempts to adjust user behavior by revealing S_s and B_s to the user via a message m.

In our future work, we will incorporate these simple empirical-based estimations into a linear optimization framework to solve the optimization problems presented in the prior section. This will allow us to optimize over both firm and user variables to establish a set of potential optimal parameters for security nudges and policies within a particular context. Once validated, we are in the discussion stages of using these values in a live test with an end-user facing software platform.

REFERENCES

[1] [n. d.]. Back to basics: Multi-factor authentication (MFA). https://www.nist.gov/itl/tig/back-basics-multi-factor-authentication
[2] Baruch Fischhoff. 1984. *Acceptable risk.* Cambridge University Press.
[3] Nicole Kobie. 2017. Why Two-Factor Authentication Is So Important. https://www.teenvogue.com/story/why-two-factor-authentication-is-important
[4] Elissa M Redmiles, Michelle L Mazurek, and John P Dickerson. 2018. Dancing Pigs or Externalities?: Measuring the Rationality of Security Decisions. In *Proceedings of the 2018 ACM Conference on Economics and Computation.* ACM, 215–232.