Agent-Based Modeling of User Circumvention of Security

Position Paper

Vijay Kothari
Department of Computer Science
Dartmouth College
vijay.h.kothari.gr@dartmouth.edu

Jim Blythe
Information Sciences Institute
University of Southern California
blythe@isi.edu

Sean Smith
Department of Computer Science
Dartmouth College
sws@cs.dartmouth.edu

Ross Koppel
Department of Sociology
University of Pennsylvania
rkoppel@sas.upenn.edu

ABSTRACT

Security subsystems are often designed with flawed assumptions arising from system designers' faulty mental models. Designers tend to assume that users behave according to some textbook ideal, and to consider each potential exposure/interface in isolation. However, fieldwork continually shows that even well-intentioned users often depart from this ideal and circumvent controls in order to perform daily work tasks, and that "incorrect" user behaviors can create unexpected links between otherwise "independent" interfaces. When it comes to security features and parameters, designers try to find the choices that optimize security utility—except these flawed assumptions give rise to an incorrect curve, and lead to choices that actually make security worse, in practice.

We propose that improving this situation requires giving designers more accurate models of real user behavior and how it influences aggregate system security. Agent-based modeling can be a fruitful first step here. In this paper, we study a particular instance of this problem, propose user-centric techniques designed to strengthen the security of systems while simultaneously improving the usability of them, and propose further directions of inquiry.

1. INTRODUCTION

At a relatively simple level, we can look at security as making a design choice that optimizes some overall security goal (while staying within economic constraints). For example, a security officer concerned with reducing the risk of some adversary logging into a password-protected interface might choose to force users to have long, complex, non-personally meaningful passwords which must be changed on a regular basis. In other words, the more the officer "dials up" her control knob, the more secure the system is.

ACySe '14, May 06 2014, Paris, France

ACM 978-1-4503-2728-2/14/05 ...$15.00.
http://dx.doi.org/10.1145/2602945.2602948

However, fieldwork (e.g., [4], [6], [7], [8], [14]) shows that human users behave in ways that subvert this model. For example, if a password is too complex, human users will write it down (thus *increasing* compromise risk); if forced to change a critical password, some users will change passwords for other accounts (outside the officer's purview) to be the same—thus increasing both risk *and* consequence of compromise. Thus, we see a descent into an *uncanny valley*: where implementing a more stringent security mechanism may reduce the overall security posture [12].

Addressing this paradox requires understanding how and why user behavior differs from "optimum." People behave differently for a number of reasons: first, they may have a different model of their environment from the security designers, with typically a richer set of goals and action costs. Second, they may have a different model of the security risks. The designer may typically have a more accurate model but this is not always true. Third, even with correct models, humans often make suboptimal choices, for example due to biases in decision making, distractions, emotion and fatigue.

Adams and Sasse [1] challenged the belief that users are averse to engaging in secure behaviors and motivated the need for "user-centered design in security mechanisms." Through interviews, Beautement et al [2] discovered that "the key factors in the compliance decision are the actual and anticipated cost and benefits of compliance to the individual employee, and perceived cost and benefits to the organization." They also identified numerous examples of these factors. Perceived costs of compliance included increased physical load, increased cognitive load, embarrassment, missed opportunities and the 'hassle factor'. Perceived benefits of compliance included avoiding the consequences of a security breach and protection from sanctions. Finally, external factors of compliance included design, awareness, training, and education, the culture of the organization, monitoring and sanctions. Furthermore, they postulated that users will comply when the task cost of compliance (individual cost - benefits) does not exceed the compliance threshold. Fieldwork also indicates security systems (e.g., anti-virus software, backup systems) often provide users with a false sense of security, allowing users to believe such systems act as safe-

guards against their own circumventions and security lapses.

Agent-based modeling can provide important insights into the question of the optimal security posture by taking the potential human response into account. Some might classify circumvention as a *wicked problem* (e.g., [15]) – an extraordinarily challenging problem that does not lend itself to definitive formulation, which requires novel ideas and techniques to address. We argue that agent-based modeling can nonetheless provide significant value. Even when it is impossible to define *a priori*, how users will circumvent controls, agent-based simulations may be able to predict when users would be inclined to circumvent given the opportunity, e.g. by estimating frustration levels. In some situations, we may be able to address sources of frustration, and, in doing so, address the problem of circumvention, without identifying a particular method of circumvention. Furthermore, agent-based simulations may enable us to look at not just the causes for surprising behavior, in the aggregate, but also at the effects, in the aggregate. In other situations, where we can identify *a posteriori*, a method of circumvention, agent-based modeling may provide a deeper understanding of the factors conducive to circumvention (such as the types of password misbehavior documented in fieldwork surveys).

One can view the knowledge of a circumvention and the propensity to use one as diffusing through a network of agents. In this case agent-based simulations may help identify which population groups are most susceptible to engaging in workarounds, and where the tipping point is between between the occasional workaround and ubiquitous circumvention. Using this information, it may be possible to re-structure a network of users to reduce the prevalence of workarounds, e.g. by isolating groups or through education about the organizational threats that are reduced by security mechanisms. Additionally, agent-based simulations may help predict the efficacy of countermeasures to circumvention without the costs and risks associated with implementing such countermeasures in practice.

In order to make effective predictions, an agent platform must capture some of the behaviors that might be expected of human users. In particular it should capture the mental models of end users where they may differ from those of the system designers, and the potential effects of known biases and of emotions such as frustration on user compliance. We are in the process of developing an agent platform that captures these aspects. In this position paper we briefly present the approach and describe an example where agent models may be used to predict the best timeout value for automatically logging users off in multi-user environments. A model that assumes user compliance may find that a short timeout is optimal, because it minimizes the chance of another user either accessing material inappropriately or accidentally entering information to the wrong account. However our preliminary model predicts that a longer timeout may provide better performance in an environment where the logout procedure can be actively defeated—as actually and unexpectedly happened in a real-world deployment at a partner institution.

In Section 2 we describe this scenario in more detail. Section 3 describes an agent-based approach that allows us to explore the impact of workarounds and compliance on the optimal design choices. Section 4 discusses future work and other domains in which to apply the approach. Section 5 concludes.

2. AN ANECDOTE REGARDING TIMEOUTS

In a partner institution (a large hospital), security officers were concerned that *COWs* (*computers on wheels*) and desktop machines were too often left logged-in but unattended by clinicians, creating the risk for confidentiality and integrity problems in stored patient data.

To address this problem, officers attached proximity sensors to the machines with controls that would automatically log off any user after a fixed duration of inactivity. If the officers chose a timeout that was very short (say, 1 second), the system would become unusable—massively frequent interruptions of workflow would cause users to noisily complain. However, beyond this short window, one would assume that longer timeouts yield worse net exposure (e.g., total minutes of logged-in but unattended machines). Anything between too-short and infinite would be an improvement on the original scenario.

What happened was unexpected: frustrated with the interruptions (and perhaps with accuracy problems in the proximity detectors), some users instead put styrofoam cups on the detectors, leading to all systems always believing a user was present. The naive designer model suggested a monotonic decrease in net exposure as the timeout increases; a more accurate model would require taking into account the aggregate effects of user behavior, such as: (1) how frustrated different users might get with timeouts perceived as too short; (2) how quickly frustrated users might seek a workaround, such as the styrofoam cups; (3) how many of the remaining users, who might not actively break the system themselves, might happily use a system where someone else had left a styrofoam cup; (4) whether having a security professional walking the wards periodically and removing cups would actually improve things. (How many patrols would be needed to make a real difference?)

2.1 A Different Approach to the Timeout Decision

Enabling designers to make better security decisions requires enabling them to reason about these issues. Our goal is to build an agent-based model, as a first step away from the naive, incorrect model. Such a system will allow designers to explore in simulation the effects of different strategies, estimating the net benefits to security and overall organizational efficiency, and considering variables such as user frustration, to the extent they can be well modeled.

The naive approach to timeouts uses a fixed timeout threshold and neglects factors that are indicative of frustration experienced in the event of a timeout. These factors, which include the intended use of the system, fatigue, and stress, motivate choosing a timeout value that is sensitive to the user's state, actions, and working environment rather than choosing a constant. Given feedback, for example, a security tool might learn to estimate the user's expected frustration caused by a timeout based on the open appli-

cations on the computer, time of day, and domain-specific indications of workload such as the patient roster. It might also learn to estimate the the probability that a very frustrated user might execute an unforeseen workaround, and the probability that a less frustrated user might copy such a workaround if she sees it. Such a tool is likely to require data about the operations where it is to be deployed, but the parameters of interest and initial values could be set by learning within the simulation.

It is often counterproductive and in some situations even dangerous to consider security objectives in isolation. Security mechanisms implemented to realize security goals often impact other organizational goals in a significant way that is often left unaccounted for. However, the impact of security mechanisms on other organizational goals is not limited to scenarios involving workarounds. Even in the absence of workarounds, stringent security mechanisms can induce stress, fatigue, and changes in mood that impact workflow and hinder progress toward numerous organizational goals.

In general, an optimal security strategy will depend on the interplay between security objectives and organizational objectives, and between different individuals in the organization. The complexity of the problem is one motivating factor for an agent-based model of the system to be secured, that captures the objectives of individual agents and factors that influence their likelihood of compliance with security protocols. We envision such a platform being employed by security designers to test various security mechanisms when it is infeasible to run real experiments for various reasons.

3. MODELING WORKAROUNDS

The agent-based platform we are developing builds on DASH, a framework for modeling human agents [3]. DASH combines a dual-process approach with a BDI agent: at each time step, an instinctive module may suggest a plausible action to be taken directly by the agent or it may defer to a deliberative reasoning module, which employs a BDI reactive planner to choose an action based on explicit goals. The instinctive module maintains an activation strength on nodes in memory that can be modified by an appraisal-based emotion mechanism [21]. This approach provides a natural model for frustration, as a strong activation produced by a negative appraisal of entities that are seen as conflicting with the agent's plan. In field work, frustration has emerged as a significant factor in the application of workarounds.

Regarding the styrofoam cup example, clinicians may perceive the following as organizational and individual goals: minimizing medical errors, ensuring patients are treated with dignity and attended to promptly, minimizing unnecessary exposure of patient data. These goals give rise to a utility function that can be used to evaluate the perceived outcomes of executing plans, which is done through the reasoning module. When using deliberative reasoning, a DASH agent by default chooses among alternative actions by projecting the plan associated with each one and picking the plan whose outcome has highest utility according to its beliefs.

However, planning achieved through the reasoning module may also be bypassed using the instinctive module when the agent is subject to certain emotions such as frustration. For example, an agent under stress caused by time pressure may skip a step to check that it is logged into the computer rather than another agent, and proceed to entering prescription information, since this more material step receives high activation from the instinctive module. Since the instinctive module provides input on goals as well as actions, frustration may also impact action choices when the reasoning module is employed. Suppose a clever user is frustrated with a policy requiring routine password change. If the user is aware that the system only prohibits the current password during a reset, and the perceived burden of remembering a new password is sufficiently high, she might create a plan to call the help desk after a reset so as to have her old password restored. This matches behavior seen in fieldwork. Other attributes (e.g. fatigue, stress) may also have an impact on the user's perception, beliefs, and mental processes.

We have implemented a prototype model as a first step to exploring the interaction of security goals and user behavior in the timeout scenario, and students in a special-topics course have built models for other scenarios. We are now in the process of extending the models to account for more of the behaviors found in fieldwork or by some of the richer models available in DASH.

Validation is a significant challenge. In a two-step approach, we will begin by showing that our models can duplicate behaviors seen in our and others' field studies. Next, we are planning experiments to uncover and where possible manipulate internal states such as frustration ([9], [10], [11], [13]), stress and fatigue, to both test hypotheses about the role they play and show that our agent models can capture the behavior at a deeper level.

4. SOME OTHER APPLICATIONS

In this section, we discuss future work and potential applications of DASH to understanding other behavior that would be considered undesirable from a security perspective along with other future work.

The approach we have described here is limited in that workarounds must be explicitly described in the model in order to be used, and therefore the simulation could never be used to predict workarounds that are completely unexpected by the designers. It may be useful to relax this assumption by allowing agents to search a space of plan library modifications to find potential abstract workarounds. The agent could then analyze the way in which the security protocol reduced effectiveness of the agents plan and hypothesize an action or a change to the effects of an action that would in turn defeat this disruption. In the example of this paper, we might hypothesize that agents may stop the timeout from taking effect if they can find a way to nullify it that is less costly than making numerous logins. Examples with more detailed security protocols might lead a number of steps that might be open to attack.

As we discussed in the introduction, one promising avenue might be the exploration of policies for password-based authentication. Fieldwork (e.g., [5, 6, 7, 8, 14, 16, 17]) provides us a number of interesting user behaviors regarding

passwords. Most users re-use a small number of unique passwords across a large number of sites. Some users circumvent password complexity rules by writing down passwords. Others circumvent by constructing passwords that are all small modifications of each other. Most users never change their passwords unless they are forced to. Many users cannot accurately recall seldom-used passwords, at least within the first few guesses. Many users have easily-guessed answers to security questions, even if their passwords are strong. Many users simultaneously know what good password practices are, but fail to follow them. Some users try to choose stronger passwords when they perceive compromise of an account might hurt them personally. Many users come up with the right passwords for the wrong usernames or for the wrong services.

What is the aggregate security effect of a decision regarding password policy, if users behave as the surveys suggest in the proportions the surveys suggest? What shifts in user demographics (e.g., from law or better training) might yield the best results?

Another set of avenues might be exploring behavior-based workarounds and errors in enterprise *authorization* (e.g., [18, 19, 20]) Commercial enterprises tend to *over-entitlement*, as the perceived costs of under-entitlement are too high. Enterprises also tend to over-entitlement, because users tend to accumulate permissions over their career path (even keeping irrelevant ones across promotions and transfers). Users may solve *under-entitlement* by circumventing the system completely—so the de facto access permitted by an enterprise's system may end up much larger than what the infosec managers perceive. For security officers, the actual costs of under-entitlement—personally dealing with and assuaging angry users—may be much higher than the costs of over-entitlement. Many managers provision new employees by copying-and-pasting entitlements from current employees, rather than thinking in detail.

One infosec officer at an investment bank reported that potential clients would judge his bank's security by the question "how many of your employees will be able to see my data?" Realistically reasoning about this question, or the net amount of exposure, the costs of under-entitlement, how much exposure could be reduced by hiring N more officers or switching to scheme Y, all requires modeling the aggregate behavior of humans.

5. CONCLUSION

In this paper, we have argued that to realize systems security goals we must first fully understand the nuances of users and user behavior. We have also argued that we must stop looking at security objectives in isolation — incorrect assumptions by security designers can have very real repercussions, e.g. due to circumvention, that impact non-security goals. Agent-based modeling can help on both fronts. In particular, we believe that the DASH framework can assist system designers in understanding user behavior, predicting the prevalence of workarounds, and measuring both security and organizational benefits of various systems. We described a particular scenario in timeouts and believe the agent-based approach will be useful in a number of other applications, including password-based authentication and authorization.

6. ACKNOWLEDGEMENTS

This material is based in part upon work support by the Army Research Office under Award No. W911NF-13-1-0086.

7. REFERENCES

[1] Anne Adams and Martina Angela Sasse. Users are Not the Enemy. *Communications of the ACM*, 42(12):40–46, 1999.

[2] Adam Beautement, M Angela Sasse, and Mike Wonham. The Compliance Budget: Managing Security Behaviour in Organisations. In *Proceedings of the 2008 Workshop on New Security Paradigms*, pages 47–58. ACM, 2009.

[3] Jim Blythe. A Dual-Process Cognitive Model for Testing Resilient Control Systems. In *Resilient Control Systems (ISRCS)*, pages 8–12. IEEE, 2012.

[4] Jim Blythe, Ross Koppel, and Sean W Smith. Circumvention of Security: Good Users Do Bad Things. *Security & Privacy, IEEE*, 11(5):80–83, 2013.

[5] S. Brostoff and M.A. Sasse. Ten Strikes and You're Out: Increasing the Number of Login Attempts Can Improve Password Usability. In *Proceedings of CHI 2003 Workshop on HCI and Security Systems*, 2003.

[6] R. Dhamija and A. Perrig. Deja Vu: A User Study Using Images for Authentication. In *Proceedings of the 9th USENIX Security Symposium*, 2000.

[7] Dinei Florencio and Cormac Herley. A Large-Scale Study of Web Password Habits. In *International Conference on World Wide Web*, pages 657–666. ACM, 2007.

[8] Shirley Gaw and Edward W Felten. Password Management Strategies for Online Accounts. In *Symposium on Usable Privacy and Security*, pages 44–55. ACM, 2006.

[9] Richard Hazlett. Measurement of user frustration: A biologic approach. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 734–735. ACM, 2003.

[10] Ashish Kapoor, Winslow Burleson, and Rosalind W Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724–736, 2007.

[11] Jonathan Klein, Youngme Moon, and Rosalind W. Picard. This computer responds to user frustration:: Theory, design, and results. *Interacting with computers*, 14(2):119–140, 2002.

[12] Masahiro Mori. The uncanny valley. *Energy*, 7(4), 1970.

[13] Carson Reynolds. *The Sensing and Measurement of Frustration with Computers*. PhD thesis, MIT, 2001.

[14] Shannon Riley. Password Security: What Users Know and What They Actually Do. *Usability News*, 8(1), 2006.

[15] Horst Rittel and Melvin Webber. Dilemmas in a General Theory of Planning. *Policy sciences*, 4(2):155–169, 1973.

[16] Stuart Schechter, AJ Bernheim Brush, and Serge Egelman. It's No Secret. Measuring the Security and Reliability of Authentication via ?Secret? Questions. In *IEEE Symposium on Security and Privacy*. IEEE, 2009.

[17] Stuart E Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The Emperor's New Security Indicators. In *Security and Privacy, 2007. SP'07. IEEE Symposium on*, pages 51–65. IEEE, 2007.

[18] S. Sinclair and S.W. Smith. Preventative Directions for Insider Threat Mitigation via Access Control. In S. Stolfo et al., editors, *Insider Attack and Cyber Security: Beyond the Hacker*, pages 173–202. Springer-Verlag Advances in Information Security 39, 2008.

[19] S. Sinclair and S.W. Smith. What's Wrong with Access Control in the Real World? *IEEE Security and Privacy*, 8(4):74–77, July/August 2010.

[20] S. Sinclair, S.W. Smith, S. Trudeau, M.E. Johnson, and A. Portera. Information Risk in Financial Institutions: Field Study and Research Roadmap. In *International Workshop on Enterprise Applications and Services in the Finance Industry (FinanceCom)*, 2008.

[21] Marc Spraragen. Modeling the effects of emotion on cognition. In *Proc. AAAI*, 2012.