Agent-Based Modeling of User Circumvention of Security

ABSTRACT

Security subsystems are often designed with flawed assumptions arising from faulty mental models held by system designers. 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, and that 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.

However, fieldwork (e.g. [4], [6], [7], [8], [10]) 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, what seems to be a simple monotonic relationship may in fact harbor what graphics specialists call *uncanny valleys*: places where dialing up the controls actually decreases overall security posture.

Solving this problem requires understanding how and why user behavior differs from "optimal."

People behave differently for a number of reasons: (1) they have a different model of their environment from the security designers, with typically a richer set of goals and action costs, (2) they have a different model of the security risks. The designer may typically have a more accurate model but this is not always true, (3) even with correct models, humans often make suboptimal choices, for example due to biases in decision making, distractions, emotion and fatigue.

Agent models of human behavior might be an effective way to better predict the impact of a security system in a given environment. These models may predict otherwise unexpected outcomes due to individual circumvention or emergent effects when a group of users collaborate over a network. 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. 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.

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 are 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 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), clinicians used *COWs* (*computers on wheels*) and desktop machines. The security officers were concerned that these systems were too often left logged-in but unattended, creating the risk for confidentiality and integrity problems in the stored patient data.

To address this problem, the officers attached proximity sensors to the machines and added controls that, when detecting that a computer had been left unattended but logged-in for a predetermined, fixed, period of time, the user would be automatically logged off. 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 linear curve; a more accurate model would require taking into account the aggregate effects of aggregate user behavior, such as: (1) how frustrated different users might get with timeouts perceived as too-short; (2) how quickly such "edgy" users might find 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 intern walking the wards once a day and removing cups would actually improve things. (How many interns would there have to be 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 would 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.

For example, the simple approach to timeouts uses a fixed timeout threshold and neglects numerous factors that contribute to user frustration in the event of a timeout. One indicative factor of frustration experienced is the intended use of the system. A home computer user who is timed out due to a lack of keyboard interaction while watching a video on, say, NetFlix, is likely to experience a greater level of frustration than one who is timed out while checking the weather.

This motivates choosing a timeout value that is sensitive to the user's 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 applications on the computer, time of day, and domain-specific indications of workload such as the patient roster. At the same time it might learn to estimate the probability of a vulnerability caused by a lengthened timeout in the same way and use the two values to reason about the tradeoffs of user frustration and near-term security. Such a tool is likely to require extensive 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 measures implemented to realize security goals often impact other organizational goals in a significant way that is unaccounted for. For example, Koppel et al [9] demonstrated ways that circumvention relating to BCMA (Barcode Medication Administration) systems led to medical errors. Even in situations where workarounds are not employed, medical errors may arise due to sincere attempts by clinicians to comply with a system that interferes with workflow.

In general, the optimal security strategy will depend on the interplay between security objectives and organizational objectives, and between different individuals in the organization, even in cases where workarounds are not available. 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 measures when it is infeasible to run real experiments due to ethical costs, monetary costs, or for other 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 [16]. 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.

We have developed a model for the COW anecdote discussed earlier, capable of running simple simulations, based on a subset of DASH. This allows us to measure various security effects as a function of variables in the model, the most pertinent being the timeout threshold.

During each time step, each agent (clinician) begins or continues an action in accordance with a short plan that requires the agent to be logged in to a COW. During the simulation, agents may be timed out of a COW that they had previously been using. These timeouts occur when users are logged into a COW that they had not been using for a duration of time, the timeout threshold.

Under normal operation, the agent will not interfere with the timeout process and will log back in when necessary. However this action increases the agent's frustration level, a parameter that is tracked for this simulation and that decreases over time, unless an action causes it to be increased. When the agent's frustration level reaches a threshold, the agent may choose an alternative plan that defeats the autologout sensor with a styrofoam cup, if the agent is aware of this workaround. Some agents begin the simulation with this knowledge, while we model other agents as gaining the knowledge by witnessing another agent use the workaround.

We have not yet performed real-world validation of this model. It is described here as an implemented illustration of the agent modeling approach. However, we intend to perform simple lab experiments to validate the model by observing the correlation of timeout values and other factors that increase frustration with the frequency with which the workaround is employed.

In the future, we plan to build upon this model by incorporating richer alternative plans, explicit appraisal and a more faithful model of the way agents of different types interact in this environment. We would also like to implement a module that tracks the value of an organizational objective function, dependent upon the medical errors, monetary costs of plans, and security effects experienced during a given simulation, instead of naively considering a security objective in isolation. We believe that we would arrive at a vastly different timeout threshold if we optimize for organizational objectives instead of solely optimizing for a security objective.

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, 10, 6, 11, 12, 7, 8]) 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.

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., [15, 13, 14]) 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 the model the aggregate behavior of humans.

5. CONCLUSION

In this paper, we have argued that in order to realize systems security goals it is imperative that we first fully un-

derstand 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 have studied 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. 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), 2012 5th International Symposium on*, 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 *Proceedings of the 16th 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 *Proceedings of the Second Symposium on Usable Privacy and Security*, pages 44–55. ACM, 2006.

[9] Ross Koppel, Tosha Wetterneck, Joel Leon Telles, and Ben-Tzion Karsh. Workarounds to Barcode Medication Administration Systems: Their Occurrences, Causes, and Threats to Patient Safety. *Journal of the American Medical Informatics Association*, 15(4):408–423, 2008.

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

[11] Stuart Schechter, AJ Bernheim Brush, and Serge Egelman. It's No Secret. Measuring the Security and Reliability of Authentication via ?Secret? Questions. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 375–390. IEEE, 2009.

[12] 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.

[13] 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.

[14] 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.

[15] S. Sinclair, S.W. Smith, S. Trudeau, M.E. Johnson, and A. Portera. Information Risk in Financial Institutions: Field Study and Research Roadmap. In *3rd International Workshop on Enterprise Applications and Services in the Finance Industry (FinanceCom 2007)*, pages 165–180. Springer-Verlag Lecture Notes in Business Information Processing 4, 2008.

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