

Targeted Risk Communication for Computer Security

Jim Blythe
USC
Information Sciences
Institute
blythe@isi.edu

Jean Camp
Indiana University
School of Informatics
ljcamp@indiana.edu

Vaibhav Garg
Indiana University
School of Informatics
gargv@indiana.edu

ABSTRACT

Attacks on computer systems are rapidly becoming more numerous and more sophisticated, and current preventive techniques do not seem able to keep pace. Many successful attacks can be attributed to user errors: for example, while focused on other tasks, users may succumb to 'social engineering' attacks such as phishing or trojan horses. Warnings about the danger of these attacks are often vaguely worded and given long before the dangers are realized, and are therefore too easy to ignore. However, we hypothesize that users are more likely to be persuaded by messages that (1) leverage mental models to describe the dangers (2) describe particular vulnerabilities that the user may be exposed to and (3) are delivered close in time before the danger may actually be realized. We discuss the design and initial implementation of a system to achieve this. It first shows a video about a potential danger, then creates warnings tailored to the user's environment and given at the time they may be most useful, displaying a still frame or snippet from the video to remind the user of the potential danger. The system uses templates of user activities as input to a markov logic network to recognize potentially risky behaviors. This approach can identify likely next steps that can be used to predict immediate danger and customize warnings.

Author Keywords

Modeling and prediction of user behavior; Planning and plan recognition; Help intelligent assistants for complex tasks

ACM Classification Keywords

H.5.2 User Interfaces: Theory and Methods

General Terms

Human Factors, Security

INTRODUCTION

Attacks on computer systems are rapidly becoming more numerous and more sophisticated, and current preventive techniques do not seem able to keep pace. Many successful attacks can be attributed to user errors: for example, while

focused on other tasks, users may succumb to "social engineering" attacks such as phishing or trojan horses. Usability is therefore key to security, and in the field of *usable security*, security tools are evaluated in terms of how well users can secure their data and systems with the tools [7].

Usable security is distinct from many other usability challenges in that security is rarely the user's primary goal. The challenge is not to enable the individual's mastery of an application so much as to convince the individual to avoid digital risks by adopting appropriate security tools and application settings, despite the financial and time costs of doing so. A second characteristic interaction design challenge is that security should be neither entirely opaque nor entirely transparent. In usable security design, opaque systems allow the user to take an action seamlessly rather than requiring some understanding of the underlying system design. However, some information must be presented to the user in order for them to make security choices, or the default would be to prevent all risky behaviors without communicating. In fact, blocking desired activities without communication is one reason that users abandon security technologies even when the risks these technologies mitigate are known.

Conversely, a completely transparent security design would overwhelm the user with information about configuration, the nature of the security technology and the elements of a risk that are mitigated. An example of overly transparent design is the provision of hash information and public keys in certificate information given to users in an ubiquitous and almost universally ignored pop-up.

In this paper we describe an approach for *translucent* tools for security, that communicate risk choices only to the degree necessary to avoid inadvertent high-cost choices, and that therefore remain in use. Since security is not the user's primary goal it is important to limit the level of communication as far as possible, and to make the warning timely, to the point and effective. We combine several technologies in order to achieve this. We employ plan recognition and probabilistic reasoning to improve the tool's awareness of when the danger to the user is highest, in terms of the likelihood and cost of risky behavior. This allows the tool to restrict communication to situations when the potential for danger crosses a threshold and the dangerous actions will take place soon. We also use ideas from risk communication to inform the user effectively about the dangers and the relative costs and benefits of proposed mitigation actions. In particular we adopt a mental models approach from the environmen-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IUI'11, February 13–16, 2011, Palo Alto, California, USA.

Copyright 2011 ACM 978-1-4503-0419-1/11/02...\$10.00.

tal and health studies, another domain where the dangers of an action may not be immediately apparent. We also consider the modality of the warnings. Risk communication for personal computers has traditionally involved pop up boxes with text that is usually too technical for the user to comprehend. Studies in cognition suggest that the use of videos over text would lead to better comprehension [9, 15].

The contributions of this work are a set of mental models for communicating security risk, presented as videos, and a plan recognition tool that can create a specific, timely warning linked to a snippet or still from the video. In the next section we describe the mental models and videos that are used to give warnings about risks to the user. Following this we discuss in more detail our approach to user modeling and plan recognition that allows more targeted warnings. We end with a discussion of future work.

RISK COMMUNICATION AND SECURITY

Risk communication is the first step in enabling users to make good security decisions [4]. Previous studies have shown that users' expressed security preferences deviate from their behavior from real life [1]. It has been argued that this is because the user is unaware that they are taking risks at all. Thus there is a need for tractable feedback. Several efforts have been made to leverage mental models of users to provide them with real time information about their risk-taking behavior. For example, Web Of Trust¹ is an effort that informs the user whether a website is trustable or not. The drawback of many of these mechanisms has been the static nature of their feedback. With static risk communication users can become indifferent to the message being delivered. Thus the strength of communication must be appropriate in response to the risk being faced.

Risk communication technologies also need to take into account the decision making heuristics [18]. In particular, human judgment under uncertainty is affected by the valence effect (overestimating the likelihood of favorable versus adverse outcomes), availability (overestimating the probability of an event that is easier to recall), representativeness (categorizing an object based on how closely it resembles the category rather than the statistical likelihood) and other cognitive barriers [2]. Risk perception is also important: risks can be underestimated if they are perceived as voluntary, controllable, lacking in severity and the impact is not immediate [8]. Security risks are often not perceived differently from offline risks. There is also evidence that commonly accepted theories of offline decision making, *e.g.* Prospect Theory [12], may not hold true online [17]. (Prospect Theory posits a heuristic ordering of potential outcomes of a decision and a reference point from which to decide "losses" or "gains").

This creates unique challenges for risk communication as traditional techniques used for offline risks might not be effective for online risks. Previous studies have explored the use of graphics and symbols in risk communication messages to alter risk perception. However, there were no statistically valid results [14]. Users may also use incorrect

¹<http://www.mywot.com/>

signals to measure risk. For example, Jakobsson et al. [10] found that end user trust is based not on authentic phishing stimuli but rather on inconsequential indicators like document layouts, relevance, well formed URLs. It is important that we communicate the right information, at the right time, to the right stakeholder, framed in the right context.

Risk communication for personal computer security has traditionally used pop up boxes with text that is usually too technical for the user to comprehend. Studies in cognition suggest that the use of videos over text would lead to better comprehension, in particular with a story-based approach [9]. Risk communication is usually designed by computer scientists and thus tends to leverage the mental models of experts. Mental models refer to internalized representations of external reality [11]. There is evidence to show that mental models of security experts and users are not the same [3]. Camp [6] states that security experts predominantly use five mental models: physical, criminal, medical, warfare and market. End users find physical/criminal mental models to be most accessible [8]. Thus framing risk information such that it is grounded in physical analogues may be much more accessible for end users. Physical mental models can both leverage affect heuristic and also provide grounding of abstract online risks in a more tractable context.

We developed a video to convey information about phishing emails to the user. In it, an older adult is approached by a person claiming to be from the IRS (Figure 1), who has discovered discrepancies in his accounts. He then asks the older adult for information such as his SSN, bank account numbers etc. In the first part of the video the older adult readily gives this information out. The user is then told that the older adult got phished. In the second part of the video the older adult is more suspicious and decides to call the bank before he gives out any information. At this point the "agent" leaves, claiming he needs to attend to other issues. The user is told that the older adult made the right decision this time. The user is then informed that just like the agent, phishing emails can appear to be legitimate and are trying to get to the user's financial data (Figure 2). Here we leverage the story-telling capability of the video and use physical analogies to create mental models easily accessible to users. We also provide alternative responses that they can make when faced with a similar threat.

ACTIVITY RECOGNITION FOR EFFECTIVE COMMUNICATION

We attempt to optimize the timing and extent of warnings by combining plan recognition and user modeling. We develop simple, generic models of common risky tasks, such as paying bills on a bank web site, in order to predict when the user will take a potentially dangerous action and to make a pre-emptive warning. We also build a model of the potential danger of a task based on the state of security of the computer. We combine these processes into one prediction of future actions and their risks using a Markov logic network [16]. The network compiles both the action models and the user model and continually updates the probability that an action in the near future will compromise security.



Figure 1. The older adult with the “IRS agent”

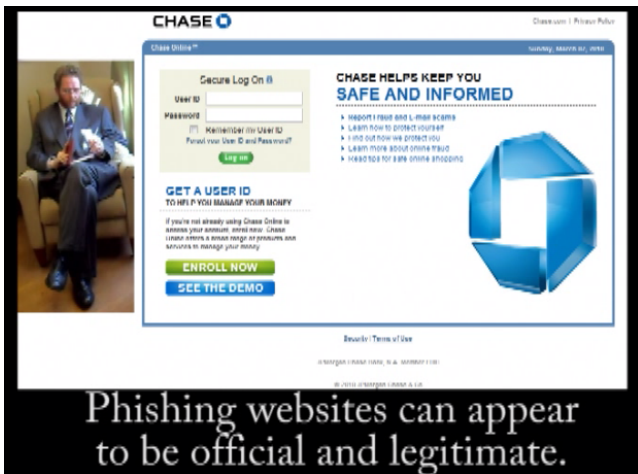


Figure 2. The “agent” is a physical analog of a phishing website

Consider the following scenario. A user takes a short break while balancing his checking account. He clicks on a link in his email, but sees that the page it opens is not the usual one and quickly kills it. However, since he has not recently patched his browser, the site installed a keylogger. He continues to browse a variety of more innocent sites for 30 minutes, and then returns to his banking site and enters his information.

Our tool attempts to save users from such disasters by providing warnings and offering to create patches and clean installations. Our aim is not to develop new security tools, but to make existing tools relatively easy to use and persuade users of their benefits by delivering timely, pertinent warnings. To illustrate this approach, consider two different warnings the user could be given in the scenario above. When the user first follows the link from the email, a warning could be given that the final site, after forwarding, was not a trusted site. However the danger is vague and will occur at some time in the future. e.g. “Warning: this site is not trusted and may have compromised security. It is recommended that you bring your browser up to date before accessing potentially sensitive information.”. The user may

well decide to put this off until after browsing, by which time it may be forgotten.

In contrast, a warning delivered just as the user is about to log onto the bank site is both more timely and can be more specific, e.g. “Warning: you may be about to enter sensitive information in your browser. However, the site you visited from your email with subject “NFL action” was not a trusted site and may be able to pass on this information. It is recommended that you bring your browser up to date and refresh before entering this information. This will take approximately three minutes.” Once the user has viewed an initial video such as the one described in the previous section, the tool will also use key snippets or stills from the video to remind the user of their earlier appreciation of the dangers involved.

It would be very hard to recognize every case where the user is about to access or provide sensitive data. Our approach is to model a number of standard tasks that users perform and attempt to identify the user’s current task and assess its risks. Our base action models are similar to those of a hierarchical task network or reactive planner, e.g. [13]. We translate these models into a knowledge base encoding a Markov logic network (MLN) [16], in order to use observed actions and background knowledge about user activities to predict next steps. A MLN combines elements of logical and probabilistic reasoning. A knowledge base consists of a set of weighted logical formulae that can be viewed as a template for constructing a Markov network. The higher the weight, the greater the likelihood that the formula holds. Our approach is similar to that of Blythe et al. [5], which performs probabilistic abduction by translating weighted Horn clauses into a MLN KB. However, we have tailored the translation to logical descriptions of HTN actions and use prior probabilities on different activities.

In this case, an activity to balance a checking account may have substeps of opening the bank’s page, logging in, accessing the account and finally inspecting each returned check. In order to support reasoning about the next step we recast the procedure as follows:

$W1$ $balanceChecking \ \& \ occurs(openBankPage,N)$
 $\quad \rightarrow \ occurs(logInToBank,N+1)$
 $W2$ $balanceChecking \ \& \ occurs(logInToBank,N)$
 $\quad \rightarrow \ occurs(accessAccount,N+1)$
 ...

Here the symbols W_i refer to the weight given the i^{th} clause in the KB and N is a variable representing the time step. We include a rule that the observed step is probably the step performed and also allow inferring unobserved steps with low probability, so an activity can still be recognized if not all steps are observed. We include rules about the danger of steps in different circumstances, e.g.
 Wd $occurs(logInToBank,N) \ \& \ compromised \rightarrow \ danger$
 with similar probabilistic rules for when the system may have been compromised.

Since the underlying Markov chain is undirected, each observation consistent with an activity such as balanceChecking increases the belief that this activity is present. When the probability of danger reaches a threshold we create a warning for the user. The maximum a posteriori solution to the MLN includes the likely next action and reasons for the tool's belief of imminent danger. This is used to word the warning and select a snippet from a previously-seen video, in this case on phishing.

This strategy to delay warning about potential dangers contains some risk, of course. By waiting as late as possible before a potential vulnerability is discussed, the tool may give the warning too late, having missed an earlier visit to a sensitive site. This risk is outweighed by the greater chance that the warning will be heeded, however.

DISCUSSION

We have described an approach combining probabilistic plan recognition and risk communication to improve the usefulness of security tools by making their warnings timely, specific, graphical and grounded in effective mental models. The contributions of this work include reasoning explicitly about the security consequences of possible user actions and developing a vocabulary of mental models that can be used to inform the user about possible risks. These approaches are independent of particular security tools and can be used in a system that marries a set of open-source tools as appropriate. We currently have an initial implementation of the system and are planning user tests. Our observations of user actions are currently limited to actions taken inside a web browser, such as opening specific URLs.

One advantage of the MLN approach for plan recognition and inferring likely danger is the flexibility of the representation. For example we can easily include information about potential mistakes the user may make, coded as probabilistic consequences of actions that may further compromise security or have other side effects. Activity recognition is more general than simply categorizing URLs by vulnerability and captures context. For example, other visits to the bank site to check interest rates would not trigger a warning. In the long run we aim to include representations of user affect such as tiredness and task urgency and models of bounded rationality to improve our model of the user's reaction to warnings.

ACKNOWLEDGEMENTS

This research was supported by NSF EAGER award 0943382. We thank the members of the security group at Indiana University and the intelligent games group at USC for helpful discussions.

REFERENCES

1. A. Acquisti. Imagined communities: Awareness, information sharing and sharing on facebook. *PETS*, June 2006.
2. A. Acquisti and J. Grossklags. Uncertainty, ambiguity and privacy. In *Fourth Annual Workshop Economics and Information Security (WEIS 2005)*, MA.
3. F. Asgharpour, D. Liu, and L. Camp. Mental models of computer security risks. In *Workshop on the Economics of Information Security (WEIS)*, 2007.
4. V. Bellotti and A. Sellen. Design for privacy in ubiquitous computing environments. In *Proceedings of the third conference on European Conference on Computer-Supported Cooperative Work*.
5. J. Blythe, J. Hobbs, P. Domingos, R. Kate, and R. Mooney. Implementing weighted abduction in markov logic. In *International Workshop on Semantics of Computing*, 2011.
6. L. Camp. Mental models of privacy and security. *Technology and Society Magazine*, 28(3), 2009.
7. L. F. Cranor and S. Garfinkel. *Security and Usability*. O'Reilly, 2008.
8. V. Gargv and J. Camp. How Safe is Safe Enough: Online Version. In *Workshop on Security and Human Behavior*, 2010.
9. C. Herron, H. York, C. Corrie, and S. Cole. A comparison study of the effects of a story-based video instructional package versus a text-based instructional package in the intermediate-level foreign language classroom. *CALICO JOURNAL*, 23(2):281, 2006.
10. M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y. Lim. What instills trust? A qualitative study of phishing. *Lecture Notes in Computer Science*, 4886:356, 2008.
11. P. Johnson-Laird. *Mental models: Towards a cognitive science of language, inference and consciousness*. Harvard Univ Pr, 1986.
12. D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.
13. D. Nau, T. C. Au, O. Ilghami, U. Kuter, J. Murdock, D. Wu, and F. Yaman. Shop2: An htn planning system. *JAIR*, 20:379–404, 2003.
14. M. Pattinson and G. Anderson. How well are information risks being communicated to your computer end-users? *Information Management & Computer Security*, 15(5):362–371, 2007.
15. D. Podszabka, C. Conklin, M. Apple, and A. Windus. Comparison of Video and Text Narrative Presentations on Comprehension and Vocabulary Acquisition. *Geneseo Annual Reading and Literacy Symposium*, 1998.
16. M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*.
17. N. Schroeder and U. Capt. Using prospect theory to investigate decision-making bias within an information security context, 2005.
18. A. Tversky, P. Slovic, and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Social Cognition: Key Readings*, page 167, 2005.