Promoting A Physical Security Mental Model For Personal Firewall Warnings

Fahimeh Raja
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
fahimehr@ece.ubc.ca

Kai-Le Clement Wang
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
w.kaile@alumni.ubc.ca

Kirstie Hawkey
Dalhousie University
Halifax, NS, Canada B3H 1W5
Hawkey@cs.dal.ca

Konstantin Beznosov
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
Beznosov@ece.ubc.ca

Steven Hsu
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
h.steven@alumni.ubc.ca

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
We used an iterative process to design personal firewall warnings in which the functionality of a firewall is visualized based on a physical security mental model. We performed a study to determine the degree to which our proposed warnings are understandable for our participants, and the degree to which they convey the risks and encourage safe behavior as compared to warnings based on those from a popular personal firewall. Initial results show that our warnings facilitate the comprehension of warning information, better communicate risk, and increase the likelihood of safe behavior. Moreover, they provided participants with a better understanding of both the functionality of a personal firewall and the consequences of their actions.

Keywords
Usable security, firewall, warning, mental model

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces-Evaluation/Methodology; D.4.6 Software: Security and Protection-Information flow controls.

Introduction
Even though personal firewalls are an important aspect of personal computer security, little attention has been given to their usability. Prior research [6] revealed that users' interaction with firewalls is mainly limited to

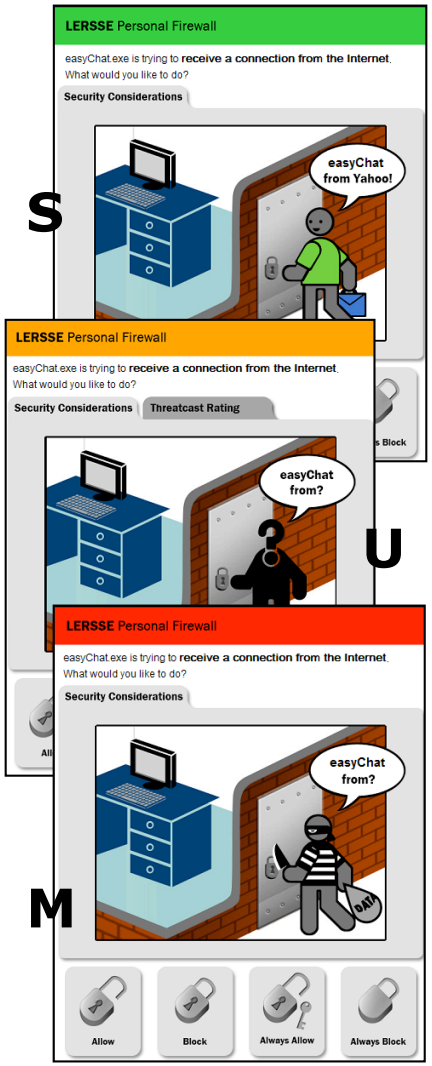

Figure 1: Our proposed warnings, which are based on a mental model of physical security (S:Safe, U: Unrecognized, M: Malicious).

responding to warnings that ask them to allow or block a connection. Thus, it is crucial to design warnings that are understandable for users and properly communicate risk to them so that they can make informed decisions. However, security risk communications to home computer users has been largely unsuccessful to date [3][5][7]. In the warning science literature, one successful technique for risk communication is the mental model approach: a risk communication method based on the recipients' mental model [4]. This approach has been successfully applied in areas such as environmental and medical risk communications, but not in computer security.

Risk communications in computer security have been based on experts' mental models, which are different from non-experts and may not be good for typical users [2]. This gap could lead to ineffective risk communications to non-experts. The goal of our research is to evaluate the mental model approach for firewall warnings. An open question is which mental model to use for this evaluation. Prior research [2][8] shows that the most common mental models of security are the physical security and burglar mental model; for firewalls, the physical security mental model is the closest to both expert and non-expert users' mental models. This suggests that it could be appropriate for risk communication to non-expert users in computer security, particularly for firewalls.

In this paper, we present our initial evaluations of a mental model of physical security in firewall warnings. We used an iterative process in the design of firewall warnings in which the functionality of a personal firewall is visualized based on a physical security mental model and the metaphor of a firewall, a fireproof wall that separates the parts of a building most likely to have a fire from the rest of it. We compared our warnings with warnings based on those from the Comodo personal firewall, which is the most popular personal firewall for both its protection and for its warning design [1]. Our initial results showed that our proposed warnings were more understandable for participants, and they helped them develop a better mental model of the functionality of a personal firewall. They also better communicated the risk and increased the likelihood of safe behavior. Finally, they were preferred by the majority of participants.

Prototype Interface Design

We designed two sets of warnings: our proposed one, which is based on a physical security mental model (P-warnings), and one based on the Comodo warnings (C-warnings). Comodo categorizes applications in three categories depending on the level of risk: safe, unrecognized, or malicious. Based on that, the firewall provides "security considerations'' in its warnings to help users make informed decisions. This is inline with recommendation for designing security warnings [3]. Based on this classification, we designed six interfaces, three for P-warnings: P-safe, P-unrecognized, and P-malicious (Fig. 1.S,U,M), and three for C-warnings: C-safe, C-unrecognized, and C-malicious (Fig. 2.S, U, M).

To design C-warnings, we removed Comodo warning's technical information (i.e., protocol, IP address, and port). As recommended in usable security literature [7] security warnings should be jargon-free. We also removed the recommended action warnings to eliminate the effect of this parameter on the users' intention and allow us to focus on the impact of the mental model in our study.

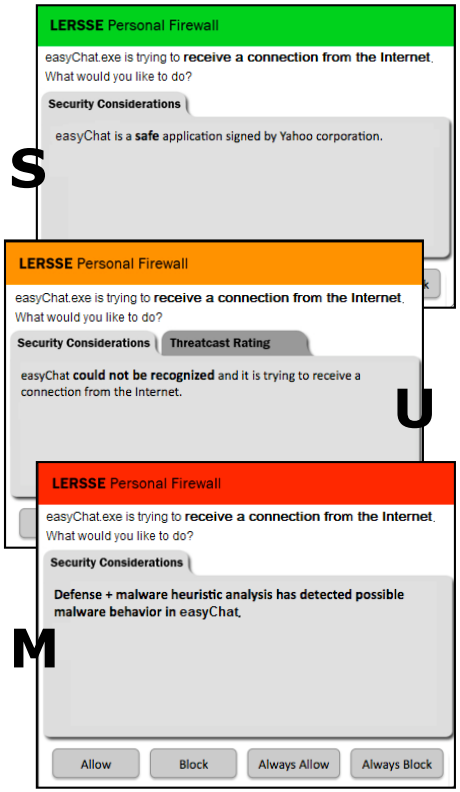

Figure 2: Warnings, which are based on the Comodo's firewall's warnings (S:Safe, U: Unrecognized, M: Malicious).

For P-warnings, we mimicked Comodo's layout. After performing two formative studies, we finalized the design of the warnings. In the final P-warnings, we used a brick wall and a metal door to resemble a physical firewall and a fire exit, which are the actual metaphors for a computer firewall. We added a lock on the door, which was the most familiar metaphor for our formative study participants for controlling access in physical security. For the actions (Allow, Block, Always Allow, Always Block), we added icons corresponding to the lock. One would unlock the lock to allow access once through the door, keep it locked to deny access, give the key to grant permanent access, or have a lock without a keyhole to permanently block access.

We used a figure to represent an application that wants to make a connection through the firewall. We also used a cloud to show the name of the application and its developer (e.g., in Figure 1 it represents Yahoo!). For safe applications, we used a figure wearing a green shirt with a smile on his face. Our formative study found this figure was friendly, trustworthy, and gave participants a positive impression that suggest they should unlock the lock and grant access. For unrecognized applications, we used a black silhouette with a question mark as its head to show that the application is unidentified. For malicious applications, we used a figure dressed in a prisoner's uniform, with a knife and a thief's bag. According to our formative study participants "the message that the interface conveys is very clear: It is very dangerous!"

Methodology

Our research questions were: (1) Do our participants understand what the warnings mean when they encounter them for the first time? (2) What are their misunderstandings or confusions about the warnings? (3) How do the warnings affect their intention to act? (4) Which kind of warnings would participants prefer to have for their personal firewall?

To evaluate comprehension and the degree of initial clarity of the warnings, we used open-ended questions because they provide more information about any sources of confusion and the types of errors people make. To reduce the effects of subjective judgments and increase reliability, we used three evaluators to code the data. As it is common in warning science, we used risk perception to measure intention. There are multiple known contributing variables to risk perception, including level of hazard, likelihood of damage or loss, and severity of potential damage or loss. We used the most common approach for evaluating warnings on each dimension, which is using Likert-type scales ranging from 0 to 7, followed by an interview for clarification. We also used the self-reported likelihood of participants choosing any action as the behavioral intention for performing that action.

Study Design and Protocol

We performed a within-subjects study to compare P-warnings with C-warnings as individual differences could affect users' understandings and intentions making a between subjects study inappropriate. To reduce practice effect, we counterbalanced the presentation order of the warnings. We had two conditions; in the P-C condition, participants saw P-warnings first; in C-P, they saw C-warnings first. We also counterbalanced presentation order of safe, unrecognized, and malicious interfaces in each condition. Initially, we presented all safe, unrecognized, and malicious interfaces of one warning design (P or C)

Table 1: Presentation order of the warnings. P: our proposed warnings based on a physical security mental model, C: warnings based on the Comodo firewall warnings (s: safe, u: unrecognized, m: malicious).

P-C	Order1	PsCs-PuCu-PmCm
	Order2	PsCs-PmCm-PuCu
	Order3	PuCu-PsCs-PmCm
	Order4	PuCu-PmCm-PuCu
	Order5	PmCm-PsCs-PuCu
	Order6	PmCm-PsCs-PuCu
C-P	Order7	CsPs-CuPu-CmPm
	Order8	CsPs-CmPm-CuPu
	Order9	CuPu-CsPsCmPm
	Order10	CuPu-CmPm-CsPs
	Order11	CmPm-CsPs-CuPu
	Order12	CmPm-CuPu-CsPs

Table 2: Six security tasks used to assess participants' security knowledge and expertise.

Tasks
Installing updates
Scanning for viruses, spyware, and other potentially unwanted software
Changing security settings of Internet browsers
Deleting browsing history and cookies
Setting different security controls for different users
Managing browsing add-ons

before the other. However, piloting revealed that presenting one full set of warnings (P or C) to participants primed them about the existence of three levels of risk, which impacted their responses in the second condition. We therefore opted to show one interface (safe, unrecognized, malicious) from one set and then the corresponding interface from the other set. During analysis, we examined participants' understanding for the first interface they saw from each set. Based on this study design, we had 12 presentation orders of the interfaces (see Table 1).

Each participant completed a one-hour session in a meeting room in our department. They first completed a consent form and background questionnaire. This included an assessment of their security knowledge and experience with six tasks taken from the Security Center of Windows Vista (Table 2). We then described a scenario for them to provide context (Table 3). We then presented them with the interfaces using one of the orders in Table 1. After presenting each interface, we evaluated their understandings and risk perceptions. We repeated the same procedure for other interfaces.

Participants

We recruited a diverse set of 60 participants from both the university and general community via messages sent to email lists of several departments in the university and posted on two public online classifieds. Moreover, we posted flyers both at the university and local public places. Table 4 shows participants' demographics. They had a wide range of backgrounds and occupations (e.g., physician, diamond trader, teacher). All except one were daily users of computers, but their expertise varied.

Results

Warning Understanding

When we asked participants to describe what they understood from warnings, most of them started by repeating the text at the top of the warnings. Further assessment of their comments revealed P-warnings made them more aware of the protection of a firewall. With P-warnings, most participants (80%: 23 in P-C, 25 in C-P) explained what would happen if they chose each option: "your computer is presumably safe in a locked up space. This is what your firewall does, now there is this new software which is trying to access the computer through the door, and you have the control of the lock, you can either allow or block it."

With C-warnings, 57% of participants talked about the prevention provided by a firewall; however, most of them had seen P-warnings first (23 in P-C, 11 in C-P). They mainly mentioned that the warning conveyed a similar message as the P-warnings and just emphasized the different presentations of the message (17), and the different levels of risk conveyed (16). From the remaining 26 participants, most (18) read the text and emphasized the bold terms. The rest (8) either had misunderstandings or said they did not understand the warning, why they would get it, and what would happen if they clicked on allow or block.

Some participants misunderstood the warnings. For P-warnings, 9 thought the warning was generated by easyChat, asking if they wanted to chat with someone from Yahoo! (safe), or an unknown (unrecognized) or malicious (malicious) person. Three also thought it was asking them if they want to have connection to the Internet. For C-warnings, besides those who did not understand the warning and those who only read the

Table 3: The context we provided for the participants.

Scenario
Assume that you are in an urgent need of using a chat application that gives you the ability to do video conferencing with four other people in different locations. Assume that you want an application, which provides videos of all the four other people with whom you are chatting. For this purpose, do a research on the Internet, and find an application, called easyChat, with these specifications. You download and install the application and send it to others to start using it. But, when you want to use the application, you get this warning.

Table 4: Participants' demographics.

Condition		P-C	C-P
N		30	30
Age	Mean	31	31
Gender	F	16	14
	M	14	16
Security level	High	2	3
	Med.	8	9
	Low	20	18
Education level	HS	5	6
	BS	13	15
	MS	10	7
	PhD	2	2

text, we had one who thought the message was from easyChat; one who thought there was a problem in his internet connection; and one who thought it was for the security of his Internet connection: "if I allow it, my Internet is safe, and I can check my emails securely."

Risk Perception and Intended Action

To examine if our within-subjects study design had affected our results, we performed a preliminary analysis. A two-way ANOVA with two between-subjects factors: warning type order (P-C, C-P) and threat order (SUM, SMU, USM, UMS, MSU, MUS) did not reveal a significant effect for any of the factors on participants' risk perception and intended action. We also compared participants' risk perception and intended action when they saw each interface first (before seeing any interface) and last (after seeing all the interfaces). Our results showed no significant difference between the first and last exposure to each interface.

We conducted a one-way repeated measures ANOVA to evaluate the effect of warning type (P-warnings or C-warnings) on participants' risk perception and intended action. For safe applications, there was no significant effect for warning type on participants' perceived level of hazard, likelihood of damage or loss, severity of the potential damage or loss, and likelihood of allowing or blocking the program. For both warnings, participants' perceived level of risk was appropriately low and they were more likely to allow the program (See Table 5).

For unrecognized applications, we found a significant main effect for warning type on participants' perceived level of hazard ($F(1,59)=7.79$, $p<.01$), likelihood of damage or loss ($F(1,59)=10.01$, $p<.01$), severity of the potential damage or loss ($F(1,59)= 4.88$, $p<.05$), and probability of allowing the program ($F(1,59)=8.62$, $p<.01$). With C-warnings, participants' risk perception was lower, and they were more likely to allow the application than with P-warnings (See Table 5). For malicious applications, warning type had a significant effect on participants' perceived level of hazard ($F(1,59)=36.00$, $p<.01$), likelihood of damage or loss ($F(1,59)=37.74$, $p<.01$), severity of the damage or loss ($F(1,59)=28.53$, $p<.01$), and probability of allowing the program ($F(1,59)=10.54$, $p<.01$). These results show that P-warnings convey more risks to the participants than C-warnings (See Table 5).

Warning Preference

Most participants (40) preferred P-warnings; 13 noted P-warnings provide a mental model of the functionality of a firewall: "this one (C-warning) is just a warning of a firewall. [The] brick wall and the locked door is very good. It tells me the theory of the firewall" (P60). Some found P-warnings more intuitive (16), and easier (37) and faster (9) to understand. P38 also noted when "multitasking, [such as] talking to your friends, this [P-warnings] is very effective. It tells you everything at a glance, you make less mistakes.'' Some (11) said that P-warnings more clearly convey the risk and the consequences of allowing the program. Twelve thought P-warnings would grab attention better. Five also stated they are more universal: "this one [P-Warning] is better, especially for old people that cannot see clearly or children that may not understand security, or those who do not know what a firewall is" (P19)."

Only one third of our participants (20) preferred C-warnings. Most of them thought C-warnings were more professional (11) and they would take them more seriously (4). They also found C-warnings more

Table 5: Participants risk perceptions in a scale of 0..7, and the probability of choosing each action. P our warnings based on a physical security mental model, C: warnings based on the Comodo firewall warnings.

	Warning	P	C
Safe	Hazard level	1.48	1.33
	Loss	1.52	1.27
	Loss severity	1.83	1.43
	Allow	65.6	53.2
	Block	8.7	10.0
	Always allow	21.7	35.3
	Always block	4.0	1.5
Unrecognized	Hazard level	3.75	3.17
	Loss	3.64	3.00
	Loss severity	3.58	3.13
	Allow	50.0	62.7
	Block	28.0	22.7
	Always allow	5.5	5.3
	Always block	16.5	9.4
Malicious	Hazard level	6.12	4.73
	Loss	6.03	4.60
	Loss severity	5.95	4.72
	Allow	11.4	24.8
	Block	41.5	43.3
	Always allow	3.8	2.6
	Always block	43.2	29.2

informative (7) and descriptive (2). Five noted they understood C-warnings better. Two said C-warnings are more specific; they thought people could have different interpretations for P-warnings.

Discussion

Our findings suggest that supporting a mental model of physical security in firewall warnings could be a promising approach for improving users' understanding of the warnings, their perceived level of risk, and intention for safe behavior. Our results show that applying known metaphors, such as the bandit figure, in the warnings is very effective in conveying risk to the user. Our participants could relate the potential risks of the warning to the risks from the physical world; this resulted in better understanding the consequences of their potential actions. However, we did have participants who mentioned they would take our warnings less seriously than the textual warnings.

We conducted a post-hoc analysis of the one third of our participants who preferred C-warnings to our warnings. Their demographics showed that all but one had a high or medium level of security knowledge, indeed all participants with a high-level security knowledge were in that group. These results indicate that warning design may need to be customizable for different groups of users. We are doing a more in-depth analysis of our data to find out: (1) how participants' demographics affect their comprehension and intention, and whether they affect their warning preference, (2) if there is any relationship between participant understanding of warnings and their intention to act. Further research is also required to determine participants' actual behavior in response to our proposed warnings rather than their stated intentions.

Conclusions

We presented a study in which we evaluated a novel approach for designing personal firewall warnings. We used an iterative process to visualize the functionality of a personal firewall based on a mental model of physical security and the metaphor of a physical firewall. We compared our warnings with those designed based on the warnings of one of the most popular personal firewalls. Our initial results showed our warnings were more untreatable for participant; they helped them develop a better mental model of the functionality of a firewall. They better communicated the risk and increased the likelihood of safe behavior.

References

[1] 2010 personal firewall software review. http://personal-firewall-software-review.Toptenreviews.com/

[2] Asgharpour, F., Liu, D., and Camp, L. J. Mental models of security risks. In *Proc. USEC 2007*, 367–377.

[3] Egelman, S., Cranor, L., Hong, J. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proc. CHI 2008*, 1065–1074.

[4] Morgan, M. R*isk communication: A mental models approach*. Cambridge Univ. Press, 2002.

[5] Motiee, S., Hawkey, K., Beznosov, K. Do windows users follow the principle of least privilege? Investigating user account control practices. In *Proc. SOUPS 2010*, 1-13.

[6] Raja, F., Hawkey, K., Jaferian, P., Beznosov, K., Booth, K. It's too complicated, so I turned it off!: expectations, perceptions, and misconceptions of personal firewalls. In *Proc. SafeConfig 2010*, 53-62.

[7] Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., Cranor, L. Crying Wolf: An empirical study of SSL warning effectiveness. In *Proc. USENIX 2009*, 399–432.

[8] Wash, R. Folk models of home computer security. In *Proc. SOUPS 2010*, 1-16.