Poster: Validating and Extending a Study on the Effectiveness of SSL Warnings

Andreas Sotirakopoulos, Kirstie Hawkey, Konstantin Beznosov

University of British Columbia
Vancouver, Canada
{andreass,hawkey,beznosov}@ece.ubc.ca

1. INTRODUCTION

We recently replicated and extended a 2009 study that investigated the effectiveness of SSL warnings. The original study was conducted at CMU by Sunshine et al. [2], and we will refer to it as the CMU study. As in the CMU study, we required participants to perform a series of tasks; and we observed their reactions to SSL warnings that were presented to them. After they completed the tasks, we asked them to complete an online questionnaire where we asked about their reasoning behind their actions during the study's tasks. We designed our experiment in such a way so as to mitigate some of the limitations of the prior work, including allowing participants to use their web browser of choice and recruiting a more representative user population.

Our research is a work in progress, thus data collection and analysis are still underway. However, our preliminary analysis indicates interesting findings and differences with the findings of the CMU study. We have observed an impact of components we introduced in the study (i.e., broader population and usage of the browser of participant's choice) on the results. We plan to have completed data collection and analysis by the time the poster session will be held.

2. EXPERIMENTAL DESIGN

Our study is a between subjects experiment with four conditions based on the warning presented and browser used by the participant. The four initial conditions are as follows:

- Firefox 3.5 browser presenting its native SSL Warning
- Firefox 3.5 browser presenting an SSL warning designed by us
- Internet Explorer 7 (IE7) presenting its native SSL Warning
- IE7 presenting an SSL warning designed by us

Following the CMU design we did not want participants to be primed for security, so we did not reveal our study purpose during the recruitment and laboratory tasks. Instead, we advertised our study as one investigating the difficulties people face while trying to retrieve information online. For that purpose, we had four tasks: 1) find the total surface area of Greece in square kilometers, 2) report the last two digits of their bank account balance, 3) retrieve the price of the book, and 4) create a new email account. The first and third tasks were dummy tasks that were not relevant to our study and were present only in order to obfuscate further the real purpose of the study. In addition, each task had a primary and a secondary way to retrieve information so as to avoid a task focus effect on participant's actions.

Our experiment was designed so it would mitigate some of the limitations of the CMU experimental design, both these acknowledged in the CMU study and those which we felt should be addressed. The first limitation we identified was that participants in the CMU study were drawn almost exclusively from the CMU student body. The second limitation was that participants were randomly assigned to the browsers investigated. This might have caused them to alter their normal behavior and become more cautious about SSL warnings as the warning interface was unfamiliar to them. Finally, in the CMU study, custom warnings designed for IE7 were radically different in colors, wording, and layout from the native IE7 warnings. We believe this unfamiliarity may also have elicited a more cautious reaction to warnings.

In an effort to mitigate these limitations we sought participants in the broader Vancouver population instead of limiting ourselves to UBC students. In order to limit the surprise effect a previously unseen browser interface and warning would have, we assigned users to our conditions according to the browser they normally used and redesigned the custom SSL warnings that were presented to the users. Moreover, we needed to substitute the CMU study's task of having participants accessing the CMU library to retrieve the call number of a book. We could not merely change this to the UBC library due to the diversity of our participants. We instead had them create an email account at Hotmail as it is a similar task in that it poses a smaller threat to the personal data of the participant than the banking task.

3. RESULTS

Our goal is to recruit 20 participants in each of the four conditions, but recruitment has been challenging. While 174 individuals have responded to our recruitment notices, to date 75 participants have taken part in our study. When scheduling the study session, we send a second email revealing that they will need to use their bank credentials so as to ensure that they arrive at the study able to perform the tasks. Sixteen potential participants explicitly stated that they were no longer interested in participating because they feel that their personal information will be at risk.

Our participants are equally divided by gender and have a broader age distribution than in the CMU study where 98% of all participants where CMU students and all between the age of 18 and 30, as shown in Figure 1. Although we have a younger population when compared to Canadian statis-

	FF3	FF3 custom	IE7	IE7 custom
CMU	55% (11/20)	N/A	90% (18/20)	45% (9/20)
UBC	79% (15/19)	84% (16/19)	70% (14/20)	70% (12/17)

Table 1: Percentage of participants in both studies who chose to ignore the SSL warning in the bank task.

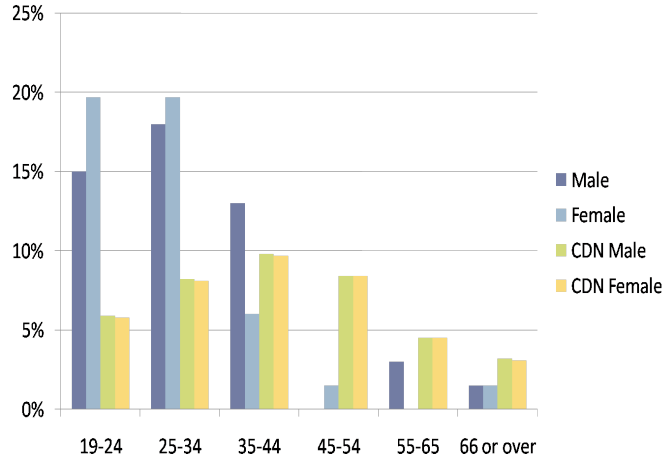

Figure 1: Gender age distribution in UBC study, compared to the general Canadian statistics

tics of internet usage, we believe that we have achieved an adequately distributed population in terms of age groups, particularly given the difficulty of recruiting working non-student participants. Our participants are moderately technically sophisticated, scoring a 3.16 on a 1 to 5 Likert scale asking them to self evaluate their technical skills based on how often they ask for help from others with computers (1) or others ask for help from them (5).

The data we have collected to date indicate that there are differences in the way participants respond in the two studies in the banking task (see Table 1). The percentage of our participants who ignored the native warning in Firefox 3 is larger than in the CMU study, whereas in the IE7 case, fewer participants ignored the warning in our study. We believe that these differences are due to the changes we have introduced to the experimental design of the CMU study. We have recently added a fifth condition of assigning respondents to use an IE7 browser regardless of their regular browser of use to tease out whether the differences we are observing are attributable to random assignation of participants to web browsers in the CMU study.

4. DISCUSSION

Our preliminary analysis indicates that our choice to use a broader, non-student population and also having participants using an interface that they are accustomed to has had an impact on the reactions we observed to the warnings. It also confirms that solely altering the warning's wording and colors is not sufficiently effective to deter participants from ignoring the warnings. Habituation to the warning's layout and presence seems to play a major part in the decision to ignore or obey it. In addition, we have also observed that although almost 50% of participants understood what actions the warning was urging them to take (i.e., not to continue) they ignored the warning. This behavior is evident in how users in every day tasks consider security. Even if they understand the messages they receive, security is a secondary issue compared to their primary task of retrieving or viewing the information they are looking for.

As we move ahead with our analysis, we must also consider the impact of the study design on participant's "normal" behaviours. After the study tasks, participants completed a questionnaire and we probed their reasoning behind heeding or ignoring the warnings. As discussed in our USER workshop paper [1], the responses of several participants indicated that their behaviour may have be influenced by their trust in us to not put them in a dangerous situation during a study approved by UBC's research ethics board.

5. CONCLUSION

We are currently replicating a CMU study on the effectiveness of SSL warnings. While analysis is not yet complete, there appear to be some differences in our results, which we believe are relevant to our choice to differentiate components of that study in order to introduce greater realism in the study design. However, there are also questions about whether our participants truly believed that they were at risk during the study tasks as they were taking part in a study approved by UBC's research ethics board.

Once we finish recruiting participants, we will perform a comparative analysis with the original CMU study results and identify areas where our results differ. We will also investigate whether any differences are attributable to the impact of using a broader population and maintaining increased ecological validity in the experimental procedures. Finally, we will analyze our results in light of comments made by participants about their trust in us in order to determine whether or not our results are truly indicative of behaviours in the wild. We hope that our findings will not only provide us with a basis for making recommendations on ways to make SSL warnings more effective in maintaining user's information security, but will also provide methodological insights valuable to others evaluating similar types of usable security problems.

6. REFERENCES

[1] A. Sotirakopoulos, K. Hawkey, and K. Beznosov. "i did it because i trusted you": Challenges with the study environment biasing participant behaviours. In *SOUPS Usable Security Experiment Reports (USER) Workshop*, 2010.

[2] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying Wolf: An empirical study of SSL warning effectiveness. In *Proceedings of 18th USENIX Security Symposium*, pages 399–432, 2009.