Does MoodyBoard Make Internet Use more Secure? Evaluating an Ambient Security Visualization Tool

Alexander De Luca[1], Bernhard Frauendienst[1], Max-Emanuel Maurer[1], Doris Hausen[1],
Julian Seifert[2], Niels Kammerer[1], Heinrich Hussmann[1]
[1]Media Informatics Group, University of Munich [2]MHCI, University of Duisburg-Essen
{alexander.de.luca, max.maurer, doris.hausen, hussmann}@ifi.lmu.de
{frauendi, kammerer}@cip.ifi.lmu.de, julian.seifert@uni-due.de

ABSTRACT

Internet users are targets for ever-advancing phishing- and other attacks. The risks are, for example, to disclose credit card information or passwords to unauthorized instances. One approach to help users with insecure situations is provided by MoodyBoard, which uses ambient information to highlight potential risks. In this paper, we present findings from an evaluation of this system. Two user studies were conducted in order to find out whether an ambient security tool can protect users during sensitive tasks. We designed a pilot study to find out whether users understand the warnings and a security study to see if it helps to protect users from phishing attacks. Results show that MoodyBoard users behaved significantly more secure.

Author Keywords

Security, Awareness, Ambient visualization, MoodyBoard.

ACM Classification Keywords

H5.2. [Information interfaces and presentation (e.g., HCI)]: User Interfaces – Evaluation/methodology.

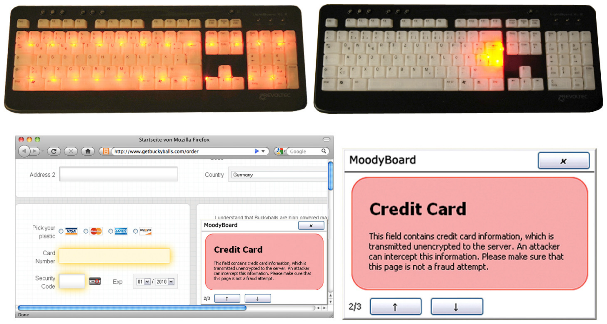
General Terms

Experimentation, Human Factors, Measurement.

INTRODUCTION

Frauds on Internet services like phishing are widely covering the media these days. Scams are improving every day but simple attacks already do the trick. From habituation effects [2] to overlooking important warnings [10] there are different reasons why people fall for phishing. Lack of interest in security [9], wrong mental models [4] or lack of knowledge [7] can open security holes. We cannot expect users to solve these problems. We have to work with them together to do so [1].

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
Copyright 2011 ACM 978-1-4503-0267-8/11/05...$10.00.



Credit Card

This field contains credit card information, which is transmitted unencrypted to the server. An attacker can intercept this information. Please make sure that this page is not a fraud attempt.

Figure 1: Top: The MoodyBoard prototype can glow in arbitrary colors. The return key can be lit seperately. Bottom: Message window after pressing the help button.

Current work on supporting users to make secure decisions when using Internet services often focuses on improving warnings. They are either blocking – forcing the user to decide upon an action [5] – or non-blocking, leaving it to the users to heed the warning or not [5,10]. While blocking warnings can become quickly annoying and fail due to habituation effects, non-blocking warnings are often simply overlooked or not understood. Another approach is teaching users to behave more securely and identify threats [6].

In this work, we introduce a fourth approach. Instead of relying on teaching, blocking or non-blocking warnings, ambient security notifications are used. Ambient notifications could be categorized as non-blocking warnings. However, there is one major property that sets them apart: a non-blocking warning can usually only occupy limited space of the user's screen. This limitation does not exist for ambient information. Thus we can use a very intense warning while not actively blocking the user's current task.

As the tool of choice, we decided to use MoodyBoard, which has been developed in several recursive steps to support users in security sensitive tasks [3]. In short: MoodyBoard is a keyboard that can glow in arbitrary colors (see figure 1, top). The concept is promising but has not been evaluated yet. In this work, we therefore provide an evaluation of the concept based on two consecutive user studies. The results of the studies show that this kind of ambient warning can significantly improve secure behavior.

MOODYBOARD

The prototype is a stripped down version of a Revoltec Lightboard XL standard layout keyboard, extended with strips of RGB-SMD-LEDs, which illuminate the whole key area (figure 1, upper left). The Return key is separately lit (figure 1, upper right). The wrist rest holds a small vibration motor. The original backlight switch was remodeled into a help button. Pressing the help button displays a message containing information on why the specific notification was triggered (figure 1, bottom). The LEDs allow us to change the luminosity and colors in arbitrary ways, which is necessary since only hard changes of those values are likely to get the user's attention [8].

A Firefox extension connects the hardware with the browser interface. It allows for MoodyBoard notifications to be "attached" to DOM-events on arbitrary HTML elements using XPath expressions. The improved version for the security study additionally displays messages for each notification when the help button is pressed. For more information on MoodyBoard, please refer to [3].

PILOT STUDY

It is essential for security notifications that the user understands them in order to work. Therefore, a first user study was conducted to gather information about the participants' interpretations of MoodyBoard notifications and whether they match their actual meaning.

Study Design and Procedure

The physical setup resembled a standard desktop environment and consisted of an LCD screen, mouse and MoodyBoard (without help button) connected to a laptop.

The study involved four tasks whose order was perfectly counterbalanced. For each, the participant had to complete a task on a web page, three of them included payment and entering credit card information, the fourth was to enter a comment. One of the payment sites was SSL encrypted and resulted in green light upon focusing the credit card field. The other pages were unencrypted and showed red illuminations when sensitive data was entered or submitted, with exception of the "comment" task, which also showed green light due to the non-sensitive nature of the data.

The study was set up in an empty room. Each participant was handled separately. After entering the room, they were provided a small introduction to the scenario and a list of tasks already in the appropriate order. During each task, after the MoodyBoard feedback was triggered, users were asked to share their opinion as to why this particular feedback was provided in this situation. They also had to rate its appropriateness on a five-point Likert scale (1="not appropriate at all", 5= "very appropriate"). Their answers were written down and matched against two central points – encryption status and sensitivity of data – in order to calculate to which extent they coincided with the actual implementation. Finally, after completing the tasks, participants were asked to fill out a questionnaire.

Participants

24 participants (eight female) were recruited. Their average age was 25 years (21 - 31). About two thirds were students with a background in computing science, the rest had a diverse background.

Results

The results showed that users often had problems interpreting the feedback MoodyBoard gave. We calculated an overall match ratio of 35%. Match ratio refers to the intersection of the users' interpretation and the actual technical reasons for a notification. Even more, only five participants had a perfect match, each on exactly one task. On the other hand, all users recognized the red messages to be warnings, and over half associated them with the security of private data.

In the post-study questionnaire, users voted average on a five-point Likert scale that the information given from MoodyBoard was sufficient (Median=4). However, the need for more information was rated only slightly less important (Median=3).

Essentially, our main findings from this study are that MoodyBoard is capable of attracting the user's attention, but more – preferably context-sensitive – information needs to be provided to produce a sensible reaction. That is, a help button as envisioned in the concept seems highly necessary. Additionally, (red) warnings were not only more efficient, they were also interpreted more accurate. That is, we decided to abstain from positive notifications since they caused more confusion than adding benefit.

SECURITY STUDY

The security study took place at our premises and had a similar physical setup as the pilot study. The main goal was to find out whether MoodyBoard can provide appropriate feedback to make Internet use more secure.

Study Design and Procedure

At the beginning of the experiment, the participants got an introduction to the study which was disguised as an Internet surfing behavior experiment. Before starting with the practical part, the users were asked to fill out a pre-questionnaire, which contained questions regarding their Internet expertise. Based on these, the participants were either marked as experts or not and were assigned to the two groups respectively. This allowed for an equal amount of experts in both groups.

The investigator asked the participants to perform some urgent tasks for a "good friend" who had to go to his grandmother's funeral. The tasks were done "at the friend's home" whose computer was equipped with MoodyBoard (not in the control group). Four different tasks, together with the required input data, were handed out to the participants in written form. Two tasks required passwords (for an auctioning site and a payment service), the other two the use of credit card information (payment for a travel and a gift order). For each participant and each data type, one website was randomly assigned as being a phishing site. These used slightly modified URLs and were unencrypted. All sites, phishing or not, were hosted on our own servers (spoofing the URLs) to avoid connectivity problems.

To the experimental group, MoodyBoard was introduced as being able to notify about different situations (security was never mentioned) and that the reasons for the notifications could be looked up by pressing the help button. In this group, the use of unencrypted websites in combination with sensitive data triggered warnings: red glowing keyboard for password and credit card fields and a red blinking return key plus vibration for submit buttons. Alarms were triggered whenever an input field got the focus (e.g. was clicked) or the submit button was hovered by the mouse pointer. Pressing the help button displayed a warning that data would be submitted over an unencrypted channel and that the page should be checked for integrity. As opposed to the pilot study, "secure" situations did not trigger notifications. The interaction was filmed with a camera for later analysis (mainly verifying if phishing websites were identified or not). In the end, the investigator briefed all participants in both groups about the experiment and asked them to fill out a final questionnaire containing questions related to phishing.

We applied a mixed-model design with two groups. The between-group variable was *MoodyBoard* (yes or no). Within the groups, a repeated-measures design was used. The independent variables were *data type* (password and credit card number) and *phishing* (yes or no). The dependent variable measured was *security* (number of identified phishing websites). To minimize learning effects, a 2x(4x4) Latin square was used. The "2x" refers to the fact that firstly the phishing sites were evenly assigned to a 4x4 block, which was then duplicated, inverted, and its last two rows and columns swapped. This required a minimum of eight users per group. For each possible setup, a bookmark set was provided (and loaded) that had to be used by the participants.

The goal was to measure whether MoodyBoard helped to identify more phishing websites. The only condition under which a phishing website was counted as recognized was when the participant aborted the task. This was possible since in the introduction, it was emphasized that the personal data should be treated with care as if it was a real situation. The only means for the control group to identify a phishing website were standard browser methods: certificates were displayed or not and the URLs could be checked. Thus, the main hypothesis was that MoodyBoard users would identify significantly more phishing websites than users in the control group.

Participants

We recruited 32 participants, 16 per group. Group 1 (control) had an average age of 23 years (21 – 27, eight female), group 2 (MoodyBoard) of 24 years (22 – 27, seven female). Having 16 users per group allowed for perfectly applying the Latin square design as described in the previous section. Expert users were mostly computer science students, while the other participants were diverse. Not only did both groups include the same number of experts, additionally there were equally as many as non-experts (eight in each group).

Results

Phishing

Overall, 24 (38%) out of the 64 phishing sites were identified. 18 (56%) out of 32 in the MoodyBoard group were found, compared to only six (19%) in the control group. In nine instances, participants aborted the task after using the help button and reading the explanations. There was no difference between experts and non-experts (twelve phishing sites each). Most identified phishing websites were in the credit card conditions (16 in total; eleven of them with MoodyBoard) while only eight (seven with MoodyBoard) password phishing sites were identified. Taking a closer look at the data reveals that only users that found the credit card phishing site were able to identify the password phishing site. Additionally, only one user in the control group found the password phishing site. MoodyBoard performs better than what we know from current non-blocking warnings [5] and similar to advanced blocking warnings [7] while not interrupting the current workflow of the user.

A mixed-model ANOVA confirmed these results. It revealed a significant main effect for the between-group variable *MoodyBoard* ($F_{1,30}$=9.956; *p*=.004). This means that participants in the MoodyBoard group were significantly more likely to identify a phishing website which supports the main hypothesis. The main effect of *data type* was highly significant as well ($F_{1,30}$=16.351; *p*<.001; Greenhouse-Geisser corrected). Therefore, the results that more credit card phishing sites were identified did not come up randomly. This is very interesting since the password-protected sites in the study were both highly sensitive too (e.g. online payment). It seems that users still considered credit card information as more sensitive. This might explain, why, in contrast to password sites, five people found the credit card phishing sites without MoodyBoard support. Finally, no significant interaction effects were found (all *p*>.05).

The questionnaire (experimental group) gives further insights. Even though usefulness of MoodyBoard notifications was rated averagely (Median=3), the participants that correctly identified phishing websites rated it very high (most 4 or 5). Furthermore, all four participants that did not use the help button rated usefulness very low (1,1,1,2). This supports the finding of the pilot study, that this button is essential. It is thus not surprising that some participants mentioned that the help messages should always be displayed at least for the first visit of a website.

False positives

Another interesting finding was on falsely identified phishing websites. Despites happening only three times, all cases occurred in the MoodyBoard group and all for credit card. As mentioned before, genuine websites did not trigger any notifications. So why did these false positives occur? In one case, a user stopped the task (which we defined as "phishing identified") since she was "*not sure whether this is really the trip my friend wanted to book*". The other two were users that also found the real phishing sites. Finding them might have made them suspicious with respect to phishing attacks in the experiment.

Why MoodyBoard users fell for phishing attacks

Finally, we wanted to know what reasons made users fall for phishing attacks, especially in the MoodyBoard group. One user stated that she recognized the wrong URLs but thought they were necessary for the experiment (see limitations section). More interesting are reasons based on wrong interpretation of the notifications. As stated in the study description, participants were only briefed that MoodyBoard would give them feedback during surfing the Internet. Security was not mentioned at all. Thus, three users interpreted the warnings with respect to the current task. For instance, two participants thought that the red light indicated a mistake during input (e.g. a spelling mistake). Another user was sure that the glowing enter button told him not to forget to submit the form. None of them used the help button, which avoids such misinterpretations. We argue that these situations would not happen in a real situation, in which a user knows what MoodyBoard is designed for. Still, evaluating a "worst case" scenario reveals the most interesting findings.

Limitations of the Results

Testing the security of a system in a lab study is rather hard. Users tend to behave more carelessly and do not feel in real danger, which is supported by the artificial setup. It is likely that the results for both groups are affected by this fact. It is also hard to hide the real meaning of a study and we cannot say for sure that none of the users knew what it was really about. Nevertheless, the experiment is a good indication on how MoodyBoard or a similar ambient notification system can unobtrusively support users in security relevant tasks.

The lack of a condition using only software notifications in the experiment only allows for theoretically comparing the system to such approaches. Thus, we can only state that the current MoodyBoard concept is a valuable tool to protect users but we cannot say how much more beneficial it is compared to a software only solution.

Even though neither in the pilot nor the security study did participants overlook the ambient notifications, we did not explicitly test required intensity, or compare different setups for notifications. Thus, we cannot state which parameters make an ambient security notification visible and "important enough" for users to interrupt their tasks.

CONCLUSION AND FUTURE WORK

In this work, we presented an evaluation of MoodyBoard based on two consecutive user studies. The results show that MoodyBoard positively influenced secure behavior even though misinterpretations of notifications occurred. In an informed user group however, such interpretations as seen in the security study, are less likely.

The main advantage of MoodyBoard is that it does not block any screen real estate but at the same time can deliver very intense, noticeable non-blocking warnings without interrupting the user's current task. The notifications are also visible to anyone nearby which might raise privacy issues in shared working environments (like offices). For instance, repeated red glowing keyboards might raise suspicion about colleagues.

For future work, it would be interesting to see how MoodyBoard performs in a real world setting. However, it is extremely hard or impossible to measure its efficiency in such a setting. Nevertheless, qualitative results of a real world long-term study could be very insightful.

REFERENCES

1. Adams, A. and Sasse, M. A.. Users are not the enemy. Commun. ACM 42, 12 (Dec. 1999).
2. Amer, T.S., Maris, J.B. Signal words and signal icons in application control and information technology exception messages – hazard matching and habituation effects. Tech. Rep. Work. Paper. Series 06-05, Northern Arizona University, Flagstaff, AZ, October 2006.
3. De Luca, A., Frauendienst, B., Maurer, M., and Hausen, D. 2010. On the design of a "moody" keyboard. *In Proc. DIS '10*. Aarhus, Denmark, August 16 - 20, 2010.
4. Dhamija, R., Tygar, J.D., Hearst, M. Why phishing works. *In Proc. CHI '06*. Montréal, Québec, Canada, 2006.
5. Egelman, S., Cranor, L. F., and Hong, J. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. *In Proc. CHI '08*, Florence, Italy, 2008.
6. Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., and Hong, J. Teaching Johnny not to fall for phish. *ACM Trans*. Internet Technol. 10, 2, Article 7, June 2010.
7. Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., Cranor, L.F. Crying Wolf: An empirical study of SSL warning effectiveness. *In Proc. USENIX '09*.
8. Ware, C. Information Visualization, Second Edition: Perception for Design. Morgan Kaufmann, 2004.
9. Whitten A., Tygar J.D. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. *In Proc. USENIX '99.*
10. Wu M., Miller R.C., Garfinkel S.L. Do security toolbars actually prevent phishing attacks? *In Proc. CHI '06*. Montréal, Québec, Canada, 2006.