Towards the Systematic Development of Contextualized Security Interventions[1]

Steffen Bartsch
CASED – TU Darmstadt
Darmstadt, Germany
steffen.bartsch@cased.de

Melanie Volkamer
CASED – TU Darmstadt
Darmstadt, Germany
melanie.volkamer@cased.de

Current warnings during daily Web browsing demonstrate how difficult it is for developers to craft precise and comprehensible security interventions. While researchers have found that personal contextualization of interventions help in security-critical applications, taking this approach leads to an overwhelming range of options of how and when to intervene as well as which factors to consider. To make contextualized security interventions feasible, we need to support developers in selecting the relevant factors for their applications and support them in deriving the appropriate intervention strategy and content. In this paper, we propose a security intervention framework and methodology which provides such a support.

Security intervention, security communication, usability engineering

1. INTRODUCTION

Warnings of self-signed certificates in Web browsing are an example of how difficult it is to craft precise and comprehensible security interventions[2]. These warnings occur independently from the user's intention like browsing for information (a low risk) or transferring money (a high risk). This imprecision results in habituation that might cause users to ignore warnings in critical situations, since they have ignored them several times without any negative consequences. The problem here lies in the *precision* of this particular security intervention: The warning about the security of the connection should ideally only occur if there is a risk for the user from continuing to use the Web service as intended. Moreover, the warning is formulated in a technical language that is not comprehensible by the user: The inadequate *content* of the intervention prevents the user from understanding the risks of continuing to the Web site.

Several studies have shown that common Web browsing warnings (certificate warnings, Sunshine et al. (2009)) as well as other security indicators (passive indicators, Whalen and Inkpen (2005)) are not effective. The ineffectiveness is caused by the traditional approaches on security interventions that take the form of generic hazard warnings: warn a broad audience with static texts and symbols (Wogalter 2006). Accordingly, researchers propose to personalize and contextualize security indicators (De Keukelaere et al. 2009). The idea is to employ additional information on the context (e.g. user intention) and the user (e.g. expertise) so as to make better decisions on when to warn, how, and with what content.

However, this is challenging for developers of security-critical applications, because they need to take into consideration both the contextual factors and user characteristics in order to implement the correct *intervention strategy* (whether, when, and how to intervene) and *intervention content* (what content to convey). Developers need to evaluate available contextual factors, particularly for their availability and impact. They also have to combine the factors and balance whether the risks justifies a blocking warning or whether the negative effects (habituation, annoyance) are too high (cf. Böhme and Grossklags 2011). De Keukelaere et al. (2009) proposed an architecture for contextualized warnings that evaluates factors so as to decide which type of warning to display, but did not provide a methodology to select the relevant factors. Moreover, the content of the intervention should relate to the user to make the warning more convincing – for example, by taking the mental model into account (Bravo-Lillo et al. 2011).

[1]The work presented in this paper is supported by funds of the Federal Ministry of Food, Agriculture and Consumer Protection (BMELV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme.

[2]We consider security interventions as signals to humans that influence security-relevant decisions, e.g. a green location bar (positive intervention) or warnings (negative intervention).

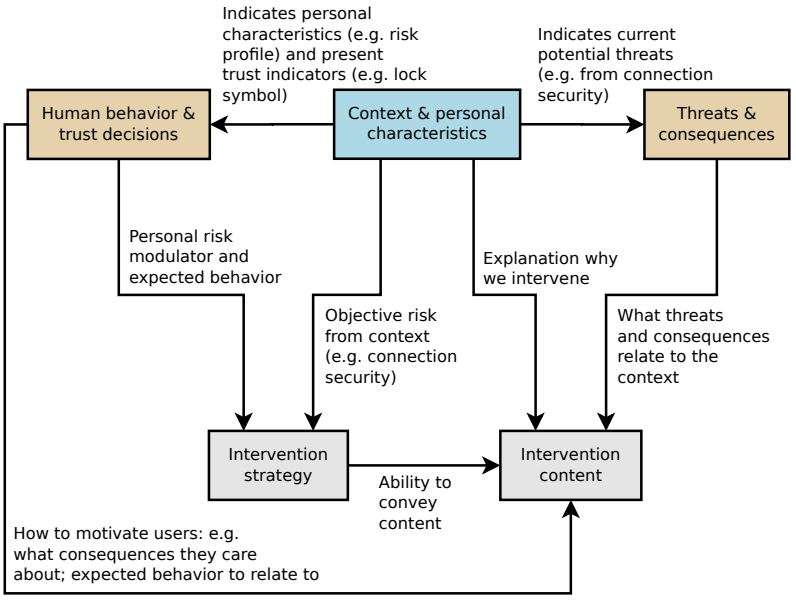

Figure 1: *Intervention framework with influences of the factors on intervention strategy and content*

One part of this challenge is to assess which threats and consequences are present in the actual situation, and whether to emphasize the technical threat (e.g. man-in-the-middle attack) or potential consequences (e.g. financial losses).

Developers need to combine different kinds of interventions (e.g. passive symbols, active warnings) for optimal results so that developers need broader support that considers a range of intervention options. However, prior research on interventions primarily focused on individual (types of) interventions. The broad Human-In-The-Loop framework (Cranor 2008) shows many factors that influence the effectiveness of security interventions, but remains on a descriptive level of theory: The framework only describes the factors, but does not guide the developers on how to arrive at an intervention.

To better support developers in future, we propose an intervention framework that relates the intervention strategy and content to the context and human factors of a corresponding situation as well as knowledge bases of human behavior and threats (cf. Figure 1). We further present a methodology to derive the relevant factors and knowledge bases, which employs user studies, literature reviews, and expert consultations. While applicable in various domains, we focus on Web browsing as one important application area when giving examples in this paper.

2. PRIOR RESEARCH ON SECURITY INTERVENTIONS

Among the areas that security-intervention researchers have focused upon is that of *intervention strategies*, that is, when and in which form to intervene. For example, Whalen and Inkpen (2005) showed how symbols as a passive form of interventions are seen, but not interacted with by the users. Wu et al. (2006) argued that the right timing is important for interventions. Generally, active warnings have been shown to be more effective than passive indicators (Schechter et al. 2007). However, overly frequent warnings (e.g. from false positives) lead to habituation effects (Amer and Maris 2006).

Further research occurred on the *content of interventions*. Bravo-Lillo et al. (2011) showed empirically that warnings are not understood – for example, due to technical terminology. Biddle et al. (2009) found that their reformulated warnings made users more responsive to different levels of connection security. Wogalter (2006) argues that warnings need to inform about or remind of the *threats and consequences*. Downs et al. (2006) showed that phishing warnings are more often ignored if the threats and consequences are unknown. Furthermore, Kauer et al. (2012) found that individuals are more likely to heed warnings if they perceive personal consequences.

Wogalter (2006) also argues that warnings need to fit the audience and that *personal characteristics* should be taken into account when designing security indicators. Lin et al. (2011), for example, found that domain highlighting helps a subset of their study participants, depending on the participants' expertise. Bravo-Lillo et al. (2011) apply the Human-in-the-Loop framework (Cranor 2008) to warnings to describe the various factors that influence the

Indicator	Scope	Measurement
Trustworthiness of operator	Web site	Recommendations
Connection encryption	Connection	Browser
User expertise	User	Questionnaire

Table 1: *Examples of context indicators in Web browsing*

behavior of the user – for example, they show that behavior depends on expertise and prior experience.

To warn in an adequate form and achieve the necessary impact, De Keukelaere et al. (2009) proposed to adapt the intervention to the *context*; they found improvements from considering the security risk and prior actions of the user.

3. SECURITY INTERVENTION FRAMEWORK

The goal of the framework is to support the development of suitable security interventions. The literature in the previous section points to the primary concepts of the framework, depicted with their most important interrelations in Figure 1. The outcome for the developer is whether, when, and in which form the intervention appears (*intervention strategy*, e.g. active as a warning or passive as a symbol), and what *content* it conveys (e.g. technical threats or personal consequences).

Both the appearance and the content of the intervention is in our framework primarily influenced by the *context* and *personal characteristics* of the user. Context indicators measure the security and further aspects of the context, and vary concerning scope and type of measurement (see Table 1 for examples from Web browsing).

The information about the context needs to be interpreted and modulated according to two knowledge bases. The first concerns *human behavior*, particularly the *trust decisions* – e.g. under which circumstances users will trust a Web site enough to enter a password (green location bar, professional design, user expertise). Combining the information from the context (e.g. whether the location bar is green) and the knowledge on trust decisions will allow us to estimate the behavior of the current user and to what degree the user needs to be influenced through an intervention. Herein, we take the existing trust signals (e.g. green location bar) as the base line.

As the second knowledge base, the structure of relevant *threats and consequences* enables a more effective formulation of the content of the intervention. By interrelating context indicators, threats, and consequences, the specific consequences relevant to the situation can be identified. In combination with the knowledge on trust decisions, those threats and consequences can be selected that are considered most effective for the user (e.g. those with highest personal value or those unaware of), depending on her experience and expertise.

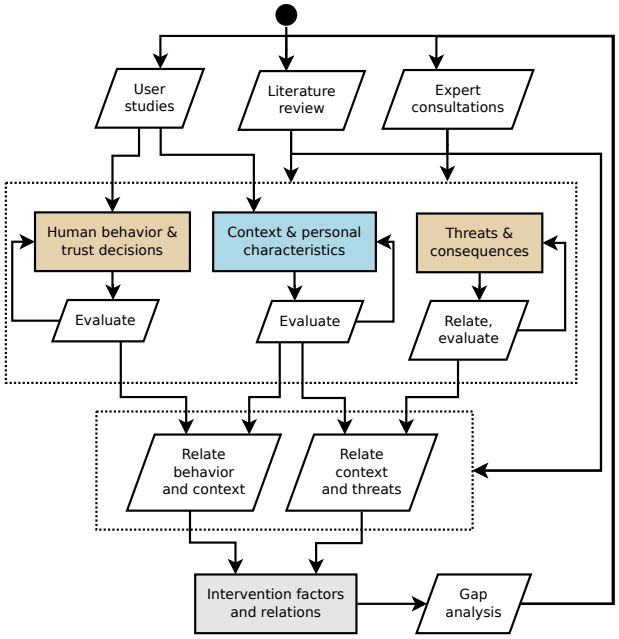

Figure 2: *Intervention-factor elicitation methodology*

4. ELICITATION OF INTERVENTION FACTORS

To operationalize the framework, developers need to elicit the intervention factors and build the knowledge bases for their application area (e.g. Web browsing). We propose the methodology depicted in Figure 2, which employs expert consultations, literature reviews, and user studies to elicit the factors (context indicators, factors in human behavior, and threats and consequences). Context indicators are then evaluated for their availability (e.g. how can we elicit the user's Web browsing intention). Human-behavior factors and threats and consequences are evaluated for their influence on the intervention. These sources are also used to interrelate context indicators and factors to derive graphs for the decisions on the intervention strategy and the dynamic construction of intervention content: How threats may lead to specific consequences, and how context indicators signal specific behavior and threats. Lastly, a gap analysis is performed based on the graphs (e.g. for missing context indicators to identify specific threats) and may trigger an additional iteration.

Not all of these activities will be necessary in their entirety for each newly developed intervention. We foresee general and domain-specific knowledge bases, which, for example, are provided by researchers and which developers then tailor to the specific application.

5. DISCUSSION AND FUTURE WORK

We collected first experiences on applying the methodology in two application areas: Contextualized warnings for Web browsing, which, amongst other aspects, consider the intentions of the user; and for email communication, e.g. addressing malicious email content (phishing) and attachments. In both cases, we conducted literature reviews for human behavior (to identify how indicators influence users), context indicators (what indicators exist and how they can be measured), and threats (which threats and consequences exist in the application). In addition, we applied expert consultations for threat analysis (how are the various threats and consequences interrelated and how do they relate to context indicators) and user studies (how can we elicit the intentions of users in Web browsing as a context indicator). By applying the methodology, we could already derive promising dynamic warnings for concrete situations.

The primary goal for future work is to evaluate the practical applicability of the methodology on two levels: regarding the resulting intervention (efficiency and effectiveness of the intervention for the users) and regarding the development process for the developer. Since the evaluation of generative theories is generally challenging (experimental settings are difficult), we will analytically evaluate the developer effort as the first step and conduct user studies on the resulting interventions. The evaluation will also include the level of complexity of the factors that is necessary to arrive at superior interventions. Lastly, we will study in which ways we can generalize knowledge bases and algorithms that build upon the framework to derive interventions.

REFERENCES

T.S. Amer and J.B. Maris. Signal Words and Signal Icons in Application Control and Information Technology Exception Messages – Hazard Matching and Habituation Effects. Technical Report 06-05, Nothern Arizona University, 2006.

R. Biddle, P. C. van Oorschot, A.S. Patrick, J. Sobey, and T. Whalen. Browser interfaces and extended validation SSL certificates: an empirical study. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, CCSW '09, pages 19–30, New York, NY, USA, 2009. ACM.

R. Böhme and J. Grossklags. The security cost of cheap user interaction. NSPW '11, pages 67–82, New York, NY, USA, 2011. ACM.

C. Bravo-Lillo, L.F. Cranor, J.S. Downs, and S. Komanduri. Bridging the Gap in Computer Security Warnings: A Mental Model Approach. *IEEE Security & Privacy*, 9(2):18 – 26, Mar–Apr 2011.

L.F. Cranor. A Framework for Reasoning About the Human in the Loop. In *UPSEC 08*, Pittsburgh, Pennsylvania, 2008.

F. De Keukelaere, S. Yoshihama, S. Trent, Y. Zhang, L. Luo, and M. Zurko. Adaptive Security Dialogs for Improved Security Behavior of Users. In Tom Gross, et al., editors, *Human-Computer Interaction – INTERACT 2009*, volume 5726 of *Lecture Notes in Computer Science*, pages 510–523. Springer Berlin / Heidelberg, 2009.

J.S. Downs, M.B. Holbrook, and L.F. Cranor. Decision strategies and susceptibility to phishing. In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security*, pages 79–90, New York, NY, USA, 2006. ACM.

M. Kauer, T. Pfeiffer, M. Volkamer, H. Theuerling, and R. Bruder. It is not about the design – it is about the content! Making warnings more efficient by communicating risks appropriately. In *GI SICHERHEIT 2012 Sicherheit – Schutz und Zuverlässigkeit*, 2012.

E. Lin, S. Greenberg, E. Trotter, D. Ma, and J. Aycock. Does domain highlighting help people identify phishing sites? CHI '11, pages 2075–2084, New York, NY, USA, 2011. ACM.

S.E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The Emperor's New Security Indicators. In *S&P '07: Proceedings of the 2007 IEEE Symposium on Security and Privacy*, pages 51 – 65, May 2007.

J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L.F. Cranor. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In *USENIX Security 2009*, 2009.

T. Whalen and K.M. Inkpen. Gathering evidence: use of visual security cues in web browsers. GI '05, pages 137–144, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2005. Canadian Human-Computer Communications Society.

M.S. Wogalter. *Handbook of warnings*. Routledge, 2006.

M. Wu, R.C. Miller, and S.L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 601–610, New York, NY, USA, 2006. ACM.