Guidelines for Usable Security Interventions

Benjamin Maximilian Berens, Mattia Mossano, Maxime Veit, Anne Hennig, and Melanie Volkamer

SECUSO, Karlsruhe Institute of Technology , 76133 Karlsruhe, DE
https://www.secuso.aifb.kit.edu/
{firstname,lastname}@kit.edu

Abstract. Security interventions are vital to protect users against evolving cyber-threats. However, their effectiveness is often limited by usability shortcomings. This paper examines six fundamental obstacles that undermine the impact of security interventions: lack of visibility, habituation & warning fatigue, lack of awareness, long explanation, lack of transparency, and UI attacks. Drawing on results from human-computer interaction, behavioral psychology, and usable security studies, we synthesize evidence-based principles to guide developers in designing usable security interventions. We outline as future work both 1) to empirically validate the proposed guidelines and 2) to extend them for special purposes.

Keywords: M. Angela Sasse · Usable Security Intervention · UI developers.

1 Introduction

Thanks to outstanding usable security researchers such as M. Angela Sasse, the world started acknowledging that blaming and just educating users will not end-up in a more secure digital world [1]. Research has shown that blaming the developers and administration does not help either [32]. Furthermore, M. Angela Sasse and her co-authors showed that existing advice is scattered and finding recommendable, consistent advice to design usable security interventions is a challenge for developers [26].

We aim to close this gap by consolidating the scattered guidance into comprehensive guidelines that help the developers to design usable security interventions. Accordingly, this paper contributes to support developers when applying the 'holistic design approach for usable and effective security' outlined by Sasse et al. [35].

There are different types of security interventions, such as: phishing and malware warnings, permission and consent prompts, and password meters. Our

focus is 1) on security interventions that communicate the current risk and/or the risk to continue, and, *in particular*, 2) on security interventions requiring users to make a decision due to the technology in place not having enough information to reliably do so itself.

We collected several usability issues within existing security interventions from the academic literature. Our focus is on those issues that are context independent. Future work will focus on context and content dependent issues, e.g., those related to URLs being displayed in security interventions to indicate which webpage to open or from which webpage a file is downloaded. We categorized the issues into six categories: lack of visibility, habituation & warning fatigue, lack of awareness, long explanations, lack of transparency, and UI attacks.

Each issue is discussed in its own section. Each section starts with a description of the actual problem. It then presents the solutions discussed in the literature to solve the problem, and some background information. These sections, and in particular the solution paragraphs of each section, serve as guidelines for more usable security interventions.

2 Lack of Visibility

Problem. Security is often not the main focus of users [16, 40], but rather a secondary task that occurs in parallel to the users' primary goal [33]. Research shows that passive security indicators not placed in the users' attention focus, such as the 'lock' icon in a web address bar, are not effective in protecting users [2, 13, 31, 43]. The lack of visibility causes hard-to-notice security interventions to be ineffective.

Solution. Enhance the visibility of security interventions by placing them right next to the respective hazard, i.e., *just-in-place* (e.g., close to a clickable link), and by making them appear only at the time in which they are required, i.e., *just-in-time* (e.g., right before a link is clicked) [29].

Background Information. Positioning and saliency have been important aspects of warning research even outside of the information technology field. For example, ergonomics research shows that warnings are most effective when placed close to the hazard they warn about [41]. Furthermore, education research shows that providing learners with just-in-time information helps them to better understand it, and to apply it in similar future scenarios [24]. In the context of information security, research shows that just-in-time interventions help to be more secure in the moment and to avoid similar issues in the future [5, 22].

3 Habituation & Warning Fatigue

Problem. User habituation and warning fatigue often result from the over-presentation of identical interventions, e.g., warning appearing multiple times over the same issue, leading to reduced intervention effectiveness [2, 34].

Solution. Developers should employ an adaptive approach where interventions are risk-sensitive, engaging users only when essential decisions or actions are required due to the system inability to determine the risk level on its own [3]. The number of risk levels should be only as high and as granular as needed, but it should never be lower than three risk levels (to cover the three minimum scenarios):

- High Risk Level - High risk levels should be used when a system can automatically determine the risk with certainty. No user action should be required and functionalities should be automatically blocked. Security interventions should be used to inform the user about the blocking (e.g., inform them that a website is not secure and will not be available to them).
- Unknown Risk Level - Unknown risk levels should be used when a system cannot automatically determine the risk and the user is required to make a decision. Security interventions should be used to inform the user about the systems' uncertainty. Furthermore, developers should consider employing a time delay mechanism, i.e., require users to wait for a certain amount of time before being able to proceed.
- Low Risk Level - Low risk levels should be used when a system can automatically and with certainty determine that the risk is low. In this case, users are not prohibited to proceed and are not asked to make a decision.

Background Information. Compartmentalizing interventions based on risk levels reduces warning fatigue by distinguishing between varying degrees of urgency [15]. Namely, distinguishing scenarios that require user actions from those that are automatically dealt with by the system, avoiding users becoming desensitized by frequent notifications that have no or only a low impact on their actions [10]. The length of the delay depends on the context of use. We recommend a time delay between one and three seconds, as research has shown this is enough to introduce a significant increase in users' phishing detection [6, 30].

4 Lack of Awareness

Problem. Previous research consistently demonstrates that users frequently lack the required awareness to fully benefit from security interventions [6, 39]. Although such interventions can foster more secure behavior, their effectiveness is substantially diminished when users fail to understand either the risks they mitigate, or the rationale behind the protective actions. This missing awareness is not merely a lack of knowledge but also reflects limitations in the users' mental models and threat understanding [8]. For instance, users may misinterpret security cues or fail to recognize the relevance of a security intervention if they do not perceive themselves as potential targets [12].

Solution. To maximize the impact of security interventions, developers should enhance contextual awareness, such as dynamic explanations of threats (e.g.,

explain that the link text and the URL behind the link mismatch), or just-in-time education (providing information about URL verification checks by users when they need it) that reinforce learning over time. Furthermore, awareness can be raised through a targeted tutorial and/or additional awareness measures.

Background Information. Safety research results show that targeted awareness integrated into the workflow of users, which explain the reasons behind their existence, support both the behavioral change and the safety compliance of workers [18]. In the security field, previous research showed that achieving secure behavior not only requires a targeted theoretical foundation, but also a practical aspect, and an ongoing effort to remind and recall the awareness built so far [36]. As such, research showed that the most effective way to support users is the employment of both: short, targeted awareness measures and usable security interventions [6].

5 Long and Complex Explanation

Problem. Research shows that wordy and technically complex security interventions are unlikely to be understood by the average user [7,9]. Furthermore, those interventions induce fatigue (covered in Section 3). This leads users to ignore security interventions because deemed too time-consuming and worth spending the time as they anyway do not understand the content [2].

Solution. Security interventions should contain the minimum amount of information necessary to address a situation. The amount of information presented should depend on the user's current awareness maturity. For users with low awareness, the explanation should be more supportive and provide enough context to understand the risk and the recommended action. For users with high awareness (e.g., due to recent training), the message can be kept concise and action-oriented. Consequently, security interventions should be salient and actionable, with the degree of emphasis on these attributes scaling with the risk level—particularly in cases of unknown risk. Any further information should be presented only on request, and should be adjusted to the users' awareness maturity [20,36].

Background Information. The length and complexity of an explanation should always be tailored to the demographic it is aimed at [11]. This is especially important when treating technically complex topics, such as training manuals [25].

6 Lack of Transparency

Problem. Security interventions (e.g., phishing warnings in email clients or website blocking in browsers) often hide the underlying classification logic that triggered them. This lack of transparency can lead to confusion, mistrust, or disregard of the intervention [17], especially when users deem the restriction unnecessary [2].

Solution. Security interventions should explicitly state why they appear and what triggered them. For instance, replacing a non-descriptive "blocked" indicator with a brief rationale (e.g., "Blocked due to unverified sender identity" or "High-risk file type detected under organizational policy") would increase users' understanding.

Background Information. Transparency supports informed decision-making and fosters trust, which is crucial for user compliance with security measures [4, 19]. When the classification criteria are hidden, users perceive security controls as arbitrary or obstructive, reducing their motivation to follow them [14], and encouraging risky workarounds [42]. Transparent systems, instead, allow users to form more accurate mental models of system behavior and threat detection [23].

7 UI attacks

Problem. Users may not be able to distinguish between elements belonging to the user interface (UI) of an application and content which can be controlled by an attacker [27]. As a result, attackers can create fake elements that imitate UI parts like password prompts that appear to come from the email client, but are actually part of the email body. Fake interface elements such as scrollbars or buttons can also be used for *clickjacking* [28]. In some interfaces, passive indicators (e.g., "This email is signed" or "Verified sender") are displayed in headers above the content area. If the design does not clearly separate the UI elements from the message content, such indicators could also be spoofed, misleading the user about the authenticity of the message [27].

Solution. The interface should clearly separate application-controlled elements from content areas that can be manipulated by an attacker (e.g., using separator lines or other visual dividers). Dialogues or windows, such as password prompts, should not appear directly in front of the message content, but remain visually distinct and recognizable as part of the UI. To reinforce this separation, applications should, where possible, render only external content that cannot be readily mistaken for client-generated interface cues or security information—for example, by suppressing content features that can mimic tooltips, dialogs, or controls [38]. Additionally, UI attacks can be mitigated by using secret images or other UI elements with an unpredictable appearance [10, 21, 37, 44]. For additional security, users should made aware of where to find the genuine security indicator and made aware of the boundary between trusted UI elements and message content, including that attackers can place misleading cues such as "signed" in the email body (e.g., through dedicated tutorials, see Section 4).

Background Information. By clearly separating the interface from non-interface content, users learn which parts they can trust and which should be treated with caution, as they may originate from an untrusted source [27]. This visual distinction helps users build correct expectations of where trustworthy information

and actions appear. In addition, restricting the rendering of external content features that can imitate client UI cues (e.g., tooltip- or dialogue-like elements) reduces the surface for UI redressing, and prevents attacker-controlled cues from competing with genuine client indicators [38]. When deception cannot be fully prevented at the rendering layer, adding UI elements with an unpredictable appearance (e.g., secret images) strengthens the trusted path by making it harder for attackers to visually replicate security-relevant UI elements [10, 21, 44].

8 Conclusion and Future Work

Our paper aims to help developers create security interventions that are both robust against attacks and genuinely usable. The guide details actionable guidelines for each of six issues identified from the literature. We expect developers to be able to adopt these guidelines when designing usable security interventions. For researchers this paper provides additional background information to understand the rationale for our proposed solutions.

Future work will extend the guidelines to specific contexts such as URL-based and email-based interventions, and evaluate them with developers in real-world settings.

Acknowledgments. This work was supported by funding from the project "Engineering Secure Systems" of the Helmholtz Association (HGF) [topic 46.23.01 Methods for Engineering Secure Systems] and by KASTEL Security Research Lab.

References

1. Adams, A., Sasse, M.A.: Users are not the enemy. Communications of the ACM (1999)
2. Akhawe, D., Felt, A.P.: Alice in warningland: A large-scale field study of browser security warning effectiveness. USENIX Security (2013)
3. Alt, F., Hassib, M., Distler, V.: Human-centered behavioral and physiological security. New Security Paradigms Workshop (2023)
4. Araujo, R.M., De Amorim, R., Zanatta, D., Mattos, F., Guizani, M.: Why do users ignore security alerts? a qualitative study. NDSS (2020)
5. Bender, S., Horn, S., Loewenstein, G., Roberts, O.: Phishing feedback: just-in-time intervention improves online security. Behavioural Public Policy (2024)
6. Berens, B.M., Schaub, F., Mossano, M., Volkamer, M.: Better together: The interplay between a phishing awareness video and a link-centric phishing support tool. CHI (2024)
7. Bravo-Lillo, C., Cranor, L.F., Downs, J., Komanduri, S., Sleeper, M.: Improving computer security dialogs. INTERACT (2011)
8. Camp, L.J.: Mental models of privacy and security. IEEE Technology and Society Magazine (2009)
9. Demjaha, A., Spring, J.M., Becker, I., Parkin, S., Sasse, M.A.: Metaphors considered harmful? an exploratory study of the effectiveness of functional metaphors for end-to-end encryption. USEC (2018)

10. Dhamija, R., Tygar, J.D.: The battle against phishing: Dynamic security skins. SOUPS (2005)
11. Dialog für Cyber-Sicherheit: Leitfaden des Workstreams „Effektive IT - Security - Awareness: Wirksam ein Bewusstsein für Risiken schaffen ". Tech. rep., Bundesamt für Sicherheit in der Informationstechnik (2022)
12. Downs, J.S., Holbrook, M., Cranor, L.F.: Behavioral response to phishing risk. eCrime Researchers Summit (2007)
13. Egelman, S., Cranor, L.F., Hong, J.: You've been warned: An empirical study of the effectiveness of web browser phishing warnings. CHI (2008)
14. Eslami, M., Vaccaro, K., Lee, M.K., Elazari Bar On, A., Gilbert, E., Karahalios, K.: User attitudes towards algorithmic opacity and transparency in online reviewing platforms. CHI (2019)
15. Felt, A.P., Reeder, R.W., Ainslie, A., Harris, H., Walker, M., Thompson, C., Acer, M.E., Morant, E., Consolvo, S.: Rethinking connection security indicators. SOUPS (2016)
16. Gutfleisch, M., Klemmer, J.H., Busch, N., Acar, Y., Sasse, M.A., Fahl, S.: How does usable security (not) end up in software products? results from a qualitative interview study. IEEE S&P (2022)
17. Gutfleisch, M., Peiffer, M., Erk, S., Sasse, M.A.: Microsoft office macro warnings: a design comedy of errors with tragic security consequences. EuroUSEC (2021)
18. Hancock, P., Kaplan, A., MacArthur, K., Szalma, J.: How effective are warnings? a meta-analysis. Safety Science (2020)
19. Herley, C.: So long, and no thanks for the externalities: the rational rejection of security advice by users. New Security Paradigms Workshop (2009)
20. Hielscher, J., Kluge, A., Menges, U., Sasse, M.A.: "Taking out the Trash": Why Security Behavior Change requires Intentional Forgetting. New Security Paradigms Workshop (2021)
21. Jackson, C., Simon, D.R., Tan, D.S., Barth, A.: An evaluation of extended validation and picture-in-picture phishing attacks. International Conference on Financial Cryptography and Data Security (2007)
22. Jenkins, J.L., Grimes, M., Proudfoot, J.G., Lowry, P.B.: Improving password cybersecurity through inexpensive and minimally invasive means: Detecting and deterring password reuse through keystroke-dynamics monitoring and just-in-time fear appeals. Information Technology for Development (2014)
23. Kauer, M., Volkamer, M., Braun, J., Buchmann, J.: A differentiation of feedback to support user security decisions. SOUPS (2012)
24. Kester, L., Kirschner, P.A., van Merriënboer, J.J., Baumer, A.: Just-in-time information presentation and the acquisition of complex cognitive skills. Computers in Human Behavior (2001)
25. Kincaid, J.P., Fishburne, R.P., Rogers, R.i.L., Chissom, B.S.: Derivation of New Readability Formulas for Navy Enlisted Personnel. Tech. rep., Naval Air Station Memphis (1975)
26. Klemmer, J.H., Gutfleisch, M., Stransky, C., Acar, Y., Sasse, M.A., Fahl, S.: "make them change it every week!": A qualitative exploration of online developer advice on usable and secure authentication. CCS (2023)
27. Müller, J., Brinkmann, M., Poddebniak, D., Böck, H., Schinzel, S., Somorovsky, J., Schwenk, J.: "johnny, you are fired!"–spoofing openpgp and s/mime signatures in emails. USENIX Security (2019)
28. Niemietz, M., Schwenk, J.: Out of the dark: Ui redressing and trustworthy events. International Conference on Cryptology and Network Security (2017)

29. Parkin, S., Redmiles, E.M., Coventry, L., Sasse, M.A.: Security when it is welcome: Exploring device purchase as an opportune moment for security behavior change. USEC (2019)
30. Petelka, J., Berens, B., Sugatan, C., Volkamer, M., Schaub, F.: Restricting the link: Effects of focused attention and time delay on phishing warning effectiveness. IEEE Symposium on Security and Privacy (2024)
31. Petelka, J., Zou, Y., Schaub, F.: Put your warning where your link is: Improving and evaluating email phishing warnings. CHI (2019)
32. Reinfelder, L., Landwirth, R., Benenson, Z.: Security managers are not the enemy either. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (2019)
33. Sasse, M.A., Flechais, I.: Usable security: Why do we need it? how do we get it? Security and Usability (2005)
34. Sasse, M.A., Hielscher, J., Friedauer, J., Buckmann, A.: Rebooting it security awareness–how organisations can encourage and sustain secure behaviours. European Symposium on Research in Computer Security (2022)
35. Sasse, M.A., Brostoff, S., Weirich, D.: Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. BT Technology Journal (2001)
36. Schöni, L., Carles, V., Strohmeier, M., Mayer, P., Zimmermann, V.: You know what? - evaluation of a personalised phishing training based on users' phishing knowledge and detection skills. EuroUSEC (2024)
37. Veit, M.F., Volkamer, M.: Passsec+ - an addon that protects your passwords, payment data and privacy. SOUPS Poster (2022)
38. Veit, M.F., Wiese, O., Ballreich, F.L., Volkamer, M., Engels, D., Mayer, P.: Sok: The past decade of user deception in emails and today's email clients' susceptibility to phishing techniques. Computers & Security (2025)
39. Wash, R.: Folk models of home computer security. SOUPS (2010)
40. Whitten, A., Tygar, J.D.: Why johnny can't encrypt: A usability evaluation of pgp 5.0. USENIX Security (1999)
41. Wogalter, M.S., Conzola, V.C., Smith-Jackson, T.L.: Research-based guidelines for warning design and evaluation. Applied Ergonomics (2002)
42. Woltjer, R.: Workarounds and trade-offs in information security–an exploratory study. Information & Computer Security (2017)
43. Xiong, A., Proctor, R.W., Yang, W., Li, N.: Is domain highlighting actually helpful in identifying phishing web pages? Human Factors (2017)
44. Ye, Z., Smith, S., Anthony, D.: Trusted paths for browsers. TISSEC (2005)