POSTER: Towards Ecological Validity for Password Alternative User Studies

Yasemin Acar[1], Michael Backes[1,2],
Sascha Fahl[1] Maximilian Koch[3],
Christian Stransky[1]
CISPA, Saarland University[1]
MPI-SWS[2]
Leibniz University Hannover[3]
{acar,backes,fahl,stransky}@cs.uni-saarland.de
koch@dcsec.uni-hannover.de

ABSTRACT

Passwords still constitute the most widely deployed authentication mechanism. Striking a good balance between passwords that are secure and those ones that users manage to conveniently use is still the key factor for their secure deployment: Security crucially relies on choosing sophisticated, frequently changed passwords, and on keeping distinct passwords for different services; yet such passwords are often difficult to memorize, which results in written down passwords as well as password revocation and reuse, all of which open additional attack vectors on the security of password-based authentication systems. Alternative constructions have hence been put forward – both within academy and industry – that promise to offer more memorable password replacement or enhancement systems. However, both their adoption and evaluation lack comparability and systematization. In this work, we present a study framework to evaluate the usability and security of password alternatives and improvements in a longitudinal real world scenario. This study framework allows researchers to collect ecological valid and comparable study results for password alternatives and improvements.

1. INTRODUCTION

Passwords as the most prevalent authentication mechanism lead to many security and usability problems: Users tend to choose memorable and unsafe passwords, which is sometimes countered by policies that enforce specific password compositions and/or frequent password changes. Both make the memorability issue worse, leading to even more security and usability problems with passwords: Users either write down or reuse their passwords or game the annoying policies, e.g. by adding '1!' in the end of the password. Alternately, they forget their passwords, which leads to lengthy and/or unsafe password recovery, e.g. simple security questions. This lingering problem has caused the security and HCI communities to try to educate/nudge users to make better password choices, e.g. by password meters [11]. Moreover, the community has come up with a plethora of alternate authentication mechanisms [8, 4, 3, 10, 12, 9, 2]. These lead from biometric authentication over two-factor authentication to recalls of pictures and/or faces as well as gestures. However, not all of these alternatives offer the same possibilities as textual passwords: key spaces, entry times, the possibility of shoulder-surfing and the devices required to enter them vary. Many first contact- or recall-studies have been conducted to evaluate these alternate methods [8, 3, 10, 12, 9]. However, due to their rather exploratory designs, these studies are hardly comparable and fail to map the real world use of passwords. For an alternative password system to reach widespread adoption, it is important to evaluate its real world usability and security, its strengths over passwords as well as its weaknesses [1]. Before releasing a mechanism into the real world, the following questions should be considered: *For which use case can this alternate authentication mechanism replace passwords? Is it a complete replacement, or should it be used as a backup/revocation mechanism? How does it perform memorability wise, effort wise and concerning reuse? How do security characteristics perform compared to text passwords? How long does it take to learn and set up, how long to use it to login? For systems with keys: How well do users actually use the key space? How does it perform across these questions compared to password and other password alternatives?*. Therefore, we propose a study design for password alternatives/improvements that overcomes methodological weaknesses of earlier studies and develop a study framework to make password alternative studies comparable in an ecologically valid setting without unnecessary overhead:

Browser Extension: Our study framework is implemented as a Chrome extension and hence is easily usable by all Internet users that use Chrome to surf the web.

Easily Extendable: Our study framework provides server side functionality. New authentication mechanisms to be evaluated can be implemented as clients of the framework. As of today, we have implemented three alternative authentication mechanisms/improvements for passwords: PhoneAuth [4], PassFaces [5] and a password meter [11].

Longitudinal Studies: As an improvement over the prevalent one-contact/two contacts online or laboratory studies [8, 3, 10, 12, 9], our framework allows to conduct longitudinal studies in an ecologically valid setting.

Realistic Scenarios: Instead of letting participants create artificial accounts/passwords specifically for a study, our framework allows to investigate the participants' behaviour under realistic conditions for their real login behaviour across their real accounts in a privacy protecting and respecting way.

Comparability: Our study framework allows to investigate how users behave in the study compared to their real behaviour by comparing security properties of their real passwords with security properties of the evaluated mechanism: We can analyze reuse behaviour of alternate password systems compared to their real passwords. In addition to password reuse, our framework analyzes password properties such as the alphabet, zero-order, Shannon and the NIST entropy and probabilistic password strengths on the client side. The metrics will be transmitted to our server, while the actual passwords remain on the clients.

Instant Feedback: Instead of conducting a retrospective interview/survey in the end, we apply experience sampling and collect participants' feedback during their use of the system.

Telemetry: In addition to security properties (e. g. entropy), our study framework also collects telemetry data such as the number of (successful and failed) login attempts, timestamps of logins and the websites participants login to (in case participants gave their consent) as well as reuse behaviour of their usernames.

2. STATUS QUO

Many password alternatives or improvements have not only been suggested, but were evaluated to varying degrees with the help of exploratory user studies [8, 4, 3, 10, 12, 9, 2]. In most cases, first contact studies were conducted, which suffer from many fallbacks: Long- and midterm memorability cannot be tested in this setting. It is not possible to test reuse-strategies over time, nor can real world usability and security be evaluated. These kinds of exploratory studies only offer a first insight into the uses of a new password alternative system. In some cases, these first contact studies are paired with a longer-term recall-test. However, up to now, password alternatives have not been tested for real existing accounts participants use in their real lives. In addition, many of these studies were conducted on Amazon MTurk, or in laboratory settings, which contribute to a study bias. The use of home studies has been difficult: Asking users to keep a diary has also introduced a bias to reflect more about their authentication behaviour [7]. Additionally, self-reporting biases were introduced. Based on these suggestions and exploratory evaluation approaches, the community knows that some of the suggestions *might* be *better* than passwords, but it is not clear in which ways exactly and for which use cases.

3. A BETTER FUTURE

The methodological issues discussed above illustrate that an additional study design is needed: To produce ecologically valid study results for password alternatives that extend a preliminary exploratory study, the following methodological aspects need to be considered.

3.1 Study Design

3.1.1 Longitudinal Study

Most password alternative studies are conducted as first contact evaluations, i. e. during the study, users learn/get to know the system that is to be evaluated for the first time. They then choose one or multiple passwords, which are tested for memorability either at the same time, after a few minutes, or in a two- or multisession-study. However, the latter cases are rare: Most user studies do not require their participants to use alternatives for/improvements over text passwords for a longer time [8, 3, 10, 12, 9].

With our study framework , we propose that in addition to a first-contact study, user studies should be conducted in the context of participants' real user accounts and passwords over a longer period.

3.1.2 Real Scenarios

As illustrated in our previous work [6], study results cannot be perfectly mapped onto real world user behaviour: 30% of our participants behaved differently in our password study as compared to their real password behaviour. While self-reporting did help to assuage this problem (answering the question *"Did your behaviour differ from your real world behaviour?"* slightly improved this number), it is still hard to estimate which participants represent those 30%. Therefore, authentication studies should aim to investigate real scenarios.

3.1.3 Telemetry Data

Many studies [8, 3, 10, 12, 9] include a self-reporting survey. Our study framework enables researchers to perform telemetry, circumventing a self-reporting bias.

3.1.4 Experience Sampling

It is common [8, 3, 10, 12, 9] to rely on a retrospective survey/interview at the end of the study to ask participants how they (dis-)liked certain aspects of the evaluated mechanisms. This procedure helps to achieve a rough understanding of the participants' experiences with and opinions of the evaluated mechanisms. However, self-reporting after the fact, possibly with some tasks performed between the relevant action and the corresponding questions, always introduces a memory bias. To obtain a better understanding of participants' experiences and opinions of the evaluated mechanisms, we implemented the possibility for experience sampling into our study framework to enable researchers to measure interesting aspects during the study, and gather feedback on procedures right after they were used.

3.2 Implementation

Our study framework is implemented as a Chrome extension and works as a client/server architecture. Figure 1 gives an overview of the architecture of our study framework.

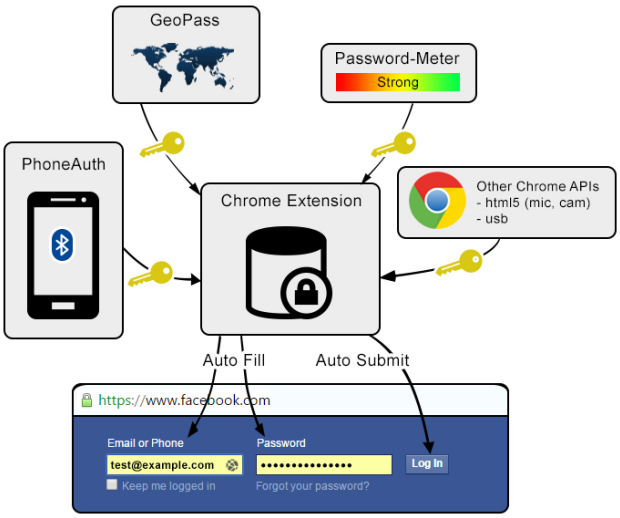

Figure 1: Overview of the study framework architecture.

Our study framework offers the functionality of a password manager: With our tool, we are able to test passwords that are meant to serve as one-to-many mappings for their usability. Presently, authentication using an Android device as a hardware-like token [4] is implemented. The framework also offers the possibility to map every password entry to an alternate password system, effectively modelling that users use the alternate password instead of their textual password login in a many-to-many mapping. We implemented this for PassFaces [5].

4. FUTURE WORK

In the near future, we plan to investigate usability and security aspects of different password alternatives and improvements and to compare our results with original research results to obtain a better understanding of the ecological validity of previous user studies in this area. As of yet, we have implemented three mechanisms:

PhoneAuth: PhoneAuth [4] allows users to use their smartphone to login to websites. It uses bluetooth to connect to our browser extension and uses an Android device as a hardware-like token to authenticate users. This mechanism removes the burden of having to remember many different passwords for many different accounts.

PassFaces: Instead of using a text-based password, PassFaces [5] asks users to select a number of facial pictures to be used as a secret for authentication. Out of a larger number of images users pick for example nine images that server as their passface. This mechanism replaces one text-based password with one passface. Hence, it does not reduce the number of secrets a user has to remember.

Password Meter: Password meters [11] are thought to strengthen text-based passwords by giving users hints on how to add extra security to a password. This mechanisms tries to increase security for a password and does not reduce the number of secrets a user has to remember.

We selected these three mechanisms above since we think they are representative candidates for mechanisms the security research community came up with over the last years. By investigating these mechanisms with our study framework, we hope to be able to answer critical questions affecting those mechanisms. Additionally, we hope to encourage the usable security research community to adopt our study design recommendations and implement alternative mechanisms or improvements as extensions of our study framework.

5. REFERENCES

[1] J. Bonneau, C. Herley, P. C. v. Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. SP'12.

[2] J. Bonneau and S. Schechter. Towards reliable storage of 56-bit secrets in human memory. USENIX Security'14.

[3] S. Chiasson, R. Biddle, and P. C. van Oorschot. A second look at the usability of click-based graphical passwords. SOUPS'07.

[4] A. Czeskis, M. Dietz, T. Kohno, D. Wallach, and D. Balfanz. Strengthening user authentication through opportunistic cryptographic identity assertions. CCS'12.

[5] P. Dunphy, J. Nicholson, and P. Olivier. Securing passfaces for description. SOUPS'08.

[6] S. Fahl, M. Harbach, Y. Acar, and M. Smith. On the ecological validity of a password study. SOUPS'13.

[7] E. Hayashi and J. Hong. A diary study of password usage in daily life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11.

[8] F. Schaub, M. Walch, B. Könings, and M. Weber. Exploring the design space of graphical passwords on smartphones. SOUPS'13.

[9] R. Shay, P. G. Kelley, S. Komanduri, M. L. Mazurek, B. Ur, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor. Correct horse battery staple: Exploring the usability of system-assigned passphrases. SOUPS'12.

[10] S. Uellenbeck, M. Dürmuth, C. Wolf, and T. Holz. Quantifying the security of graphical passwords: The case of android unlock patterns. CCS'13.

[11] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor. How does your password measure up? the effect of strength meters on password creation. USENIX Security'12.

[12] R. Weiss and A. De Luca. Passshapes: Utilizing stroke based authentication to increase password memorability. NordiCHI'08.