On Usable Security and Verified Password Managers

Carolina Carreira
Instituto Superior Técnico
Universidade de Lisboa

Abstract

Password Managers (PMs) are useful tools to manage passwords but they are not widely used. Studies indicate usability problems and distrust from users as the reasons for the low adoption of PMs. As such, we propose extending an existing PM by implementing relevant usability best practices and increasing transparency by educating users about how PMs work. This project is part of the PassCert research project, which aims to build a formally verified PM. Therefore, another goal is to explore ways that effectively convey to users the formally verified properties. We performed user studies that suggest that our solution improves the usability of the PM and that we were able to convey relevant information about its formally verified features. We contribute with the first study on users perceptions of formal verification on PMs and hope that our findings can help the formal verification security community better communicate with end-users.

1 Introduction

As Whitten and Tygar pointed out in their seminal work, security mechanisms are only effective when used correctly [28]. For example, effective use of text passwords, one of the most used security mechanisms [14], requires not reusing them across different services and not choosing simple, easy-to-guess passwords. However, this presents a challenge for users. In a study by Stobert et al. [26], only one of the 26 participants reported not reusing passwords between accounts and 73% reported reusing passwords either "always" or "frequently". Not only is password reuse a problem but users also struggle with choosing good quality passwords. Gaw et al.'s study [11] about password usage found that 51.79% (of 56 users) believed that a friend had a higher chance of guessing their password, suggesting that they used non-random passwords with personal information.

It is in this context that Password Managers (PMs) become an essential solution. Security experts and several governmental institutions, such as the European Union Agency for Cybersecurity [7], strongly recommend the usage of PMs that combine secure password storage and retrieval with random password generation. These tools can improve account security by enabling the use of strong and unique passwords, simultaneously improving the usability and convenience of text password authentication.

However, despite PMs being recommended, they are not widely used [1]. Several studies tried to find the reasons for this phenomenon and have reached different conclusions: some state users' unawareness of the existence of PMs [1, 20, 26], lack of trust [15, 21] and lack of motivation [1, 21]. One common factor that was mentioned by all studies was usability problems [1, 2, 6, 17, 20, 24].

In this short paper, we review usability challenges of PMs and we propose the use of known usability best practices and techniques to extend and improve Bitwarden [3], a widely popular open-source PM. Since this work is done in the context of the PassCert project[1], which aims to build a formally verified PM, a novelty of our work is the investigation of ways to effectively convey to users the formally verified properties and whether formal verification increases users' trust in PMs. We report on our results and propose a methodology to evaluate our extensions and determine the impact of formal verification on PMs.

The goals of this project are to:

- survey usable security techniques that can be applied to improve password managers;
- ensure that the password managers developed in the context of the PassCert project integrate best practice guidelines developed by the usable security community;
- explore ways that effectively convey the formally verified properties of the password managers;
- learn more about users perception of PMs and formal verification;

After presenting the usability challenges of PMs in Section 2, we present in Section 3 best practices for improving usability. In Section 4, we discuss usability problems in the

[1] PassCert Project Homepage: https://passcert-project.github.io

context of Bitwarden and concrete actions that can be taken to address these in our proposed extension. In Section 5, we present our testing methodology and results. We conclude the paper in Section 6, where we also discuss future work.

1.1 Research Papers

Parts of the work presented in this thesis were used in the following research papers:

- ***Carolina Carreira**, João F. Ferreira, and Alexandra Mendes. Towards Improving the Usability of Password Managers. Presented at INFORUM 2021 (Comunicação). 2021 [4]*
- ***Carolina Carreira**, João F. Ferreira, Alexandra Mendes, and Nicolas Christin. Exploring Usable Security to Improve the Impact of Formal Verification: A Research Agenda. In 1st International Workshop on Applicable Formal Methods (co-located with Formal Methods 2021). Beijing, China. 2021 [5]*

2 Usability Challenges of Password Managers

The usability of PMs is an important aspect that can increase their adoption and that has been studied by the research community. In this section, we present and discuss usability challenges documented in the literature.

2.1 Password Manager Usage

Stobert et al. [26], in a study about password usage, were surprised to find that none of their participants used a dedicated PM and that most of them were unaware of popular PMs. Furthermore, a few participants expressed distrust in PMs. The authors suggested that a good integration of PMs into operating systems and browsers would help with visibility and trust.

More recently, Pearman et al. [20] studied the usage of PMs and other password management methods. A 30-participant interview study was conducted with users who do not use PMs at all (9 people), who use PMs built into their browsers or operating systems (12 people), and who employ separately installed PM application (7 people).[2] The study found that **people who do not use PMs** rely mostly on memory or unsafe methods (e.g., saving on Excel sheets). The reason for not using PMs was mostly **unawareness** of their existence.

In this study, one of the major complaints was related to a **lack of awareness of how the tool and its security worked**. By not understanding the features offered, some users could not, for example, synchronize passwords between devices. This lack of information also made the users wary of PMs' security. These findings were also backed by the work of Ion et al. [15] where non-expert users expressed a lack of trust in PMs. The motivation for users of separately installed PMs was primarily security and even though some reported poor usability (e.g., difficulty navigating the interface), they were satisfied with the security provided.

Convenience, usability, and security were the main concerns raised in this study and a problem identified was the users' lack of information regarding how PMs work. The study calls for better usability testing and focus on non-experts.

2.2 Password Manager Usage with Older Users

The participants in Pearman et al.'s study were skewed towards young people, with a high percentage of participants with technical backgrounds. As such, Ray et al. [21] expanded Pearman et al.'s findings by replicating their protocol and interview instrument but applied to a sample of strictly older adults. A 26-participant interview study was then conducted with older adults (aged above 60) who do not use PMs at all (10 people), who use PMs built into their browsers or operating systems (9 people), and who use separately installed PMs (7 people). Across all, secure access to financial accounts was valued above other types of online accounts. Regarding users that do not use PMs, both older and younger adults were concerned about **a single point of failure** when using PMs (e.g., losing access to all passwords stored). Concerning the participants that used browser built-in PMs, both older and younger adults were worried about others having access to their passwords and about where they were stored. Similar to the findings of Pearman et al., users who adopted separately-installed PMs were motivated by their desire for better security.

Lack of self-efficacy when dealing with software was one of the main barriers to the adoption of PMs. A higher level of transparency (e.g., showing users how secure their passwords are) could also help towards increasing trust [21].

The suggestion given by Ray et al. was to encourage advocacy, particularly from family or friends, but also by trusted organizations. Another suggestion was education to convey urgency of secure practices (e.g., classes at senior centers). Erroneous and **incomplete mental models** of how PMs work (e.g., encryption, cloud storage, etc.) also surfaced in this study [21].

2.3 Password Managers in Smartphones

Usability in smartphones presents different challenges from conventional desktop interfaces. For example, in a study focused on PMs for mobile devices by Seiler-Hwang et al. [24], users' unawareness of the existence of PMs was not a rejection factor, as most of the participants knew about them. Seiler-Hwang et al. conducted a usability study comparing

[2] Two participants "were difficult to place in the aforementioned categories" [20]

4 popular smartphone PMs (Dashlane, Keeper, Lastpass and 1Password) with 60 participants. They used the *System Usability Scale (SUS)* to compare the PMs' usability. Overall, looking at the small sample of analyzed applications, PMs appear as software tools that can be subjectively considered "ok", but far from being "excellent" [24]. Participants often complained about **lack of guidance**, instructions, tutorials, or help pages. This meant that sometimes they were unable to achieve their goals within the PM. Also, for participants that were unfamiliar with PMs, this lack of guidance is translated into a lack of understanding about how PMs work. Finally, one of the most problematic areas identified in the usability of mobile PMs was **poor integration** with other applications and browsers.

Alkaldi et al. [1] investigated the factors impacting the adoption or rejection of smartphone PMs based on Play Store and App Store reviews. They found factors such as **awareness, no perceived usefulness, security, and privacy concerns** to be detrimental to the adoption of PMs. They state that even if people become aware of the apps, they might still not embark on a search process to consider installing one. Failure to reassure potential users about the trustworthiness of PMs was identified as a main factor behind their rejection.

2.4 Comparative Usability Studies

A comparative analysis of PMs usability and security was conducted by Arias-Cabarcos et al. [2] on five different mainstream PM applications. For the usability study they used a set of evaluation criteria known as the 5 Es (Efficient, Effective, Engaging, Easy to learn and Error tolerant). Although the PMs studied did not have negative ratings of usability, important differences arose when users rated PMs according to the engaging and easy-to-learn features. An interface is engaging if it is pleasant and satisfying to use and it is easy to learn if it allows users to learn without effort. KeePass was the worst evaluated manager in both these categories [2]. The best rated PM, in all categories, was Dashlane.

A usability issue related to the users' **mental maps** was about the tools' activation. Users believed that the PMs would, after an initial activation, stay working for the rest of their computer session. Inconsistency in the interface of the PM also hinders the mental model of the users. For example, this was observed in PwdHash, one of the PMs studied by Chiasson et al. [6], where a specific command was irrelevant as it would give the same output whether it was used or not.

Not all usability problems encountered by Chiasson et al. were a direct result of the PMs' interfaces. Some problems were due to bad website design. These are valid usability issues that provide context and insight into the circumstances and environments where people will be using PMs.

Control was also an important issue for users. When the PMs on the study did not show the passwords that they were generating, users felt frustration as they felt as though they had no control over their passwords.

A major problem arises from the developers' assumption that users will use the tool correctly. This is problematic as new users frequently commit mistakes and may be deceived into thinking they are safe when they are not. If the systems are very secure but do not have good usability, users may opt to use a different, less secure system that lets them do what they want [23].

3 Improving Password Managers

As we have seen in the previous section, PMs have usability problems that need to be addressed. Additionally, given PassCert's context, we consider a new set of challenges related with how information about formal verification is conveyed to users. This section presents best practices and possible solutions to address these challenges.

3.1 Usability Improvements

There are general design guidelines that can be followed to improve the usability of PMs. For example, a guideline to follow is Shneiderman et al.'s *Eight Golden Rules of Interface Design* [25]. These are intended to be used during design of systems [8]. Another good practice is *security by default*: to have default configuration settings that are the most secure settings possible [24]. This is particularly helpful for inexperienced users as they may not understand the meaning of every setting in the interface. A **well-integrated software** with bug-free features is also essential to enable users to create clear mental models of the tool [6, 24]. To achieve this, software and usability tests, and formal verification can be used. A good software has to have a clear navigation and be error tolerant (this is especially important for new users). It should be permissive and allow the users to recover and learn from mistakes [10].

Table 1 summarizes usability challenges concerning PMs and proposed solutions. Regarding the proposed help documentation and tutorials, it is important to to avoid the use of technical jargon [10, 22]. Moreover, explanations of the different security options can be achieved through the use of tooltips [10] and help icons.

3.2 Information on Formal Verification

It has been shown that formal verification is valuable when considering password security [9, 16]. Since this work is done in the context of the PassCert's project, which aims to build a formally verified PM, a novelty of our work is the investigation of ways to effectively convey to users the formally verified properties and whether formal verification increases users' trust in PMs. Therefore, a primary concern we have is **educating the users about formal verification**. This can be achieved by implementing the following:

- Provide a clear way to understand what properties the system is formally assuring with status symbols to indicate that a certain action is formally assured [10].
- Concise explanations about formal verification. It is important to use correct and simple language in order to prevent alienating users (e.g. avoid the use of jargon and unnecessary technical language) [10, 22].
- Further information may be required by the more inquisitive users and it should also be provided. This can be done by providing links to expanded documentation and further resources.

4 Extending Bitwarden

The PassCert project is using Bitwarden as a basis for creating a proof-of-concept PM that through the use of formal verification, guarantees properties on data storage and password generation [12]. Therefore, as the work presented here is done in the context of PassCert, our goal is to improve the usability of Bitwarden and explore how it can effectively convey information about the formally verified properties. This section starts by presenting usability problems of Bitwarden. It then describes several extensions already implemented and preliminary results.

4.1 Bitwarden Usability Problems

A thorough analysis of the Bitwarden interface was conducted and, considering the information presented in previous sections, the main problems found were:

- **Lack of user support.** Bitwarden does not provide access to any tutorial, which is something that participants in previous PM studies asked for [24].
- **Lack of consistent tooltip information.** Some settings had no support or tooltips associated, making it more difficult to understand some features (e.g. the input box "Authenticator Key").
- **Lack of consistent behaviour.** We found inconsistencies in buttons that look the same but present distinct behaviors. Inconsistencies in interfaces can hinder users mental models [6, 25].

4.2 New Icon Signalling Formal Verified Features

To help users become aware of the formally verified features of the PM we designed new icons to represent formal verification. We use icons because Wiedenbeck [29] suggested that users have less favorable perceptions of text only UIs. When designing new icons is important to have an unified design [27]. To ensure this, we used the same font that Bitwarden's existing icons use. Moreover, we considered variations of existing or familiar icons (see Figure 1). Some of these may serve as metaphors for security. Interface metaphors are important to convey information [25, 27]. The icon design process went through several iterations, a brainstorming session, two rounds of feedback from the team, and lastly, feedback from 25 users outside the team. The feedback was composed of an attractiveness test where users chose the icons they like more without context, followed by a preference test where we explain what the icon is trying to convey and ask users to rate it by preference [13]. The icon chosen to implement in PassCert was Icon D from Figure 1.

Figure 1: Icon variations. Icon D is the formal verification icon chosen.

4.2.1 Explanations about formal verified features.

The formal verification icon is distributed throughout the interface where a feature is formally verified. When clicked, it opens a contextual description about the formal verification of that specific feature (see Figure 2(b)).

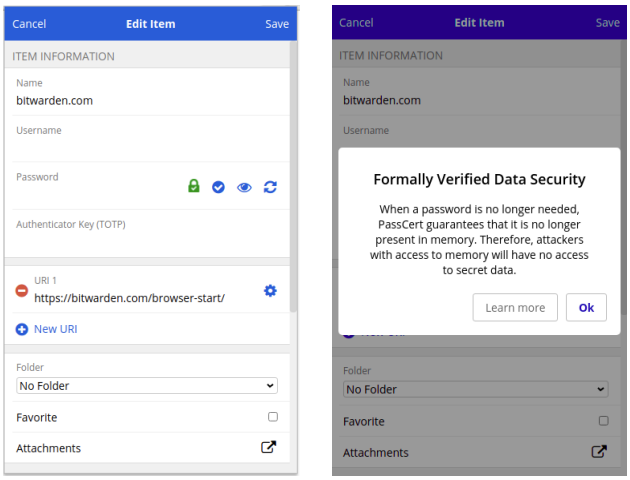
These explanations were designed in two iterations: a first definition of the features was written and improved over two rounds of feedback from the PassCert team that is implementing them. We aimed to keep the language simple and without jargon to facilitate understanding [10, 22]. An example, related to a data security property, is illustrated in Figure 2(b).

4.3 Additional Information via FAQs and Tutorials

As mentioned before, when users are using the PM they may want to learn more about certain aspects. To convey information about relevant topics, we designed a FAQ to be embedded within the PM. Although Bitwarden already provides help pages, these are exclusively online. On the other hand, our FAQ is accessible even when users want to access their passwords (and PM) offline. To implement the FAQ we followed Redish's [22] recommendation of going through every topic of interest and providing questions and answers for them.

Table 1: PMs' challenges and proposed solutions

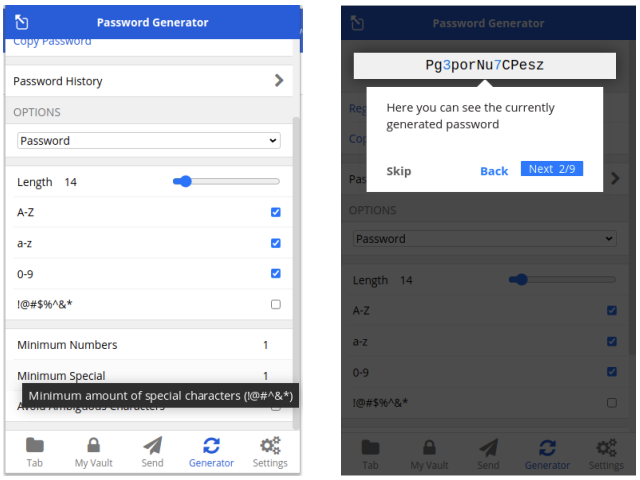
Challenge	Proposed Solution	Description of Solution
Lack of trust and understanding [15, 21, 24]	Provide a higher level of transparency (e.g., showing users how secure their passwords are)	• Educate users about how PMs' work [24]. • Advocacy from trusted organization about the use of PMs (e.g., schools) [21].
Lack of motivation to use PMs [1, 15, 21]	Educate users about the benefits of using a PM	• Provide information related to the dangers of unsafe password habits [6, 15, 20], and about the increased productivity and security of using PMs [15].
Bad performance, poor integration with other applications and browsers [24, 26]	Solid implementation of all PM's features	• Functionalities like password generation, auto-fill, and device synchronization are core and need to be well implemented [24]. • Usability testing of the PMs and their integration with other applications and browsers [24, 26].
Difficulty of use (lack of usability) [1, 2, 6, 17, 20]	Simplify the interface and provide support for users	• Tutorials about how the interface works (for beginner and expert users). These should be naturally integrated with the interface to be promptly accessible when required, but should not interfere negatively with the user experience [24]. • Explain what different options in the security settings mean [24]. • If users are unsuccessful, feedback should be short and help them address the issue [6]. • The PM should be error tolerant: this is especially important for new users. The PM must be permissive and allow the users to recover and learn from their mistakes [2].
Inadequate Mental Models [6, 15, 21]	Provide a precise interface	• Give feedback to users about the status of their actions (if they were successful or not) [6]. • Navigation should be as clear as possible [10, 23].



(a) Formal verification icon in the password vault (green icon)

(b) Pop-up after clicking formal verification icon

Figure 2: Formal verification icon and subsequent pop-up



(a) Tooltip implemented for the "Special Characters" button

(b) Example of the page tour

Figure 3: Interface extensions: tooltips and tutorial walkthrough

Users can access the FAQ pages from the "Settings" tab or by opening the formal verification icon, and clicking "Learn More" (see Figure 2(b)).

We also implemented a tutorial for users in the form of a walkthrough, which guides users through how the application works. This is in the form of a layer on top of the application [27]. The walkthrough implemented goes through the main sections of the PM: current tab, vault, password generator, and settings. Figure 3(b) shows an example step of this walkthrough (password generator).

4.4 Improved Tooltips

As stated in Section 4.1, Bitwarden's native tooltips can improve. We categorized existing tooltips as **Well implemented**, **Non-descriptive**, or **Missing**. Examples of **non-descriptive** tooltips were found in the bottom toolbar used to navigate in the PM. For instance, the tab "My Vault" has an icon and a label, which is a good practice according to Wiedenbeck [29]; however, its tooltip has the same text as the icon label. This does not help the user as it is redundant. All these tooltips were replaced with more descriptive ones. Lastly, there were some tooltips **missing**, such as the one illustrated in Figure 3(a).

4.5 Lack of consistency

Inconsistencies were found in the behavior of certain buttons that redirect the users to Bitwarden's website. These buttons are in the settings tab and can be separated in two groups: the first group includes the buttons *Premium Membership* and *Two-step Login*; the second group includes, among others, *Import Items* and *Bitwarden Web Vault*. Even though the buttons in these two groups look the same, they present distinct behaviors. When users click a button from the first group, Bitwarden warns them that it will redirect them to its web page and asks for their permission. However, in the second group, Bitwarden redirects to its web page without asking users for their permission. This inconsistency goes against two of the "Golden Rules" of Interface Design as stated by Shneiderman et al. [25]: **Strive for consistency**, stating that actions sequences should be consistent; and **Keep users in control**, stating that some users desire the sense that they are in control. Moreover, it is known that inconsistencies in interfaces can hinder the user experience and users' mental models [6, 25].

To rectify this problem all the buttons from the second group were expanded with a prompt asking for users' permission to redirect them.

5 Evaluation

The work presented in this paper has been tested with user studies and non-structured interviews. In this section, we present the testing methodology and the results.

5.1 User Studies Methodology

To evaluate the success of the solution proposed, we performed user studies to gather insights on: the usability of the solution (if the best practices implemented were successful); if we were able to communicate with users about formal verification; and participants' perceptions on formal verification and PMs.

The user studies are divided into 4 parts. First, we provide users with a brief introduction on what this study is about; tell

them the goals of the study; and how long it will last. After this introduction, we ask users to fill the "Pre-Task Questionnaire", so that we can learn information about their past experience with PMs and demographics. Then, we go through the tutorial and begin the tasks (see Section 5.1.1), they are presented in a random order, and in between each one we ask users to fill a quick "Task Questionnaire" relating to it. When participants finish they fill a "Final Questionnaire". Before ending the session, further feedback from users is collected. Each of the previously mentioned questionnaires is used as a base for an informal interview where we discuss with users the reasons for their answers.

To evaluate the users' perceptions of formal verification in the PM we include questions about it in the questionnaires mentioned above and in the interviews. The answers to the questionnaire use a Likert scale [8, 10] and the answers to the SUS on the "Final Questionnaire" are aggregated to reach a usability score from 0 to 100.

We also register user interaction with the implemented features (e.g. if the user clicks on the formal verification icon and spends time in that screen). We have also performed pilot tests with the aim of refining the testing protocol and script. An example of an improvement suggested was to reduce the number of tasks.

We implemented the previously described protocol in two phases the: i) extended interface, and; ii) base interface. This evaluation is heavily focused on qualitative data and not quantitative. Our quantitative results are subjective, and, as such we used **15 users in total, 10 for the extended interface and 5 for the baseline interface**.

5.1.1 Tasks

In each session, the participants perform the following set of tasks:

- **Login in the PM**: use the primary password and login in the PassCert. In this task, the user goes through the formal verification icon by the primary password field.
- **Register in a website**: register a new user on a website and save the credentials in the PM's vault. This task is one of the most commonly done tasks in a PM, with Bitwarden a pop-up appears prompting users to save the password to the PM, nonetheless, they can dismiss the pop-up and choose to manually add the credential. In this task, users may be exposed to the secure formally verified vault.
- **Generate a random password**: use the PM's generator to generate a new random password. As the password generator is a formally verified feature this task allows us to understand how they integrate with this key feature of the PM.
- **Log in to a website**: login to a website that has a previous password saved in the PM. In this task, users explore the autofill feature of the PM.
- **Update password**: update a password saved in the PM's vault to a new one. Here they explore Bitwarden's vault and can also see the formal verification by the password field.

These tasks are based on the ones implemented by Chiasson et al. [6] in their usability study of PMs. In the user test, the tasks were presented to the users in a random order to prevent bias.

5.2 Usability Results

Our goal is to gather insights on topics that could be used for posterior large-scale quantitative studies.

Regarding the usability of the PM we found that PassCert scored higher than the baseline in SUS. Although these results suggest an improvement they should be repeated with more users to get a more significant statistical result. Nonetheless, the expanded PM's score suggests that our end product is usable. We strongly recommend Bitwarden's team implements some of our extensions such as the tooltips and the tutorial.

Our results also suggest that participants from the extended interface benefited from having gone through the tutorial in the beginning. Tutorials are usually directed at beginner users and most of our users had no previous contact with a browser-extension PM. Moreover, during the think-aloud, users of the extended PM mentioned remembering where a certain feature was because of the tutorial (e.g. where the generator was) and users of the base PM during the informal interview stated that they would have liked to have more support during their first-time use of the PM.

5.3 Perception on formal verification and PMs

In regards to users' perception of formal verification, we found that most (90%) were unfamiliar with the concept before the study. We studied users' perceptions of formal verification before and after using the PM and found that by the end of the study most (80%) associated the concept with *security* and some (20%) were able to give an accurate (but non-technical) explanation about what formal verification is. Formal verification does not always imply security so future research should study in more depth users' perceptions of on this topic.

We wanted to make sure users understood what was formally verified in the PM and this was the case as some participants correctly identified the generator and the storage as formally verified.

A different perspective was found when we asked users to state what would be important for them in a PM, the majority of users stated that they did not want to use a PM that was

open-source, and the reason why revealed a lack of understanding of what open-source software is. We suggest future research should also look into users' perceptions of open-source security software.

Like in literature [20, 26] we found that some participants (60%) reported that they were not aware of the existence of PMs before the study. One user stated "I never saw any add (...) like I see for anti-virus for example". The majority of users also stated that they were open to trying to use PMs in the future. This suggests that unawareness is a barrier to the use of PMs. One of the participants had used a PM in the past but stopped using it because they were not able to understand how it worked stating, "I felt everything was very complicated to do". This is corroborated by the literature where lack of usability was identified as a barrier to effective use of PMs [6, 20].

Adoption of PMs. Our results demonstrate that most users correctly identified the formally verified features. The results also suggest that we have improved the usability of the PM. By improving the usability of the tool and providing formal verification, user adoption of PMs may improve, but further long-term studies must be done in order to gather insights on the impact of formal verification in the adoption of PMs.

6 Conclusion and Next Steps

The advantages of using PMs are undeniable. As such it is important to make users trust and want to use them. This project is a part of the PassCert research project that will build an open-source, proof-of-concept formally verified PM. Previous work from members of the PassCert project identified the need for better user-experience design and thorough usability test of password managers [20].

We have surveyed usable security techniques that can be applied to improve password managers; aimed to ensure that the password managers developed in the context of the PassCert project integrate best practice guidelines developed by the usable security community; explored ways to effectively convey the formally verified properties of the password managers; performed user studies to determine the impact of formal verification on the adoption of password managers; and finally gathered insights on users perception of PMs and formal verification.

In our solution, we proposed ways to educate users about formal verification and increase their trust in the software. Regarding usability, we wanted to implement the relevant usability best practices in the solution such as informing them about what each security feature does.

After implementing the proposed solution we performed user tests where we learned about the usability of the solution, users' perception of PMs, and of formal verification. Our results suggested that our solution has better usability than the base PM. Additionally, some of the insights gathered suggest that there is a general unawareness of both PMs and formal

Figure 4: Formal Verification icon

verification. Moreover, our results suggest that we were able to effectively convey the formally verified features as most users were able to successfully identify them. While users did not present a formal understanding of formal verification in general most associated the concept with security.

PassCert's PM is composed of our interface extensions and the work of other project members to ensure a full PM that aims to provide a usable and secure experience for users. We contribute with first user study on perceptions of formal verification on PMs. We hope our insights can help the formal methods security community better communicate with end-users about its assurances.

6.1 Threats to validity

User studies such as these may suffer from bias. Bias can arise from the questions asked, the questionnaires, the description of formal verification, the formal verification icon (it looks like a green lock, see Figure 4) or even because the name formal verification in plain English (users may associate it with some concept resembling the definition of formal verification without fulling understanding it).

To mitigate these risks we went through a long design process from the descriptions, the icon, and having reviewed the testing protocol with members of the PassCert team. We also made sure to randomize the order of the tasks and provide approximately the same experience for all participants. Nonetheless, problems related to bias can still occur, such as the Dunning-Kruger Effect where users overestimate their skills and knowledge in self-assessments [18]. To mitigate this problem, after asking users to self-assess their knowledge, we always ask them to explain what is their understanding of the topic (e.g. if they state they know what is a PM we ask them to explain what it is to assert their knowledge).

Other biases we must take into consideration when analyzing the results include the Hawthorne effect where users may be inclined to agree with the researchers [19].

Additionally, the sample of participants can also induce bias in the results, and we have a small sample of users in the users' studies. To mitigate this problem we aim for a diverse sample of users. Nonetheless, due to the sample size, we are not able to gather strong statistically significant findings. As stated before this evaluation is heavily focused on qualitative data as we are trying to gather insights on relevant research paths for the future.

6.2 Future Work

Future work could include a translation into Portuguese and this was something mentioned by members of the PassCert team and from users in the user's tests as a barrier to the adoption of PMs.

First time users that want to move all their passwords to a PM face serious barriers and may have to save all their password manually (something that is time and work intensive). So we suggest that further work should be done to study user's transition into PMs and if this has an impact on their adoption.

It is important to mention that our work focused on qualitative data and not quantitative as we have a small sample of users. As such we have gathered insights on topics that could be used for posterior large-scale quantitative studies. Future work could include topics such as users' unawareness of formal verification, misconceptions about PMs, and users' perceptions of formal verification in PM.

Our results seem to indicate that most users are unaware of formal verification as such we suggest that future research should study users' pre-conceptions on formal verification in general. This study could be done in a qualitative way with semi-structured interviews and by focusing on formal verification in different domains.

Formal verification in this study was strictly applied to PMs and the method of transmitting information was also specific to PMs (in this case the formal verification icon). Future research on this topic should study different approaches of conveying formal verification in different contexts (i.e. other formally verified software).

Finally, due to the limited time frame of a master thesis, we were not able to perform longitudinal user studies to measure PM adoption. To understand the impact that formal verification has on adoption and user retention in PMs, future work on this topic should include long-term user studies.

Acknowledgments. This work was partially funded by the PassCert project, a CMU Portugal Exploratory Project funded by Fundação para a Ciência e Tecnologia (FCT), with reference CMU/TIC/0006/2019 and supported by national funds through FCT under project UIDB/50021/2020.

References

[1] N Alkaldi and K Renaud. "Why do people adopt, or reject, smartphone password managers?" In: *The 1st European Workshop on Usable Security (EuroUSEC 2016)*. 2016.

[2] P. Arias-Cabarcos, A. Marín, Palacios, F. Almenárez, and D. Díaz-Sánchez. "Comparing password management software: toward usable and secure enterprise authentication". In: *IT Professional* 18.5 (2016), pp. 34–40.

[3] *Bitwarden Open Source Password Manager*. [Online; accessed 12-December-2020]. URL: bitwarden.com/.

[4] Carolina Carreira, João F Ferreira, and Alexandra Mendes. "Towards Improving the Usability of Password Managers". In: *INFORUM* (2021).

[5] Carolina Carreira, João F Ferreira, Alexandra Mendes, and Nicolas Christin. "Exploring Usable Security to Improve the Impact of Formal Verification: A Research Agenda". In: *First Workshop on Applicable Formal Methods (co-located with Formal Methods 2021)*. (2021).

[6] S. Chiasson, P. C. van Oorschot, and R. Biddle. "A Usability Study and Critique of Two Password Managers." In: *USENIX Security Symposium*. Vol. 15. 2006, pp. 1–16.

[7] European Union Agency for Cibersecurity. *Authentication Methods*. [Online; accessed 12-December-2020]. URL: www.enisa.europa.eu/topics/csirts-in-europe/glossary/authentication-methods.

[8] A. Dix, J. Finlay, G. D. Abowd, and R. Beale. *Human-computer interaction*. Pearson Education, 2004.

[9] João F. Ferreira, Saul Johnson, Alexandra Mendes, and Phillip Brooke. "Certified Password Quality: A Case Study Using Coq and Linux Pluggable Authentication Modules". In: *13th International Conference on Integrated Formal Methods*. 2017.

[10] Manuel J Fonseca, Pedro Campos, and Daniel Gonçalves. "Introdução ao design de interfaces". In: *FCA-Editora de Informática* (2017).

[11] S. Gaw and E. W Felten. "Password management strategies for online accounts". In: *Proceedings of the 2nd symposium on Usable privacy and security*. 2006, pp. 44–55.

[12] Miguel Grilo, João F. Ferreira, and José Bacelar Almeida. "Towards Formal Verification of Password Generation Algorithms used in Password Managers". In: *arXiv preprint arXiv:2106.03626* (2021).

[13] Aurora Harley. *Usability Testing of Icons*. Jan. 2016. URL: www.nngroup.com/articles/icon-testing/.

[14] Cormac Herley and Paul C. van Oorschot. "A Research Agenda Acknowledging the Persistence of Passwords". In: *Published in IEEE Security and Privacy Magazine, Volume 10 Issue 1, Jan.-Feb.* IEEE. 2012, pp. 28–36.

[15] Iulia Ion, Rob Reeder, and Sunny Consolvo. "...no one can hack my mind: Comparing Expert and Non-Expert Security Practices". In: *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. 2015, pp. 327–346.

[16] Saul Johnson, João F. Ferreira, Alexandra Mendes, and Julien Cordry. “Skeptic: Automatic, justified and privacy-preserving password composition policy selection”. In: *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*. 2020, pp. 101–115.

[17] A. Karole, N. Saxena, and N. Christin. “A comparative usability evaluation of traditional password managers”. In: *ICISC*. Springer. 2010, pp. 233–251.

[18] Khalid Mahmood. “Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect”. In: *Communications in Information Literacy* 10.2 (2016), p. 3.

[19] Frank Merrett. “Reflections on the Hawthorne effect”. In: *Educational Psychology* 26.1 (2006), pp. 143–146. DOI: 10.1080/01443410500341080.

[20] S. Pearman, S. A. Zhang, L. Bauer, N. Christin, and L. F. Cranor. “Why people (don’t) use password managers effectively”. In: *Fifteenth Symposium On Usable Privacy and Security (SOUPS 2019). USENIX Association, Santa Clara, CA*. 2019, pp. 319–338.

[21] H. Ray, F. Wolf, R. Kuber, and A. J. Aviv. “Why Older Adults (Don’t) Use Password Managers”. In: *30th USENIX Security Symposium)*. 2021.

[22] J. G. Redish. *Letting go of the words: Writing web content that works*. Morgan Kaufmann, 2012.

[23] M. A. Sasse and I. Flechais. “Usable security: Why do we need it? How do we get it?” In: O’Reilly, 2005.

[24] S. Seiler-Hwang, P. Arias-Cabarcos, A. Marín, F. Almenares, D. Díaz-Sánchez, and C. Becker. ““I don’t see why I would ever want to use it”:Analyzing the Usability of Popular Smartphone Password Managers”. In: *Proceedings ACM SIGSAC CCCS*. 2019, pp. 1937–1953.

[25] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos. *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016.

[26] Elizabeth Stobert and Robert Biddle. “The password life cycle: user behaviour in managing passwords”. In: *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. 2014, pp. 243–255.

[27] Jenifer Tidwell. *Designing interfaces: Patterns for effective interaction design*. " O’Reilly Media, Inc.", 2020.

[28] A. Whitten and J. D. Tygar. “Why Johnny Can’t Encrypt: A Usability Evaluation of PGP 5.0.” In: *USENIX Security Symposium*. Vol. 348. 1999.

[29] Susan Wiedenbeck. “The use of icons and labels in an end user application program: An empirical study of learning and retention”. In: *Behaviour & Information Technology* 18.2 (Jan. 1999), pp. 68–82. ISSN: 0144-929X, 1362-3001. DOI: 10.1080/014492999119129.