A Comparative Usability Evaluation of Traditional Password Managers

Ambarish Karole[1], Nitesh Saxena[1], and Nicolas Christin[2]

[1] Polytechnic Institute of New York University
[2] Carnegie Mellon University

Abstract. Proposed in response to the growing number of passwords users have to memorize, password managers allow to store one's credentials, either on a third-party server (online password manager), or on a portable device (portable password manager) such as a mobile phone or a USB key. In this paper, we present a comparative usability study of three popular password managers: an online manager (LastPass), a phone manager (KeePassMobile) and a USB manager (Roboform2Go). Our study provides valuable insights on average users' perception of security and usability of the three password management approaches. We find, contrary to our intuition, that users overall prefer the two portable managers over the online manager, despite the better usability of the latter. Also, surprisingly, our non-technical pool of users shows a strong inclination towards the phone manager. These findings can generally be credited to the fact that the users were not comfortable giving control of their passwords to an online entity and preferred to manage their passwords themselves on their own portable devices. Our results prompt the need for research on developing user-friendly and secure phone managers, owing to the ubiquity of mobile phones.

1 Introduction

Typical credentials employed for user authentication fall into following categories of authentication "factors": (1) "*Something You Know*," such as passwords or PINs, (2) "*Something You Have*," such as a token or a card, and (3) "*Something You Are*," such as biometrics; or combinations thereof. Of these, passwords or PINs are the most widely deployed, for authentication to remote servers, ATMs and mobile phones.

For over more than a decade, users have been asked to memorize an increasing number of passwords [1] to authenticate to various online services. While users can usually easily memorize a couple of passwords, the current explosion of the number of passwords each user has to maintain is severely testing the limits of their cognitive abilities [2]. This leads to "weak" choices in practice. For example, users often tend to choose short and "low-entropy" passwords [3, 4], enabling offline dictionary attacks and brute-forcing attempts, or they write passwords down or use the same password at multiple sites [5].

Password Managers (PMs) attempt to solve this conundrum by having a computing device, rather than the user herself, store (and optionally, generate) passwords, and then later deliver or recall them to the user whenever access is needed. To this end, a number of password management schemes have been proposed and are used currently.

We can broadly distinguish between three categories of password managers: desktop manager, online manager and portable manager. A desktop manager (e.g., Mozilla Firefox, Apple MacOS Keychain, RoboForm [6]) stores strong passwords on the user's desktop (i.e., on the terminal used for authentication) while an online manager (e.g., LastPass [7] and Mozilla Weave Sync [8]) stores them on remote third-party server(s).[3] A portable manager, on the other hand, stores strong passwords on user's portable device. Among portable managers, we can further identify two different types: phone-based password managers (e.g., KeePassMobile for J2ME enabled devices [10] and OpenIntents Safe for Android [11]) and USB-based password managers (e.g., Roboform2Go for USB devices [6]).

In each of these approaches, the strong passwords are typically protected using a master password; at the time of recalling a specific password, the user simply types in her master password. If a user is mobile and uses multiple terminals for authentication (e.g., her desktop at home and her laptop in the office), a desktop manager would not offer any portability to the user. We, therefore, do not consider desktop managers to be of much benefit on their own.

The online and portable managers have their own pros and cons. An online manager, although portable, requires the user to trust the third-party service provider(s). Since user's passwords would typically be encrypted using her master password and then stored on remote server(s), they might be vulnerable to offline dictionary attacks. Imagine if all users were to use a remote manager, the passwords corresponding to all of them might be susceptible to an adversarial break-in at the end of the server(s). Moreover, often proprietary, a remote manager might not offer the users any transparency in outsourcing their sensitive information and how this information has been protected.

A portable manager can possibly be more trusted since it can be locally managed by the user on her own trusted portable device. However, all existing phone managers typically involve displaying a (long and possibly random) password on the portable device, which the user is simply asked to copy onto the terminal. Typing in a such a password might have poor usability. USB managers do not have this drawback, but they may not offer a desired level of portability and accessibility to a modern user.

The goal of this paper is to formalize an evaluation of existing password managers, by comparing them in terms of security, ease of use, necessity and level of acceptance, as perceived by an average web user. To that effect, we present a comparative usability study of three popular password managers: an online manager (LastPass), a phone manager (KeePassMobile) and a USB manager (Roboform2Go).

Our study was performed with a sample of users controlled with respect to technical background (i.e, computer science students vs. non-technical "average" users). We find, contrary to our intuition, that users overall preferred the two portable managers over the online manager, despite the better usability of the latter. Surprisingly, the online manager was the last choice for non-technical people, who mostly preferred the phone manager. Also, technical people were more inclined towards the USB manager in comparison to the online manager. These findings can generally be credited to the fact

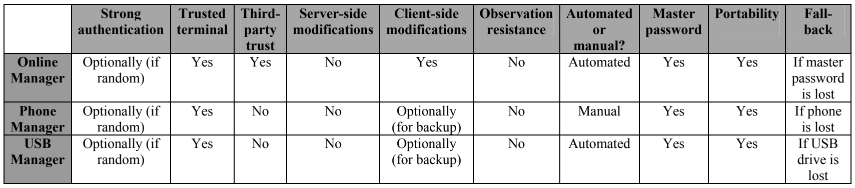
[3] Rather than storing passwords, another password management approach (e.g., PwdHash [9]) derives passwords on-the-fly, based on a master password and a specific variable, e.g. the URL of the website to authenticate to. From the usability perspective, this approach and desktop/online managers are equivalent, in that they only require a master password to be memorized/recalled.

that the users were not comfortable giving control of their passwords to an online entity and preferred to manage their passwords themselves on their own portable devices.

We note that the only prior work that directly relates to our study, to the best of our knowledge, is by Chiasson et. al [12]. The study [12] evaluates two desktop managers – PwdHash [9] and Password Multiplier [2], and points out underlying usability problems with these two managers. Our study, on the other hand, aims at evaluating and comparing three different types of traditional password management approaches, with a particular focus on mobile users.

2 Background and Research Questions

In this section, we discuss the three password managers in more details and compare them based on their usability and security characteristics. This background information will serve as a foundation to frame the research questions that we aim at answering via our study, and to come up with the usability and security measures across which the password managers will be compared. We provide a side-by-side comparison of an online manager, a phone manager and a USB manager in Figure 1.



	Strong authentication	Trusted terminal	Third-party trust	Server-side modifications	Client-side modifications	Observation resistance	Automated or manual?	Master password	Portability	Fall-back
Online Manager	Optionally (if random)	Yes	Yes	No	Yes	No	Automated	Yes	Yes	If master password is lost
Phone Manager	Optionally (if random)	Yes	No	No	Optionally (for backup)	No	Manual	Yes	Yes	If phone is lost
USB Manager	Optionally (if random)	Yes	No	No	Optionally (for backup)	No	Automated	Yes	Yes	If USB drive is lost

Fig. 1. Comparison of Password Management Methods

As discussed in Section 1, online password managers incorporate remote third-party servers for password storage. Portable managers, on the other hand, consists of a credential listing on users' personal portable devices, e.g., a mobile phone and USB drive.

One example of software that falls into the category of online manager is LastPass [7]. LastPass is a proprietary extension for the Mozilla Firefox web browser which locally encrypts user credentials using 256 bit AES prior to transmitting them to LastPass's data centers via SSL. Though their key generation algorithm is not described, LastPass's encryption and decryption is protected using a master password which is not transmitted beyond the local terminal. A similar online password management extension for Firefox is Mozilla Weave Sync [8]. Weave is an open source solution which operates by encrypting browser data with asymmetric cryptography; this allows users to share selected browser data with others if desired. Though each user's private key is stored locally as well as on remote Weave servers, in both cases this key is encrypted with a user specified passphrase. As is the case with LastPass's master password, this passphrase is used locally and not transmitted to or stored on the remote server.

These online password managers introduce some drawbacks, however. Foremost among these is the issue of trust. This class of managers asks users to trust a remote server or group of servers with their sensitive data. When a remote server is employed, the password encrypted with a master password is sent across the internet, making it much more likely for a malicious entity to capture and store it for later offline dictionary attacks (master password is still user-chosen). Furthermore, should an adversary manage to break in to one of these servers they would be able to gain access to all the encrypted passwords for every user stored on that server. Again, the fact that these credentials are stored as ciphertexts alleviates this issue somewhat, but the threat of a later offline attack on this data remains. In contrast, an offline attack on a portable password manager of a user only exposes that particular user's passwords.

An additional consideration pertaining to remote credential storage is the flexibility of authentication. Because these remote servers manage passwords for many users, authentication with a user name and password prior to credential retrieval is a necessity. Portable managers, on the other hand, never requires a user name due to the personal nature of a user's mobile device.

Also, as noted in [13], there are several flaws and challenges associated with with managing credentials through remote servers. Although users desire the additional security benefits online servers can provide, users are unwilling to compromise on usability to improve security. Thus remote servers must be careful not to add security at the cost of detracting from the overall user experience. Client side software must be easy to download and install, and should be tightly integrated with the browser or operating system to prevent users from cutting corners that could potentially lead to social engineering attacks.

Several portable managers exist for various mobile phone platforms, such as KeePassMobile for J2ME enabled devices [10] and OpenIntents Safe for Android [11]. While uncomplicated, users of these alternatives must *manually* transfer their password by reading it off their mobile device and typing it on their terminal's keyboard. This may be clumsy in terms of usability, but also restricts the security of the password management solution by limiting the length of passwords that can be used to that which a user is capable of correctly reading and typing during each authentication.

USB managers (e.g., RoboForm2Go [6]), being personal, offer a similar level of trust as provided by phone managers. One potential advantage of a USB manager over phone manager is that the password recalling process is automated. However, mobile phones appear, at first glance, potentially more appealing to users. USB devices indeed do not serve any additional purpose other than providing data storage, while mobile phones are increasingly playing the role of a "digital swiss army knife."

Strong authentication in existing passwords managers is achieved through the use of randomly generated password strings. Most existing solutions provide users with the option of either storing their pre-existing, non-random credentials or generating new random passwords at registration time. If existing passwords are stored then the solution does not provide any measure of additional security, only the convenience of password recall.

All password management approaches trust the intermediary terminal with the user's plaintext credentials, i.e., passwords. This is due to the inherent difficulty of authenticating without introducing server-side modifications.

Our discussion above raises several questions that we intend to answer through our study. These include:

- How do the three PMs compare in terms of usability? The usability can be measured with respect to perceived toughness, satisfaction and ease of use.
- How do the three PMs compare in terms of security and protection of passwords? This covers giving control of passwords to a program and perceived security.

- How do the three PMs compare in terms of their perceived necessity and acceptance? In other words, would the users be willing to adopt them in practice?
- How do the three PMs compare in terms of all security and usability measures taken together?
- How do the three PMs compare across a diverse set of users categorized based on background (technical or non-technical)? Also, what is the effect of different users' background on each PM?

3 Study Preliminaries

Password Manager Implementations: Our goal is to compare the three PMs – USB manager (denoted as USB henceforth), phone manager (Phone) and online manager (Online) – in terms of their usability and security, as perceived by average users. We also intend to evaluate each PM according to several underlying tasks, including registration, login from a personal computer, login from a remote computer, change password, and login with a changed password (these tasks will be explained in Section 4.2). This implies that each user would need to execute all these tasks to evaluate a PM, which might lead to a lengthy overall experimentation period per user. This in turn might cause user fatigue and influence the results of the study. To avoid this, it was paramount that no more than one PM of each type (USB, Phone and Online) is selected for the study. This necessitated that only those PM implementations are selected that are representative of their respective PM category.

As discussed previously, a number of commercial and popular options exist that can be used in our study. These include (to name a few) LastPass [7] and Mozilla Weave Sync [8] as Online managers; KeePassMobile for J2ME enabled devices [10] and OpenIntents Safe for Android [11] as Phone managers; Roboform2Go [6] and HandyPassword [14] as USB Managers. Numerous other implementations exist, as listed in an online survey of PMs [15]. Fortunately, the user actions involved in all PM implementations of a given category are roughly very similar to one other. In other words, for example, to login using any of these USB Managers, the user simply needs to connect her USB drive to the USB port of her computer terminal, and type in her master password to unlock the password to be recalled. To login using any of the Phone managers, the user needs to first unlock her phone with a master password and then copy the password – to be recalled – displayed on the phone's screen onto the keypad of the terminal. Similarly, in order to login using an Online Manager, the user only needs to type in her master password on the terminal, the rest of the process being farily oblivious to the user. The only distinction among these PMs are the underlying software interfaces.

According to the reviews available online [15][16], we chose Roboform2Go as our USB manager, KeePassMobile as the Phone manager and LastPass as the Online manager. Based on their popularity, we believe these three PMs are quite suitable for our usability study which aims at comparing the three PM categories (USB, Phone and Online). We also believe that our selection, and our use of existing and deployed implementations was a better approach than trying to pursue our study with our own (likely unpolished) research prototypes of the PMs.

Devices: We used common devices that most users are quite familiar with. We used Imation 2GB USB 2.0 thumb drive [17] – as our USB manager – with RoboForm2Go software. We chose Nokia 5310 mobile phone [18] as our Phone manager installed with KeePassMobile. We used a Dell Desktop as our primary authentication terminal and a Sony Laptop for the purpose of login from another terminal (see Section 4.2).

Browser: Based on its popularity [19], Mozilla Firefox was used as the Internet browser throughout our study. Participants were instructed to authenticate, using the three password managers, to a popular web email service – Gmail.

4 Usability Testing Details

Having made a selection of a password manager for each category (as discussed in Section 3), we are now ready to start the usability study. The most obvious method to record responses from a user is through the use of a 5-point Likert scale, in addition to open-ended and multiple choice personal preference questionnaires. The questionnaires were handed over to a user depending on which stage of testing he/she was at. The *During Test* questionnaire was posed after the respondent finished performing each one of the five tasks common to all the three password managers (these tasks will be discussed in Section 4.2). The *Post Test* questionnaires, on the other hand, were asked after all the three password managers had been tested by each user. Based on our discussion in Section 2, we decided to evaluate and compare the password managers with respect to the following usability and security measures. (A similar set of measures have previously been used in the study of [12]).

• During Test –

1. **Toughness**: how tough it was to execute each task? (1 question was posed)
2. **Satisfaction**: how satisfied the users felt with each task? (1 question was posed)

• Post Test –

1. **Giving Control**: how users felt while giving control into the hands of a software/tool to manage their passwords? (4 questions were posed)
2. **Perceived Ease**: did users find the password manager easy to use? (5 questions were posed)
3. **Perceived Necessity**: did users deem the password manager necessary and acceptable? (2 questions were posed for all PMs. For Phone and USB, 1 additional question was posed regarding the accessibility to mobile devices.)
4. **Perceived Security**: did users find the password manager secure? (4 questions were posed)

The users were also posed a few open-ended questions, in each of the above questionnaires, in order to poll for their opinions about any perceived problems with the password managers and suggestions for possible improvement.

Finally, a *Final Test* questionnaire was also presented to each user polling which password manger he/she preferred the most and asking about their order of preference based on the level of (1) security, (2) convenience and (3) overall experience.

The main challenge we faced was the sheer number of questions which each user needed to answer, potentially leading to lazy respondent behavior and user fatigue. Care was taken so as to minimise both the number of questions and to discard any questions which showed a tendency of not being answered genuinely (or honestly).

4.1 Study Participants

We recruited a total of 20 participants: 10 technical and 10 non-technical users. In the rest of this paper, we will refer to our technical users as Students, and non-technical users as Non-Students, because all technical people were students while all non-technical people were non-students. The participants were recruited on a first-come first-serve basis from respondents to emails. Prior to recruitment, each participant was briefed on the estimated amount of time required to complete the tests and on the importance of completing the tests in its entirety.

The student participants were all university students, studying towards undergraduate, Master's and Ph.D. degrees in Computer Science or closely related fields. This group of our users represented a fairly young, well educated and technology-savvy sample of user population.

The other group, consisting of the non-students, had an average age difference of nearly two decades from that of the students. This group was tested to gain insights into whether such a group – differing in terms of full time occupation – had any impact on the choices made with respect to the password managers.

There is an obvious concern that, if a technology-savvy group (students) does not react well to a password management approach, the approach will perform a lot worse with average users; or on the contrary, if a password manager that fares well with students, it might not perform equally well with average users. This concern was our prime motivation to categorize the respondents into students and non-students.

Our non-students ranged from help-desk personnel, technicians, real estate agents, restaurant workers to housewives. In addition to the students vs. non-students distinction, our sample was also controlled, as much as possible, in terms of other important user-centric characteristics, i.e., gender and age. This was done in order to evaluate the password managers among a diverse user population pool. The gender split was: 60% male and 40% female for both students and non-students. Our test users were divided into three age groups: 40% Young (less than 24 years old), 40% Middle-Aged (25-44 years old) and 20% Old (45-54 years old). In addition to the students and non-students category, we have also pursued gender-centric and age-centric analysis. However, due to space constraints, we only restrict ourselves to the former in the rest of this paper, which we believe is most important to our study.

Gender, age and other information was collected through a *Pre Test* questionnaire completed by our participants prior to starting the test process. None of the study participants reported any physical impairment that could interfere with their ability to complete a given task.

4.2 Testing Process

Our study was conducted at two testing locations, one on-campus (at our university) for the students and the other off-campus for non-students. These two venues were chosen solely for the purpose of convenience to the targeted participant groups. Same devices (USB drive and phone) and computer terminal (see Section 3) were used at both locations giving rise to consistent test set-up across all users. Our study lasted for a duration of about two months.

An overview of the testing process was given to each respondent prior to the study and due care was taken to minimise any scope of *explicit* "priming" of respondents considering a security-focused nature of our study.[4] Such a priming in terms of security can possibly result in skewed (over-alert) participant behavior and in biased results, as demonstrated by prior research [20].

As mentioned previously, after administering the Pre Test questionnaire, the respondents were asked to perform five tasks corresponding to each password manager. Any possible user errors in performing the above tasks were taken note of by the test administrator (no such errors were observed throughout our study, however).

1. **Register** involves registering with a password manager the password, username and other information for a particular web site.
2. **Login** involves login to a web site, whose password has already been saved with a password manager.
3. **Second Login** is similar to the Login task, only difference being the computer is not the same as the one used in the previous task. This task is aimed at judging the portability of the password manager from terminal to terminal.
4. **Change Password** involves changing the password, both with the website and password manager.
5. **Login with New Password** involves repeating the login task but with the new password.

As mentioned in Section 3, the test set-up comprised of a desktop computer which acted as the primary computer for Login, Change Password and Login with New Password, and a laptop for Second Login. This set-up, consisting of two computers, was used in order to closely mimic the tasks akin to a realistic password manager setting.

A randomly chosen 8 character master password was provided to each test user, which he/she was asked to memorize and use throughout the experiments.

The respondents were instructed, in advance, to fill-in During Test questionnaires after each of the above task was completed. As mentioned earlier, the questionnaire consisted of two simple Likert Scale type questions and two open-ended questions. The order in which the password managers were presented to the users was randomized so as to avoid any possible learning effects.

Following the During Test, the respondents were required to fill out the Post Test questionnaire for each password manager. This too comprised of Likert Scale questions followed by a few open ended questions. The purpose of this questionnaire was to judge the changes in attitudes (opinions) of the respondents towards the password managers after having worked with all three of them.

Finally, the last part of the questionnaire (Final Test) was administered to the respondents. Here the participants ranked the best of the three password managers which they felt were most appealing with respect to their overall experience with them, the level of convenience and security provided. This part of the questionnaire also consisted of a few open-ended questions. These were aimed to better understand the motivating reasons for a respondent to choose a particular password manager, which could not be captured by the Likert scales or the multiple choice questions.

[4] Since the study was about password managers, it was neither possible nor meaningful to avoid implicit priming.

5 Test Results and Analysis

In this section, we present and analyze our During Test and Post Test Likert scale logged observations. We also discuss the final preferences provided by our test subjects for the three PMs evaluated. We present two types of comparison in our analysis throughout:

- **Within-Subjects Comparisons:** This analysis would tell us how the three password managers (USB, Phone and Online) fare with one another, corresponding to the entire user sample as well as corresponding to students and non-students.
- **Between-Subjects Comparisons:** Using this analysis, we intend to understand the effect of occupation (student vs. non-student) on the usability and security of three PMs.

Recall that the During Test data is aimed at evaluating the usability of each PM in terms of two measures: Toughness and Satisfaction. The Post-Test data, on the other hand, is for investigating the PMs with respect to usability and security measures: Giving Control, Perceived Ease, Perceived Necessity and Perceived Security. We analyze the PMs based on these individual measures first, followed by a principle component and cluster analysis that evaluates the combined effect of different measures.

In the remainder of this section, we discuss our results and their interpretations. Unless stated otherwise, statistical significance is reported at the 5% level.

5.1 During Test Analysis

Figure 2 shows the *average* Likert ratings regarding Toughness and Satisfaction of the three PMs, for Students and Non-students (also shown are the collective average ratings for All Users taken together as well as those corresponding to All PMs).

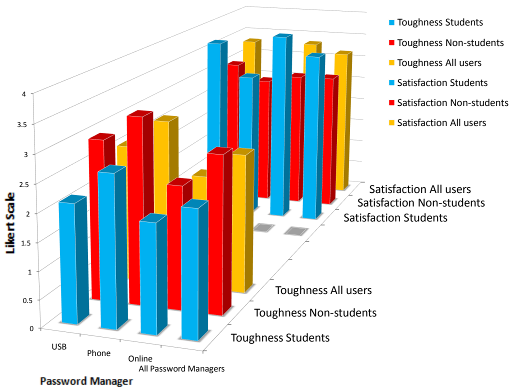


Fig. 2. During Test Toughness and Satisfaction

Within-Subjects Comparisons: Observing the bars of the graph along the X-axis, we find that Phone is deemed the toughest, followed by USB and then Online, for all our users (Students, Non-Students, All Users). In terms of satisfaction, on the other hand, among both Students and All Users, Online was preferred over USB, followed by

Phone. For Non-Students, however, USB was the first choice, preferred slightly more than Online and Phone. These results are intuitive because both Online and USB require a minimal amount of effort from the users and are supposed to be quite fast in comparison to Phone due to manual transfer of password.

Based on paired t-tests, we found the following statistical differences. Students found Phone tougher than USB ($p = 0.0103$) and tougher than Online ($p = 0.0028$). Students also found USB more satisfying than Phone ($p = 0.0006$), Online more satisfying than USB ($p = 0.0238$) and Online more satisfying than Phone ($p < 0.0001$). Non-Students also found Phone tougher than USB ($p = 0.009$) and tougher than Online ($p = 0.003$), and USB tougher than Online ($p = 0.020$). In terms of satisfaction levels for Non-Students, we did not find any statistical difference; the ratings were quite close for different PMs. For All Users, Phone was deemed tougher than USB ($p = 0.043$) and Online ($p = 0.00013$), and USB was found to be tougher than Online ($p = 0.0444$). Also, for All Users, USB was more satisfactory than Phone ($p = 0.0498$), and Phone was more satisfactory than Online ($p = 0.009$).

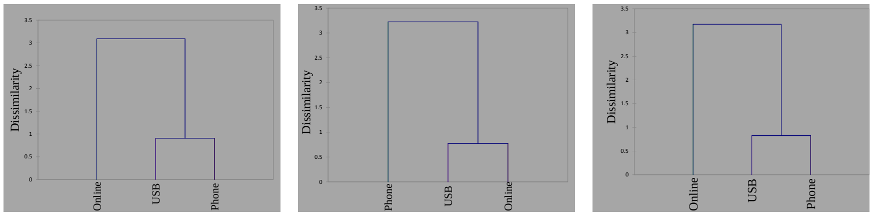
Between-Subjects Comparisons: Observing the graph of Figure 2 along the Y-axis, we notice that Non-Students consistently found the three PMs tougher to use when compared to Students. Likewise, Students found the three PMs more satisfying than Non-Students. The reason for this is simple: Students are expected to be much more technologically savvy compared to Non-Students.

Based on unpaired t-tests, following significant differences were noticed: Non-Students found USB tougher ($p = 0.0008$) and less satisfactory ($p < 0.0001$) compared to Students. Non-Students also found Phone tougher compared to Non-Students ($p = 0.003$). Students were highly more satisfied with Online ($p < 0.0001$).

Usability Measures Taken Together: In the previous subsection, we considered the usability of PMs in terms of two measures: Toughness and Satisfaction. Although the two measures were usually negatively correlated with each other (Pearson correlation coefficient was found to be -0.771, when considering data from all users), in certain cases the correlation was not entirely clear. In order to understand an overall impact of Toughness and Satisfaction on the usability of PMs, we pursued principle component (PCA) and cluster analysis. Due to ease of readability, we do not include the details regarding this analysis (a similar analysis, however, is later discussed for Post Test measures in Section 5.2). We only depict the results (using Agglomerative Hierarchical Clustering) of this analysis in Figure 3.

For All users, we obtain that Online $\succ$ USB $\succ$ Phone, and USB and Phone are clustered together (we use '$\succ$' to denote preference). A similar and independent PCA and cluster analysis for Students and Non-Students indicate the following. For students, Online $\succ$ USB $\succ$ Phone, and Online and USB form a cluster of their own. On the other hand, for Non-Students, Online $\succ$ USB $\succ$ Phone, and USB and Phone are clustered with each other. These results are intuitive and very much inline with our observations shown in Figure 2, which we discussed in the previous subsection.

Usability of Individual Tasks: As we explained in Section 4.2, in our experiments, each PM was tested for several different processes, namely, Register, Login, Second Login, Change Password, Login with New Password. Since usability of a PM depends on all these processes, we compare the three PMs based across these processes.



(a) All Users (b) Students (c) Non-Students

Fig. 3. During Test Cluster Analysis Based On Principal Components (Dissimilarity vs. PM)

Figure 4 depicts the average Likert scale Toughness ratings for different processes corresponding to Students, Non-Students and All Users. In this plot, first three bars for each process correspond to Students (USB, Phone, Online, resp.), next three bars correspond to Non-Students (USB, Phone, Online, resp.) and last three bars correspond to All Users (USB, Phone, Online, resp.). The Satisfaction ratings were generally inversely related to the Toughness ratings and are not shown in this paper.

Let us first compare the three PMs across different processes. We note that for each process, in general, Phone is tougher than the other two PMs. Between USB and Online, the former is deemed tougher, for all processes. This analysis conforms well with our overall analysis of During Test data presented in previous subsections.

There are a few exceptions to the above claim, however. Login and Change Password have the same average ratings for both USB and Online for Students Students deemed Login with New Password as equally tough for USB and Phone. Register was also rated at a equal level of toughness by Non-Students. For Second Login, Students found USB less tougher than Online. For Second Login and Change Password, Non-Students rated Phone as only slightly tougher than USB.

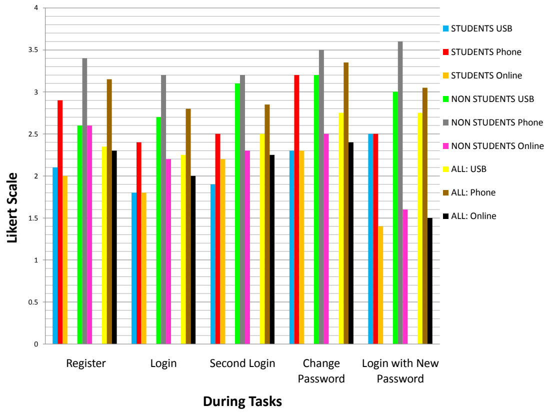
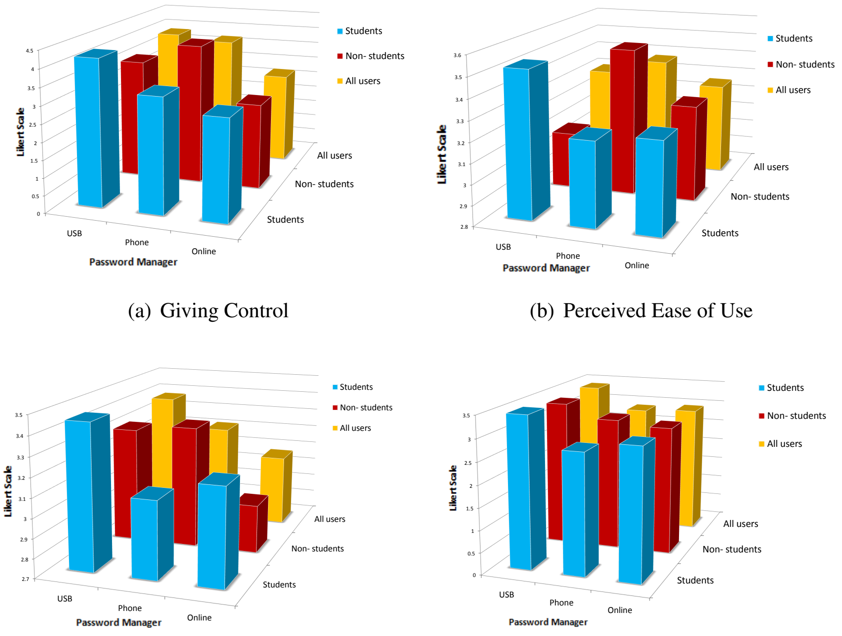


Fig. 4. During Test Toughness Per Task

5.2 Post Test Analysis

Figure 5 shows the average Likert Post Test ratings regarding Giving Control, Perceived Ease, Perceived Necessity, and Perceived Security, for Students and Non-students (also shown are the collective ratings for All users taken together). We discuss the observations made from these ratings as follows.



(a) Giving Control

(b) Perceived Ease of Use

(c) Perceived Necessity

(d) Perceived Security

Fig. 5. Post Test – Students vs. Non-Students

Within-Subjects Comparisons:

• **Giving Control:** Looking at the Giving Control ratings (Figure 5(a)), we find that Students order of preference is USB, followed by Phone and Online. Non-Students, on the other hand, like the Phone the best, followed by USB and Online. Collectively looking at All Users, USB is an overall winner, which seems slightly better than Phone, which in turn is much better than Online. In general, users felt that USB and Phone provide a better sense of control compared to Online. This is a surprising finding and is perhaps due to the fact that managing the passwords locally on their own devices gave users a sense of control and authority.

Based on paired t-tests, we found the following statistical differences. Students found USB better than Phone ($p = 0.0049$) and USB better than Online ($p = 0.016$). Non-Students preferred Phone over Online ($p = 0.0001$), and USB over Online ($p = 0.0009$). All Users prefer USB over Online ($p < 0.0001$), and Phone over Online ($p = 0.00024$).

• **Perceived Ease:** Looking at the Perceived Ease ratings (Figure 5(b)), we find that Students order of preference is USB, followed by Online and then Phone. Non-Students, on the other hand, like the Phone the best, followed by Online and then USB. Collectively looking at All Users, Phone is an overall winner, which is slightly better than USB, which in turn is slightly better than Online. Here we can see a clear split across Students and Non-Students: the former still preferred Online or Phone, whereas the latter found the Phone as the easiest. Paired t-tests, however, did not lead to any significant statistical differences with respect to Perceive Ease.

• **Perceived Necessity:** Looking at the Perceived Necessity (Figure 5(c)), we find that Students order of preference is USB, followed by Online and then Phone. Non-Students, on the other hand, like the Phone the best, followed by USB and then Online. Collectively looking at All Users, USB is an overall winner, which is somewhat better than Phone, which in turn is quite better than Online. These findings are somewhat similar to those in case of Perceived Ease, which means that necessity of a PM was based on its ease. Based on paired t-tests, Non-Students found USB better than Online ($p = 0.0448$). No other significant statistical differences were found.

• **Perceived Security:** From the average ratings corresponding to Perceived Security (Figure 5(d)), we can see that USB is generally preferred by all users, and there is not much to choose between Phone and Online (although Students prefer Online slightly more so than Non-Students, who slightly prefer Phone).

According to paired t-tests, Students found USB better than Phone ($p = 0.0132$), and USB better than Online ($p = 0.03$). All Users found USB more secure than Phone ($p = 0.0089$) and USB more secure than Online ($p = 0.012$). No other significant statistical differences were found.

Between-Subjects Comparisons:

• **Giving Control:** Looking at the Giving Control ratings (Figure 5(a)), Students prefer USB and Online more than Non-Students, however, Non-Students are more inclined to use Phone compared to Students. Based on unpaired t-tests, Non-Students preferred Phone more than Students ($p = 0.011$).

• **Perceived Ease:** From Perceived Ease ratings (Figure 5(b)), we observe that Students prefer USB much more than Non-Students, however, Non-Students are much more inclined to use Phone compared to Students. Both somewhat equally prefer Online.

• **Perceived Necessity:** Looking at the Perceived Necessity (Figure 5(c)), Students prefer USB and Online more than Non-Students, however, Non-Students are more inclined to use Phone compared to Students. This is very similar to users' perception of ease as discussed above.

• **Perceived Security:** Looking at the Perceived Necessity (Figure 5(d)), there is not much distinction between the rating of Students and Non-Students. For Phone, however, Non-Students provided higher ratings.

Accessibility to Portable Devices: In response to whether the users would have their phone and USB drive handy while accessing a web site, the average ratings (with standard deviations) were as shown in Table 1. The ratings imply that there is not much to choose between USB drive and phone when looking at All Users. Students, on the other hand, rated USB drive as more accessible compared to phone, whereas Non-Students gave higher ratings to phone. This is perhaps one of the reasons why Students had a stronger inclination towards using USB PM and why Non-Students preferred Phone.

All Usability and Security Measures Taken Together: A usable PM should perform well with respect to all (not just one of the) usability measures we discussed so far,

	All Users	Students	Non-Students
USB	3.4 (1.095)	3.8 (0.422)	3 (1.414)
Phone	3.4 (1.046)	3.1 (0.876)	3.7 (1.160)

Table 1. Average ratings (with standard deviation) for accessibility of USB and Phone

i.e., Giving Control, Perceived Ease, Perceived Necessity and Perceived Security. To this end, we performed linear cross-correlations among the PMs across these usability measures. We first present a complete analysis over data acquired from all test subjects. Table 2 shows the correlation coefficients and their respective statistical significance.

	Control	Ease	Necessity	Security
Control	1	0.287	0.130	0.337
Ease	0.287	1	0.248	0.368
Necessity	0.130	0.248	1	0.374
Security	0.337	0.368	0.374	0.287

Table 2. Cross-Correlation of Usability Measures

The coefficients from less than -0.5 and more than 0.5 are generally regarded as large [21] and in line with the findings of [22], we cannot regard any of our usability measures as sufficiently correlated with others that they could be justifiably omitted. On the other hand, since the measures are mildly correlated, it motivates us to also look at them as a whole as we show next.

Principal Component and Cluster Analysis: Table 3 lists the four principal components, denoted PC1, PC2, PC3 and PC4, that explain 100% of the variance in the logged data. As the first two components, i.e., PC1 and PC2, together explain nearly 70% of the variance, and PC3 and PC4 have eigenvalues that are less than 1, i.e., explaining less variance than one original variable [23], we disregard PC3 and PC4 in the following analysis. Table 4 shows the factor loadings of PC1 and PC2. As shown, PC1 factors in all usability measures positively and more in comparison to PC2, while PC2 has a negative rating for Giving Control and Perceived Ease. This means that high PC1 score for a PM would indicate its good usability and security, whereas low score for PC2 may indicate better control and ease.

	PC1	PC2	PC3	PC4
Eigenvalue	1.885	0.877	0.686	0.553
Proportion of Variance	47.119	21.913	17.155	13.814
Cumulative Proportion	47.119	69.031	86.186	100.000

Table 3. Principle Components of Usability Measures

Table 5 depicts how each method scores with respect to PC1 and PC2. We find that a high PC1 score for USB indicates its superiority as a PM. Online is considered to have poorest overall usability due to low PC1 score and Phone has a mediocre level of usability. Figure 6(a) shows how methods form two clusters (using Agglomerative Hierarchical Clustering), one consisting of USB and Phone together and another consisting of Online. The figure indicates that our methods can be partitioned into two classes, with good and poor usability overall. Methods with good usability are USB and Phone. Online exhibits poor overall usability and security.

	PC1	PC2
Giving Control	0.619	-0.634
Perceived Ease	0.702	-0.130
Perceived Necessity	0.621	0.671
Perceived Security	0.789	0.085

Table 4. Factor Loadings of PC1 and PC2

	PC1	PC2
USB	0.612	-0.075
Phone	-0.013	-0.244
Online	-0.599	0.319

Table 5. Scores for each PM with respect to PC1 and PC2

A similar and independent analysis for Students and Non-Students indicate the following (results shown in Figures 6(b) and 6(c), resp.). For students, USB is better compared to both Phone and Online, which form a cluster together, whereby Online is better than Phone. On the other hand, for Non-Students, USB and Phone form a cluster with each other (Phone is better than USB) which compares favorably with Online.

In summary, our Post-Test analysis shows that, for all of our users, Online was surprisingly the last choice (despite its better usability as indicated via our During Test analysis in Section 5.1). Users either preferred USB or Phone. This can be credited to the fact that users felt that managing their passwords locally on their own devices gives them a sense of control and authority, as shown by their ratings for Giving Control.

5.3 Final User Preferences

After having performed the usability experiments with the PMs, we also polled for users' final preference based on their experience. We posed the subjects three questions and asked them to rank the PMs based of their order of preference:

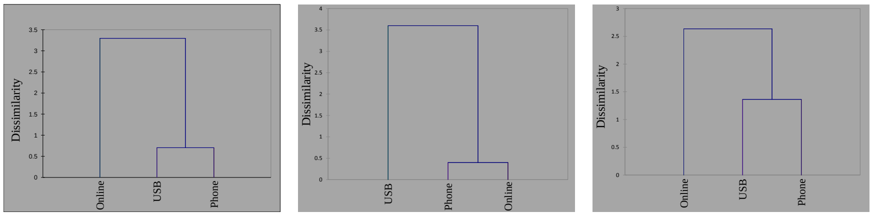
1. Which password manager do you prefer the most?
2. Which password manager according to you offers better security and protection of your passwords?
3. Which password manager according to you is most convenient to use?

The responses we received are depicted in Table 6. While a large fraction of Non-Students shows a strong liking for Phone, most Students' preference was either USB or Online (although most of them selected Phone to be most secure). In short, our overall analysis, presented in Section 5.2, conforms well with the final preferences provided by our users (i.e, All Users, Students and Non-Students).

5.4 Answers to Open-Ended Questions

Few of our test users responded to the open-ended questions that we posed. We quote below some of the interesting feedback that we received.

- **Q**: From your understanding what does Roboform2Go [USB] do?
  A: Manages, stores passwords and makes them portable. We need to know and remember only one master password, rest is taken care of by the software.



(a) All Users (b) Students (c) Non-Students

Fig. 6. Post Test Cluster Analysis Based On Principal Components (Dissimilarity vs. PM)

PM	**Prefer the Most**			**Secure**			**Convenient**		
	All	S	NS	All	S	NS	All	S	NS
USB	**40%**	**40%**	40%	25%	30%	20%	**35%**	30%	**40%**
Phone	35%	20%	**50%**	**55%**	**40%**	**70%**	30%	20%	**40%**
Online	25%	**40%**	10%	20%	30%	10%	**35%**	**50%**	20%

Table 6. % of Users Who Preferred a Particular PM (All, S, NS denote All Users, Students, Non-Students)

- **Q**: From your understanding, what does Lastpass [Online] do?
  A: Stores passwords on a central server, so makes it "real time" portable but more vulnerable towards the attacks from cyber criminals.
- **Q**: Do you have any suggestions for Lastpass? What would make it more useful or easier to use?
  A1: Lastpass server should force users to change the passwords more frequently to make it cyber-attacks proof
  A2: It is good but what if [it] does not respect our privacy and [does] not follow the code of conduct?
- **Q**: Why do you prefer this particular type of password manager [USB]?
  A: [It is] Easy to handle, portable, comparatively safe.
- **Q**: Do you think that using a password manager would make it easier to manage your passwords?
  A: Yes, but [it] is not an absolute necessity.
- **Q**: Why do you prefer this particular type of password manager [Phone]?
  A1: My Phone is the most secure of the devices and I always have it present with me wherever I go.
  A2: [It provides] Better sense of security.
- **Q**: Why do you prefer this particular type of password manager [Online]?
  A1: [It is] Easy and no need to carry anything
  A2: It is more efficient because no [additional] hardware is required.

To summarize, we find that users are aware of the importance of security of their passwords and would be inclined to use password managers. They expressed concerns regarding off-shoring their passwords to a remote entity due to security and privacy reasons and may prefer to use their own devices for managing their passwords. They, however, may not deem password managers as an absolute necessity.

5.5 Discussion and Summary

We now provide a summary of our most notable findings. Our during test analysis shows that across all users and across non-students, Online $\succ$ USB $\succ$ Phone, and USB and

Phone are clustered together. For students also, Online $\succ$ USB $\succ$ Phone, but Online and USB form a cluster of their own. This finding can be termed intuitive since online PM was expected to possess better usability than the two portable PMs. Moreover, non-students generally found the three PMs tougher and less satisfactory compared to Students. The reason for this is simple: students are much more technologically savvy compared to non-students.

Post test analysis, on the other hand, reveals surprising facts. For all of our users, the order of preference turned out to be USB $\succ$ Phone $\succ$ Online, with USB and Phone clustered together. For students, the order was USB $\succ$ Online $\succ$ Phone, whereby USB and Online form a cluster together. On the other hand, for non-students, the order of preference is Phone $\succ$ USB $\succ$ Online, and USB and Phone form a cluster. In general, we found that the portable managers are preferred over the online manager.

The above turn-around from the during test to post test can be credited to the fact that the users were not comfortable giving Control of their passwords to an online entity and preferred to manage their passwords themselves on their own portable devices. This preference reversal from during test to post test results was also confirmed by users' final preferences about the three PMs.

We also observe that the non-students had a much stronger liking for Phone compared to students while looking at overall post test data, and in terms of giving control of their passwords. Being less tech-savvy, non-students perhaps felt much more comfortable and safe while copying in their passwords (from the phone to authentication terminal) manually as opposed to letting a device (USB or remote server) doing it for them automatically.

6 Conclusions and Future Direction

We presented a comparative usability study of three notable traditional password managers. Contrary to our intuition, overall the two portable managers were preferred over the online manager, despite the better usability of the latter. Surprisingly, the online manager was the last choice for non-technical people, who mostly preferred the phone manager. Also, technical people were more inclined towards the USB manager in comparison to the online manager. These findings can generally be credited to the fact that the users were not comfortable giving control of their passwords to an online entity and preferred to manage their passwords themselves on their portable devices.

Based on our results, we can conclude that portable managers represent a more promising password management approach than online managers. The latter provide a higher degree of confidence to users in managing their passwords. However, current portable managers (especially phone managers) do not offer the usability as expected by average users, thus motivating the need for usable portable managers in the future. Owing to an ever increasing "always on, always with me" mobile phone usage trend, we believe that developing user-friendly and secure phone managers is an interesting and important research direction.

References

1. Eran Gabber, Phillip B. Gibbons, Yossi Matias, and Alain J. Mayer. How to make personalized web browsing simple, secure, and anonymous. In *Proceedings of Financial Cryptography'97*, pages 17–32, Anguilla, West Indies, February 1997.
2. A. Halderman, B. Waters, and E. Felten. A convenient method for securely managing passwords. In *Proceedings of the 2005 World Wide Web Conference*, pages 471–479, Chiba, Japan, May 2005.
3. Robert Morris and Ken Thompson. Password security: a case history. *Commun. ACM*, 22(11):594–597, 1979.
4. Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. Password memorability and security: Empirical results. *IEEE Security and Privacy*, 2(5):25–31, 2004.
5. Anne Adams and Martina Angela Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, 1999.
6. Siber Systems. Roboform password manager, 2009. Available at http://www.roboform.com.
7. LastPass. Lastpass password manager, 2009. Available at https://lastpass.com.
8. Mozilla Labs. Weave sync, 2009. Available at http://labs.mozilla.com/projects/weave.
9. Blake Ross, Collin Jackson, Nick Miyake, Dan Boneh, and John C. Mitchell. Stronger password authentication using browser extensions. In *USENIX Security Symposium*, 2005.
10. D. Reichl. Keepassmobile, 2009. Available at http://www.keepassmobile.com.
11. Openintents safe, 2009. Available at http://www.openintents.org/en/node/205.
12. Sonia Chiasson, P. C. van Oorschot, and Robert Biddle. A usability study and critique of two password managers. In *USENIX Security Symposium*, 2006.
13. R. Dhamija and L. Dusseault. The seven flaws of identity management: Usability and security challenges. In *IEEE Security and Privacy*, 2008.
14. Handypassword. `http://www.handypassword.com/login_password_manager_terms/usb_password_manager.shtml`.
15. Pc magazine: Password managers & form fillers. `http://www.pcmag.com/article2/0,2817,1791459,00.asp`.
16. Password management software review 2009. `http://password-management-software-review.toptenreviews.com/`.
17. Imation 2gb usb thumb drive: Specifications. `http://www.pcmall.com/p/Imation-Removable-Hard-Drives/product~dpno~517643~pdp.cggiicj`.
18. Nokia 5310 mobile phone: Specifications. `http://europe.nokia.com/find-products/devices/nokia-5310-xpressmusic`.
19. Browser statistics. `http://www.w3schools.com/browsers/browsers_stats.asp`.
20. Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The emperor's new security indicators. In *IEEE Symposium on Security and Privacy*, 2007.
21. J. Cohen, P. Cohen, S.G. West, and L.S. Aiken. Applied multiple regression/correlation analysis for the behavioral sciences. 1983.
22. Erik Frokjaer, Morten Hertzum, and Kasper Hornbaek. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *SIGCHI conference on Human factors in computing systems*, 2000.
23. H.F. Kaiser. The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1):141–151, 1960.