Advance Access publication on 14 November 2016 doi:10.1093/iwc/iww033

Exploring the Potential of GeoPass: A Geographic Location-Password Scheme

MAHDI NASRULLAH AL-AMEEN* AND MATTHEW WRIGHT

Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX, USA
**Corresponding author: mahdi.al-ameen@mavs.uta.edu*

Password schemes based on online map locations are an emerging topic in authentication research. GeoPass is a promising such scheme, as it provides satisfactory resilience against online guessing and showed high memorability (97%) in a single-password laboratory study. In this article, we investigate more deeply into the potential of GeoPass through four separate studies. First, in a 2-month-long field study, we found that users in a real-world setting remembered their location passwords 96.1% of the time and showed improvement with more login sessions. Then, in a study of interference effects in Geopass, in which each participant had to remember four separate location passwords, we found that memorability was <70%, with 41.5% of login failures due to interference. Based on these findings, we propose to address interference issues in GeoPass with mental stories, where users are asked to create a meaningful association between their location password and the corresponding account. We tested the efficacy of this approach through a second interference study, where the memorability rate for GeoPass was >97%, with only 3.4% of login attempts failing due to interference. We also conducted a shoulder-surfing study to examine the resilience of GeoPass against this attack. Based on our results, we identify the promising aspects of location passwords that should be further studied in future research.

RESEARCH HIGHLIGHTS

In this paper, we investigate deeply into the potential of GeoPass through four separate studies:

Field Study

- The results of our field study show the login performances of users in a real-world context, where the overall login success rate was 96.1% and the median login time of 19 seconds.
- By analyzing training effect, we found an overall improvement in login performance with more login sessions, including a 59% reduction in median login time to just 17 seconds by the 31st login session.

Interference Studies

- In the first interference study, we found that users remembered location passwords in <70% of login sessions, where 41.5% of login attempts failed due to interference effects.
- Based on our analysis in the first interference study, we hypothesized that interference effects could be reduced if participants would be asked to make a mental story during registration to create a meaningful association between their location password and the corresponding account.
- By implementing mental story approach in the second interference study, we found a 98% login success rate 1 week after registration and a 99% success rate after 2 weeks, where just 3.4% of login attempts failed because of interference effects.

Shoulder-surfing Study

- Our shoulder-surfing study on GeoPass showed a 48% overall success rate for participants playing the role of attackers.
- Our analysis found that how a user navigated to the location password, either by panning or typing the full address in the search bar, did not affect the success rate of shoulder surfing.

Keywords: usable security; authentication; field study; interference study; shoulder-surfing study

Editorial Board Member: Dr. Ian Oakley

Received 11 February 2016; Revised 10 April 2016; Editorial decision 5 October 2016; Accepted 6 October 2016

1. INTRODUCTION

Users often choose passwords that are easy to remember but also easy to guess. After decades of studies on passwords, it remains a critical challenge to address this usability–security tension in user authentication. While prior textual (Forget 2012; Shay *et al*., 2012; Wright *et al*., 2012) and graphical password schemes (Authentication 2004; Biddle *et al*., 2012; Davis *et al*., 2004; Jermyn *et al*., 1999; Wiedenbeck *et al*., 2005b) fail to fully address this problem, Thorpe *et al*., (2013) show the potential of location passwords to satisfy both security and memorability requirements. In particular, they propose GeoPass, in which a user-chosen location on an online map (e.g. Google Maps) serves as the user's password.

The history of research in user authentication schemes makes it clear that unless the primary usability and security issues of a new category of passwords are identified in the initial phases of study, the later schemes in that category might fail to address the key drawbacks of the approach. Geographic location-password schemes such as GeoPass (Thorpe *et al*., 2013) are a recent inclusion in studies of user authentication, and they show good potential in terms of single-password memorability and guessing resilience. We thus present a systematic approach to investigate deeper into the potential of location-password scheme.

Thorpe *et al.* (2013) demonstrated that GeoPass offers very good memorability (97% over a 9-day period) and high user satisfaction in a single-password laboratory study. To fully understand the usability of a scheme, however, we should study the scheme in real-world use and study how well users handle multiple passwords (Biddle *et al*., 2012). We first conducted a single-password field study that showed promise for GeoPass in a real-life scenario. We also conducted two multiple-password studies. In our first multiple-password study, we found substantial interference between the location passwords selected by users, which led to poor memorability results overall. To overcome this issue, we designed and evaluated a novel *mental story*-based approach, which contributed to gain a significant improvement in memorability for GeoPass in the second multiple-password study.

The results of security analysis show that location passwords created with GeoPass can have reasonable security against online guessing attacks, even when accounting for social engineering attacks (Thorpe *et al*., 2013). Hang *et al.* (2015) further show that GeoPass has good security, even when providing attackers direct clues about the location password, such as it is *a place where you had your first summer vacation*.

Security of a scheme goes beyond guessing attacks, however. For example, against observation attacks such as shoulder surfing, the strength of a password does not provide security guarantees, since shoulder surfers gain users' authentication secrets through direct observation. Moreover, authentication schemes with mouse input (such as GeoPass) could be highly vulnerable to this attack. Shoulder-surfing studies have been conducted on various password schemes, such as by Tari *et al*., (2006),Zakaria *et al*., (2011), von Zezschwitz *et al*. (2015b) and Bianchi *et al*., (2016).

In this article, we investigate the security of GeoPass through a shoulder-surfing study, which is the first ever experiment measuring the resilience of location passwords to such attacks. We also discuss about other security issues of GeoPass, such as phishing, malware and secure storage of location passwords at the server.

In summary, based on our findings in four separate studies, we identify the promising aspects for future research on location-based user authentication schemes.

1.1. Contributions

We now highlight the major findings of our studies with references to the corresponding sections that accommodate detailed discussions. We discuss the implications of all of our findings in Section 8.

1.1.1. Field study

A field study of a user authentication system offers better realism than laboratory studies. It provides participants with a meaningful motivation to remember their password, namely that they use it to log into a real-life account, plus it ensures that authentication is truly a secondary task, where using the system is the primary task. For example, in our study, the participants logged in using a location password to access their course study materials and their grades on examinations and assignments in a particular course.

We conducted a 66-day-long field study, including 1781 login sessions from 50 participants. Our results show the login performances of users in a real-world context, where the overall login success rate was 96.1% and the median login time of 19 seconds. We analyzed the login performance distribution among users for a detailed understanding of the usability issues (Section 4.3).

A training effect occurs when users get better at entering their password over time. Prior field studies on authentication schemes lack a detailed analysis of training effects, so it remains unknown to the research community how login performance changes during a long-term field study. We give an insight into this issue by examining the change in login performance over login sessions, where we found an overall improvement in login performance with more login sessions, including a 59% reduction in median login time to just 17 seconds by the 31st login session (Section 4.4).

Our field study does not include a control condition, and thus, we do not directly compare the login performance of GeoPass scheme with that of the traditional textual password. As reported in a comprehensive survey on 25 different password schemes (Biddle *et al*., 2012), none of the field studies conducted on these schemes included a control condition.

However, we note that given enough participants, having a control condition in the future study would let us conduct a direction comparison between two schemes.

1.1.2. *Multiple-password interference studies*

Password interference (Biddle *et al.*, 2012; Chiasson *et al.*, 2009b) occurs when users confuse the password of one account with that of another account. Each of our two interference studies was conducted over the span of 3 weeks. We used a separate group of participants for each study to prevent training effects from carrying over. Each participant had to remember four different location passwords, one for each of four different accounts. Although in real life, users have to remember >4 passwords, to keep consistency with prior multiple-password studies that asked users to remember either three or four passwords (Everitt *et al.*, 2009; Hlywa *et al.*, 2011; Wright *et al.*, 2012), the participants in our interference studies were asked to remember four location passwords.

Interference Study I: In the first interference study, we found that users remembered location passwords in <70% of login sessions (Section 5.2), where 41.5% of login attempts failed due to interference effects. For deeper understanding, we investigated both *accurate* and *non-accurate* interference (Section 5.3). In doing so, we found that the interference effect between a pair of location passwords had no correlation with the geographic distance between them (Section 5.4). In other words, participants were not confused by location passwords that were geographically near each other. Rather, they failed to associate their location passwords with the corresponding accounts and thus could not log in successfully.

Interference Study II: Based on our analysis in *Interference Study I*, we hypothesized that interference effects could be reduced if participants would be asked to make a *mental story* during registration to create a meaningful association between their location password and the corresponding account (Section 5.4). For example, in *Interference Study II* that examines the efficacy of this approach, one participant chose the Bellagio in Las Vegas as her location password for her *bank* account, and her story was: 'Bank→Money→Las Vegas→Bellagio'. In this way, participants could better memorize the location password for a particular account. In *Interference Study II*, we found a 98% login success rate 1 week after registration and a 99% success rate after 2 weeks (Section 6.3), where just 3.4% of login attempts failed because of interference effects (Section 6.4).

1.1.3. *Shoulder-surfing study*

In a shoulder-surfing attack (Florencio *et al.*, 2007; Tari *et al.*, 2006), as users enter login information, an attacker may gain knowledge about users' credentials by direct observation. Tari *et al.*, (2006) showed that neither traditional textual passwords nor commercially available graphical passwords (e.g. Passfaces (Authentication 2004)) provide sufficient resilience against shoulder surfing. In our shoulder-surfing study (Section 7.2), participants playing the role of attackers could login 48% of the time to the 'victim' account. Such a high rate of success shows that more attention is required from the research community to improve the resilience of location passwords against this attack. Our analysis found that how a user navigated to the location password, either by panning or typing the full address in the search bar, did not affect the success rate of shoulder surfing.

The rest of our article is organized as follows: we give an overview of the notable textual and graphical authentication schemes in Section 2, followed by a discussion on geographic location passwords in Section 3. The results of our field study are reported in Section 4. In Sections 5 and 6, we describe the findings from our first and second interference studies, respectively. We report the results of our shoulder-surfing study in Section 7. The implications of our findings and the potential of future research are discussed in Section 8, followed by a conclusion on our work in Section 9.

2. RELATED WORK

In this section, we give a brief overview of notable textual and graphical password schemes, in which we highlight why existing schemes are insufficient. We then discuss geographic location passwords and their potential.

2.1. Textual password schemes

Traditional user-chosen textual passwords are fraught with security problems and are especially prone to password reuse and predictable patterns (Das *et al.*, 2014; Campbell *et al.*, 2011; Shay *et al.*, 2010). Ur *et al.*, (2015) showed that misconceptions of users contribute to creating weak passwords. For example, many users believe that adding a special character at the end of a password makes it secure (Ur *et al.*, 2015). Their study also showed that users could anticipate only the targeted guessing attack, believing that it is a secure approach to use a birthday or name as a password if those data are not available on social networking sites. More recently, Ur *et al.*, (2016) showed that users have serious misconceptions about the impact of basing passwords on common phrases and including digits and keyboard patterns in passwords, which lead them to creating weak and predictable authentication secrets.

Different password restriction policies have been deployed to get users to create stronger passwords (Campbell *et al.*, 2011; Shay *et al.*, 2010, 2014). These studies report, however, that such policies do not necessarily lead to more secure passwords, but they do adversely affect memorability in several cases. In another study, Shay *et al.* (2015) found that a multi-step password-creation process that provides guidance to users is not effective enough in creating strong passwords.

While user-chosen textual passwords fail to provide adequate security, Bonneau *et al.*, (2012) suggested a set of usability and

security metrics that are required to be addressed in order to provide a viable solution to the usability–security tension in online user authentication. In their metrics, system-assigned random password schemes are more secure than the user-chosen passwords, but they fail to provide sufficient memorability, even when natural language words are used (Shay *et al*., 2012; Wright *et al*., 2012). Forget *et al*., (2008) and Forget (2012) proposed the Persuasive Text Passwords scheme as a hybrid between user-selected and system-assigned passwords, but the memorability can be poor—as low as 25%.

2.2. Graphical password schemes

Graphical password schemes can be divided into three categories (Biddle *et al*., 2012), based on the kind of memory leveraged by the systems: (i) drawmetric (recall-based), (ii) locimetric (cued-recall-based) and (iii) cognometric (recognition-based).

2.2.1. *Drawmetric*

The user is asked to reproduce a drawing in this category of graphical passwords. In *Draw-a-Secret* (*DAS*) (Jermyn *et al*., 1999), a user draws on top of a grid, and the password is represented as the sequence of grid squares. Nali and Thorpe (2004) showed that users choose predictable patterns in DAS that include drawing symmetric images with 1–3 pen strokes, using grid cell corners and lines (presumably as points of reference) and placing their drawing approximately in the center of the grid.

BDAS (Dunphy and Yan 2007) intends to reduce the amount of symmetry in the user's drawing by adding background images, but this may introduce other predictable behaviors such as targeting similar areas of the images or image-specific patterns (Biddle *et al*., 2012). DAS and BDAS have recall rates of no higher than 80%.

2.2.2. *Locimetric*

The password schemes in this category, including *Passpoints* (Chiasson *et al*., 2007; Wiedenbeck *et al*., 2005a) and *Cued Click-Points* (Chiasson *et al*., 2007), present users with an image and have users select points on the image as their password. Dirik *et al*., (2007) developed a model that can predict 70–80% of the user's click positions in Passpoints. To address this issue, Chiasson *et al*. (2008, 2012) proposed *Persuasive Cued Click-Points* (*PCCP*), in which a randomly positioned viewport is shown on top of the image during password creation, and users select their click-point within this viewport. The memorability for PCCP was found to be 83–94%. In a follow-up study, Chiasson *et al*. (2009a) found predictability in users' click points and indicate that the hotspot issue is still a security concern for PCCP.

2.2.3. *Cognometric*

In this recognition-based category of graphical passwords, the user is asked to recognize and identify their password images from a set of distractor images. *Passfaces* (Authentication 2004; Valentine 1998) is a commercial cognometric system in which users select one face among a panel of nine distractor faces and repeat this over several panels. Davis *et al*., (2004) have found that users select predictable passwords on faces (their own version of Passfaces), biased by race, gender and attractiveness of faces. As a result, the commercial Passfaces (Authentication 2004) product now assigns a random set of faces instead of allowing users to choose. Everitt *et al*., (2009) show that, unfortunately, users have difficulty in remembering system-assigned Passfaces.

In the graphical password scheme proposed by Davis *et al*., (2004), users were asked to build a story to remember a set of images in correct order. The study (Davis *et al*., 2004) found a 85% login success rate over the span of 1 week. In another study, Al-Ameen *et al*. (2015) proposed a cognometric scheme called CuedR where they combined different types of memory cues to help users memorize system-assigned passwords. In a follow-up study (Al-Ameen *et al*., 2015a), the authors examined the individual impact of different types of cues and user interaction on the memorability of system-assigned recognition-based graphical passwords and found that the combination of verbal and spatial cues for recognizing object images performed best among seven different schemes. In a third study (Al-Ameen *et al*., 2015b), the authors compared textual and graphical recognition-based schemes offering memory cues, and found that adding images to textual information contributed to gain a significant improvement in usability.

Recently, Bianchi *et al*., (2016) have proposed PassBYOP, a novel graphical password scheme for public terminals, designed to improve the security of graphical password systems against observation attacks such as shoulder surfing by integrating live video of a physical token that a user carries with them. So unlike typical graphical password schemes that use static digital images, PassBYOP uses personalized physical tokens based on digital pictures displayed on a user-owned device such as a smartphone. Users are required to present these images to a system camera and then enter their password as a sequence of selections on live video of the token, while highly distinctive optical features are extracted from these selections and used as the password. The study (Bianchi *et al*., 2016) found a 9% error rate with this scheme during task completion by the participants.

2.3. Multiple-password studies

Biddle *et al*. (2012) identified multiple-password interference as a major usability concern and found in their extensive survey that only a handful of graphical password schemes have been evaluated with an interference study. To the best of our knowledge, no interference study has been conducted yet on geographic location passwords.

In a multiple-password study on locimetric graphical passwords (e.g. Passpoints) (Chiasson *et al*., 2009b), the authors

reported that 57% (15/26) of the participants were able to recall their graphical passwords successfully after 2 weeks of registration, and the average login time varied between 18 and 47 seconds.

Everitt *et al.* (2009) performed a multiple-password study on a cognometric graphical password scheme similar to Passfaces (Authentication 2004). The study (Everitt *et al.*, 2009) demonstrated that participants using four different passwords each week had a failure rate of 15.23% after a month, when each password was used once a week.

2.4. Shoulder-surfing attack

Shoulder-surfing attacks and defenses have been studied for both textual and graphical passwords, though not for geographic location-password schemes. We now briefly discuss this prior work.

In the shoulder-surfing study conducted by Tari *et al.*, (2006), the university students played the role of attackers and the experimenter played the role of a victim. The study (Tari *et al.*, 2006) found that neither traditional textual passwords nor graphical passwords (with mouse input) provide sufficient resilience against shoulder surfing.

De Luca *et al.*, (2013) examined the impact of fake cursors in providing resilience against shoulder surfing when passwords are entered through an on-screen keyboard. They found, however, that users act predictably to identify their active cursor (e.g. moving the mouse cursor to the border of the interaction area or moving the mouse in small circles), which may make it easy for the shoulder surfer to find the real cursor and subsequently the authentication secret from on-screen keyboard entries. von Zezschwitz *et al.* (2015a) proposed a touch gesture-based mechanism to input the authentication secret gaining higher resilience against shoulder-surfing attack.

Shoulder-surfing attacks pose a particularly serious security threat for graphical passwords (Biddle *et al.*, 2012). A few techniques have been proposed to gain resilience against shoulder surfing, such as eye-gaze entry used for locimetric passwords (Forget *et al.*, 2010) and, in drawmetric passwords, 'snaking' away the strokes of the drawn password while they are being drawn (Zakaria *et al.*, 2011). These techniques suffer from usability problems (Forget *et al.*, 2010; Zakaria *et al.*, 2011). von Zezschwitz *et al.* (2015b) conducted a systematic evaluation on the shoulder-surfing susceptibility of drawmetric passwords, and found that line visibility and pattern length are the most important parameters to ensure observation resistance.

Al-Ameen *et al.* (2015) implemented *variant response* feature to gain resilience against shoulder-surfing attack for recognition-based graphical passwords, where users are required to enter a small-case letter for selecting an image and this letter changes between login sessions. An attacker needs to look at the keyboard and monitor simultaneously to gain the user's authentication secret, which has been found to be difficult in practice (Al-Ameen *et al.*, 2015; Tari *et al.*, 2006). A shoulder-surfing study on the PassBYOP scheme showed that all of the three participants were able to gain users' authentication secrets when public images are used, but the resilience to this attack was improved by using private images (Bianchi *et al.*, 2016).

3. GEOGRAPHIC LOCATION PASSWORDS

Geographic location-password schemes represent a recent category in password research. In these schemes, users select one or more locations in an online map (e.g. Google Maps) as their password. To the best of our knowledge, four schemes (Hang *et al.*, 2015; Spitzer *et al.*, 2010; Sun *et al.*, 2012; Thorpe *et al.*, 2013) in this password category have been proposed to date, where GeoPass, proposed by Thorpe *et al.*, (2013), shows most potential in terms of usability and security.

In GeoPass, the user's password is a single location on an online map (Google Maps). This secret location, known both as the *location password* and just *geopass*,[1] is selected by the user at registration by right-clicking on the map. The search bar helps to make navigation faster by enabling the user to type the name of a place. Also, typing leads to a drop-down menu suggesting locations in which the searched item may appear. Zooming and panning are also enabled via the Google Maps Application Programming Interface. Using the convention that a higher numbered zoom level represents being zoomed in closer, the initial zoom level is 2, and GeoPass allows the user to click on a location at a minimum zoom level of 16. A successful login requires the users to click within a 21×21 pixel box around the location password they had set, while the error tolerance is calculated at zoom level 16. We refer readers to Thorpe *et al.*'s (2013) paper for in-depth discussion on the features of GeoPass.

Thorpe *et al.*, (2013) performed a security analysis of GeoPass with three different types of threat models, where the most effective attack was produced when the adversary has local knowledge of users. Even though this attack model guessed 11% of location passwords, it required $2^{16.36}$ guessing attempts. The authors compared this worst-case scenario attack on location passwords to recent attacks against text passwords and found that the attackers might require up to 82 times more guessing attempts in GeoPass as compared to text passwords for gaining this 11% guessing success rate.

The security of a password scheme goes beyond guessing attacks. For example, shoulder-surfing attacks are considered a security threat for many authentication schemes, in which the attackers gain users' authentication secrets through direct observation (Bianchi *et al.*, 2016; Tari *et al.*, 2006; von Zezschwitz *et al.*, 2015b; Zakaria *et al.*, 2011). Against these attacks, the guessing strength of the passwords does not provide any security guarantee. Moreover, authentication

[1] In lowercase to avoid confusion with the system name.

schemes with mouse input, including GeoPass, could be highly vulnerable to shoulder surfing. In this work, we thus examine this aspect of GeoPass's security through a shoulder-surfing study. In addition, we discuss other security issues of GeoPass, such as phishing, malware and secure storage of location passwords at the server.

Thorpe *et al*., (2013) conducted a 9-day-long user study on GeoPass with three sessions: two in a laboratory setting and one online. The login success rate was 97%, and the median login time was found to be no >30 seconds (Thorpe *et al*., 2013). The user feedback on GeoPass was encouraging, and a majority of users showed interest to use the scheme in real life (Thorpe *et al*., 2013). Their study on GeoPass (Thorpe *et al*., 2013) did not make a direct comparison with traditional textual passwords. What we know about such passwords, however, shows even more promise in GeoPass. For example, Shay *et al*., (2014) measured the memorability of user-chosen textual passwords following different composition policies, where each user was required to remember a single password. The study found that after 2 days of registration, the login success rate was <87% (within five attempts) for all of the studied composition rules. In contrast, the laboratory study on GeoPass (Thorpe *et al*., 2013) showed 97% login success rate.[2] We thus argue that further study of location-password schemes, and GeoPass in particular, is warranted.

A comprehensive survey on graphical password schemes demonstrates that an authentication scheme showing promising usability in a preliminary laboratory study should be further studied to understand its usability in real-world use and identify how well users handle multiple passwords (Biddle *et al*., 2012). We thus conducted a field study and multiple-password interference studies for a deeper understanding on the usability of GeoPass scheme.

Other schemes: There are three other schemes that use map locations as an authentication secret: one called PassMap (Sun *et al*., 2012) and another one proposed by Spitzer *et al*., (2010). PassMap requires the user to choose two locations as the location password, while Spitzer's scheme requires the user to select five or seven locations at different zoom levels. Thorpe *et al*., (2013) have shown that GeoPass is more usable than other digital-map-based schemes (Spitzer *et al*., 2010; Sun *et al*., 2012) because of its requirement to click on a single location and normalized error tolerance to a given zoom level. The login success rate in GeoPass (97%) was found to be higher than the rate in PassMap (92.59%). Recently, Hang *et al*., (2015) designed a location-password-based secondary authentication scheme, where users are required to choose a place from Google Map as an answer to a location-based cognitive questions. Their scheme offered 84% memorability a month after registration (Hang *et al*., 2015).

[2]We note that direct comparison between different studies should be taken with caution.

4. FIELD STUDY

In the preliminary laboratory study on GeoPas, Thorpe *et al*., (2013) found satisfactory memorability for this scheme. We investigate deeper into the usability of GeoPass through a field study, which offers a good measure of login performance in a realistic setting. In particular, field studies help ensure that logging in becomes a secondary task for users as they are primarily seeking to use the system, plus it provides a realistic motivation to memorize the authentication secret.

In this section, we describe the procedure and results of our field study, where we address the following research questions:

- How usable would GeoPass be in a real-world setting in terms of login performance of users?
- How do the login performances in GeoPass change over login sessions (i.e. what is the training effect)?

4.1. Study design

We conducted the field study on a computer science class with both undergraduate and graduate students. At the beginning of study, the students were informed that we developed a website to let them access course study materials and their grades on examinations and assignments.[3] At this point, students were not told that they would be using the scheme as a part of the field study, since it could make them conscious about their login performance (Biddle *et al*., 2012; Chiasson *et al*., 2007). To note, this deception in our study was approved by IRB.

Upon collecting the students' names from the instructor, a username (`firstname.lastname`) was assigned to each user. With a projector, the students were shown the registration and login procedure with GeoPass scheme. Then, the students were asked to create a geopass for their accounts. To protect against unauthorized access, the usernames of the students were pre-stored in the system so that only students in this class could create accounts, one per username. Table 1 shows the results for registration, where maximum, minimum, median and standard deviation are cited as Max, Min, Med and SD, respectively. The median registration time was ~2.6 minutes.

The GeoPass system was active for 66 days. At the end of semester, we informed the students about our field study. Out of 57 students in this class, 50 students (10 women and 40 men with a mean age of 24) gave positive consent to use their login information for the study and signed consent forms before participating in an anonymous paper-based survey. They were compensated with extra credit in a class assignment for participating in this survey, and an alternative assignment was offered for those who did not want to participate. The participants had not taken any courses on usable security or human–computer

[3]Grades were posted in a file containing all students' grades and anonymized by replacing names with a code given to each student.

Table 1. *Field Study:* registration time (seconds).

Mean	Med	SD	Max	Min
203	154	162	593	25

interaction, nor had they participated in a password-related user study. During registration, the participants were discouraged from writing down their location passwords.

Our system recorded 1781 login sessions performed by these 50 users during the study. The users could log in at anytime from anywhere using their desktop or laptop computers. Fifty-six percent of users reported that in most cases, they logged in from a computer with a 15-inch screen, while the screen sizes of computers for other users varied from 12 to 27 inches. Thus, most of our results indicate the performance of users on typical computer screens.

During authentication, we started counting login time when the Google map interface was shown to the user after entering her username. A successful attempt required the user to enter both her username and geopass correctly. An unsuccessful attempt refers only to sessions where the username was correct but the geopass was selected incorrectly.

To reset a geopass, the participant had to send an email to the experimenter from her *.edu* email account, and in response she would receive a link through email to create a new geopass. Five participants had to reset their geopasses within the first few days of the study because of a technical problem in our system. The results reported in this article do not include any login sessions for these users performed prior to these resets, since the corrupted data had to be deleted while fixing that technical issue. Thereafter, no participant reset her geopass during the study.

4.2. Significance tests

To analyze our results, we use statistical tests and consider results comparing two conditions to be significantly different when we find $P < 0.05$, unless otherwise specified. We implement Bonferroni correction wherever appropriate. We selected statistical tests based on their appropriateness for the corresponding data sets. In this brief section, we give an overview of the statistical tests that we use in our analysis.

When comparing two conditions where the variable is at least ordinal, we use a Wilcoxon signed-rank test for the matched pairs of subjects and a Wilcoxon–Mann–Whitney test for unpaired results. Wilcoxon tests are similar to *t*-tests, but make no assumption about the distributions of the compared samples, which is appropriate to the data sets in our conditions. Whether or not a participant successfully authenticated is a binary measure. So, we use either a McNemar's test (for matched pairs of subjects) or a chi-squared test (for unpaired results) to compare login success rates between two conditions (Fig. 1).

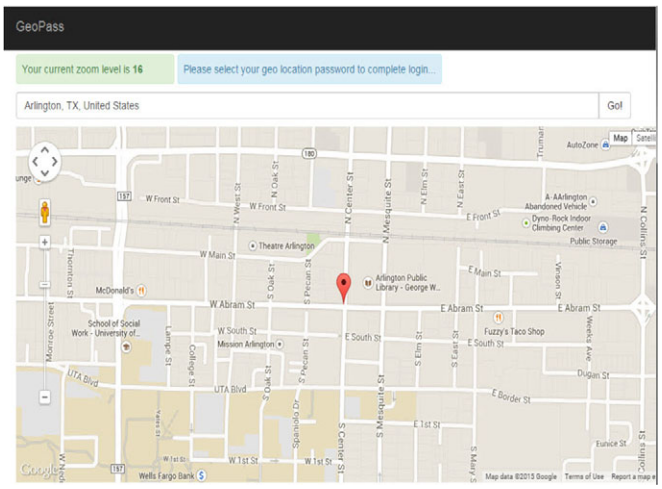

Figure 1. A partial screen shot of the GeoPass scheme.

4.3. Overall login performance

In our field study, we recorded 1781 login sessions, where a single login session (or *login*) by a participant may include multiple attempts to authenticate successfully. A login is marked as unsuccessful when the participant leaves the authentication webpage after failing to click on the correct location. To find the full distribution of the number of attempts needed for a successful login, we did not limit the number of attempts a participant can make during a login session. One attempt refers to right-clicking at a location on the Google map.

Participants performed 35.6 logins on average (minimum: 8, maximum: 130). Figure 2 shows the number of logins by the participants, where a (x,y) point represents the percentage of participants (y%) who conducted at least x logins, either successful or unsuccessful.

We measured the average login performance of each participant in her login sessions (see Figs 3–5) and calculated the overall login performances for all of the participants over 1781 login sessions (see Table 2). The overall login success rate was 96.1%. Users required 1.3 attempts (on average) per successful login, while the average login time was 32 seconds, with a median of 19 seconds. The login success rate is similar to the results from the single-laboratory study of Thorpe *et al.*, (2013) (97%) and shows that GeoPass has good memorability in a real-world setting for a single geopass.

To illustrate the login performance in more detail, Figures 3–5 show empirical cumulative distribution functions (ECDFs) of login performance statistics taken over the users in our study (in Fig. 3, the x-axis is shown with increasing success rates and thus appears reversed).

Figure 3 shows login success rates among participants. GeoPass proved sufficiently memorable for nearly all our participants. Forty-eight percent of participants had a 100% login success rate, 82% had at least a 90% success rate and 96%

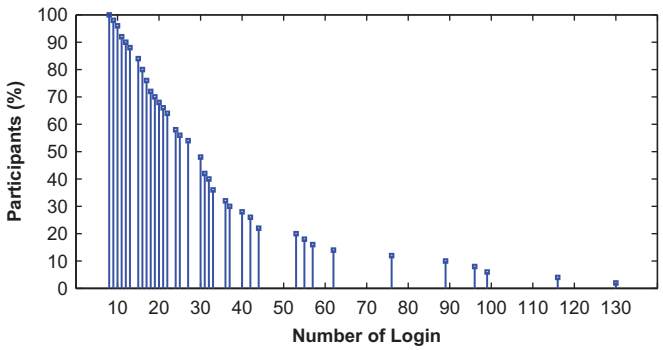

Figure 2. *Field Study:* number of logins by the participants, where a (x,y) point represents the percentage of participants (y%) who conducted at least x logins, either successful or unsuccessful.

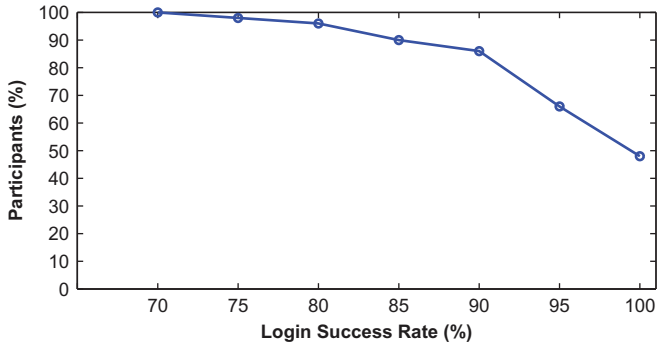

Figure 3. *Field Study*: login success rate (50 participants).

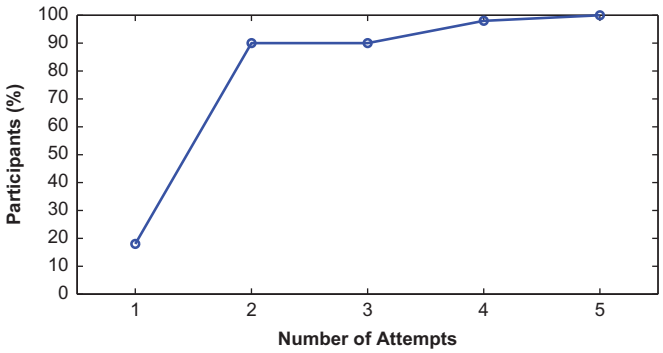

Figure 4. *Field Study:* mean number of attempts.

had at least an 80% success rate. The minimum success rate for any participant was 70%.

Figure 4 shows the average number of attempts per successful login among the participants. Ninety percent of participants made at most two attempts on average to authenticate successfully, and 100% of participants logged in successfully within five attempts on average.

The performance for login time was more mixed. Figure 5 shows the average login time among participants. The average login time was 20 seconds or less for 18% of participants and

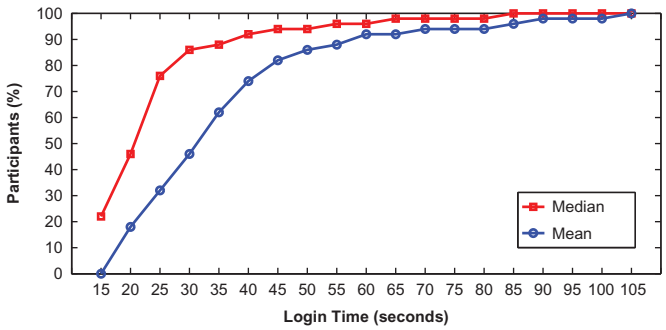

Figure 5. *Field Study*: login time.

Table 2. *Field Study*: overall login performance (login sessions: 1781; login success rate: 96.1%).

	Mean	Med	SD	Max	Min
No. of attempts	1.3	1	1.3	22	1
Login time (seconds)	32	19	59	997	5

30 seconds or less for 46% of participants. The median login time was 15 seconds or less for 22% of participants and 30 seconds or less for 76% of participants. Thus, 24% of participants had median login times >30 seconds, so a substantial fraction of users faced substantial delays in getting to their information.

4.4. Training effects

Users performing any task typically get better with practice, a phenomenon known as the *training effect*, and this may be particularly true for recalling a piece of information. Compared to first-time users, a user logging in with the same location password over time may both improve her skill in finding and clicking on the location and her memory of the location. Since the true performance of a scheme matters as summed up over the lifetime of the password, not for a single login, quantifying these improvements is very important to measuring the usability of the scheme. A short-term laboratory study cannot measure the benefits of practice in users' login performance. Thus, it remains unknown to the research community how login performances change over login sessions due to training effect. The prior field studies on graphical passwords did not present a detailed analysis of the training effect. To determine the extent of any training effects for GeoPass users in a real-world setting, we analyzed the change in login performance over login sessions. The results are shown in Figures 6–8. We illustrate the results at every sixth session, i.e. at nth login session ($n = 1, 7, 13, 19, 25, 31$). Here, we consider up to the 31st login session, since

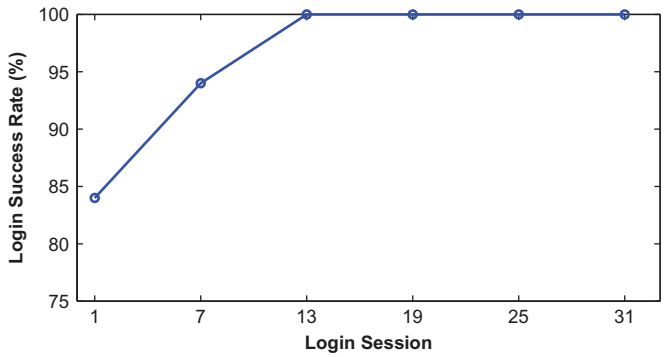

Figure 6. *Field Study*: the change in login success rate over the sessions (Table 3 shows the number of participants in each login session).

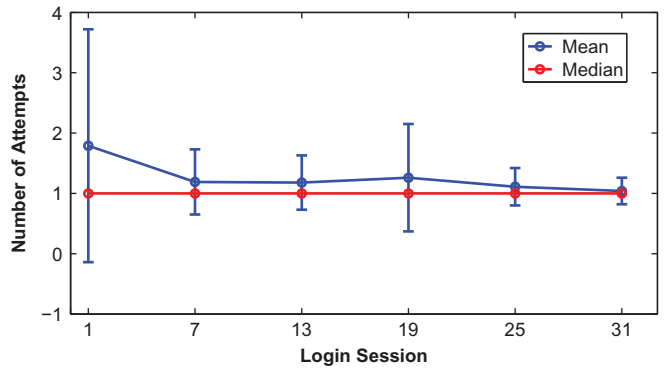

Figure 7. *Field Study*: the change in number of attempts over the sessions (Table 3 shows the number of participants in each login session).

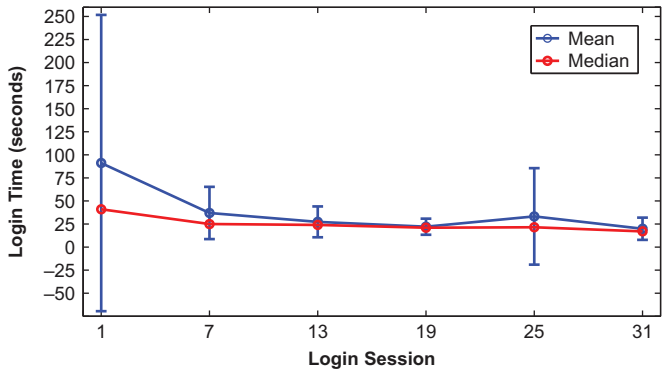

Figure 8. *Field Study*: the change in login time over the sessions (Table 3 shows the number of participants in each login session).

using the next value of n (37) would make for a rather small sample size (15 users).

We note that the sample size changes (shrinks) for each successive value of $n > 7$. As we are looking for a training effect, we may be concerned about the remaining population of users being more adept at using the system than those who have stopped logging in. Our results, however, show that the number of login sessions performed by a participant did not have a strong correlation with her login success rate ($r = 0.18$), number of attempts for successful login ($r = -0.11$) or login time ($r = -0.21$). As the strongest training effects that we found occurred by session $n = 7$, this is not a surprising result.

Table 3. *Field Study:* number of participants in the nth login session.

n	1	7	13	19	25	31
Participants	50	50	44	35	28	21

We note that the nth login session of any given participant likely occurred at a different time than that of other participants. The number of participants varied for different values of n (login session), since the participants performed different numbers of logins (see Fig. 2). Table 3 represents the number of participants in each of the nth login sessions.

4.4.1. *Login success rate*

The login success rate was 84% in the first login session, 94% in the 7th login session and 100% in the 13th login session (see Fig. 6).[4] The login success rate remained at 100% in the 19th, 25th and 31st login sessions. Thus, we see an apparent training effect in which users learn their geopasses better after a modest number of sessions.

We did not find significant difference in login success rate between any pair (m, n) of login sessions $(m, n = 1, 7, 13, 19, 25, 31)$. To note, we implement Bonferroni correction and consider results comparing two conditions to be significantly different when we find $P < 0.0033$.

4.4.2. *Number of attempts*

The mean number of attempts for a successful login was 1.79 in the first login session, which decreased to 1.19 in the seventh login session. Except for the 19th login session, the mean number of attempts for successful logins decreased over the login sessions as shown in Figure 7. Again, we see that participants more completely learned their geopasses after six sessions and made fewer mistakes. As shown in Figure 7, the standard deviation in the number of attempts for successful login was higher in the first login session as compared to other sessions.

For significance tests in this section, we used Wilcoxon signed-rank tests for matched pairs of subjects and Wilcoxon–Mann–Whitney tests for unpaired results. As we implement Bonferroni correction, we consider results comparing two conditions to be significantly different when we find $P < 0.0033$. The analysis show that the number of attempts for successful login in the 13th ($W = 795.5$, $P < 0.0033$), 19th ($W = 628$,

[4]A given (x,y) point in Figures 6–8 represents the average login performance (y) of the participants calculated over the xth login session of each individual.

Table 4. *Field Study*: significance tests for number of attempts for successful login between pairs of login sessions (Wilcoxon signed-rank tests for matched pairs of subjects and Wilcoxon–Mann–Whitney tests for unpaired results).

Login sessions	7	13	19	25	31
1	$V = 68.5$, $P = 0.01$	$\boldsymbol{W = 795.5}$, $\boldsymbol{P < 0.0033}$	$\boldsymbol{W = 628}$, $\boldsymbol{P < 0.0033}$	$\boldsymbol{W = 475.5}$, $\boldsymbol{P < 0.0033}$	$\boldsymbol{W = 333.5}$, $\boldsymbol{P < 0.0033}$
7		$W = 1059.5$, $P = 0.32$	$W = 834$, $P = 0.29$	$W = 640$, $P = 0.16$	$W = 452.5$, $P = 0.07$
13			$W = 759$, $P = 0.43$	$W = 582.5$, $P = 0.26$	$W = 410$, $P = 0.10$
19				$W = 471$ $P = 0.32$	$W = 332$ $P = 0.13$
25					$W = 276.5$ $P = 0.23$

As we implement Bonferroni correction, we consider results comparing two conditions to be significantly different when we find $P < 0.0033$.

$P < 0.0033$), 25th ($W = 475.5$, $P < 0.0033$) and 31st ($W = 333.5$, $P < 0.0033$) login sessions were significantly less than that in the first login session. As shown in Table 4, we did not find significant difference in number of attempts between any other pair (m, n) of login sessions $(m, n = 1, 7, 13, 19, 25, 31)$.

4.4.3. *Login time*

The mean login time was 91 seconds in the first login session, which decreased to 37 seconds in the seventh login session. As shown in Figure 8, mean and median login times decreased over the login sessions, where we find an exception at the 25th login session. Figure 8 shows that the standard deviation in login time in the first session was higher than the other login sessions. Some participants seem to have much more difficulty learning the new system or recalling their geopass the first time. Here, the training effect results in users getting faster at finding their geopass location in the map interface with fewer mistakes. The results of significance tests for login time between pairs of login sessions are shown in Table 5, which suggest that although most of the performance improvement occurs in the first few sessions, users continue to get moderately faster at logging in even after 13 sessions.

4.5. Summary

Our 66-day-long field study including 1781 login sessions from 50 participants provides a deeper understanding on the usability of GeoPass in a real-life setting and the change in login performance with more login sessions. The results showed promise for GeoPass scheme in a real-world scenario, where the overall login success rate was 96.1% and the median login time was 19 seconds. We also gave an insight into the training effect on GeoPass, where we found an overall improvement in login performance with more login sessions with some modest fluctuations. While Thorpe *et al.*, (2013) reported their concern about the login time of GeoPass, we found that the median login time dropped to 17 seconds by the 31st login session, a 59% reduction overall as compared to the median login time in the first session (41 seconds).

In our field study, participants were required to remember a single location password. In the next section, we report our results for the interference study on GeoPass, where we asked participants to remember multiple location passwords.

5. INTERFERENCE STUDY I

Because of password interference (Biddle *et al.*, 2012; Chiasson *et al.*, 2009b), users confuse the password of one account with that of another account while remembering multiple passwords. So, it is important to conduct interference study in order to investigate deeper into the usability of an authentication scheme (Biddle *et al.*, 2012; Chiasson *et al.*, 2009b).

Table 5. *Field Study*: significance tests for login time between pairs of login sessions (Wilcoxon signed-rank tests for matched pairs of subjects and Wilcoxon–Mann–Whitney tests for unpaired results).

Login sessions	7	13	19	25	31
1	$\boldsymbol{V = 266, P < 0.0033}$	$\boldsymbol{W = 430, P < 0.0033}$	$\boldsymbol{W = 245, P < 0.0033}$	$\boldsymbol{W = 228, P < 0.0033}$	$\boldsymbol{W = 114, P < 0.0033}$
7		$W = 867.5, P = 0.05$	$W = 592, P$=0.006	$W = 473.5, P = 0.009$	$\boldsymbol{W = 267.5, P < 0.0033}$
13			$W = 681.5, P = 0.19$	$W = 520.5, P = 0.14$	$\boldsymbol{W = 319.5, P < 0.0033}$
19				$W = 456, P = 0.32$	$W = 271.5, P = 0.05$
25					$W = 244, P = 0.16$

As we implement Bonferroni correction, we consider results comparing two conditions to be significantly different when we find $P < 0.0033$.

We conducted two separate interference studies on GeoPass, where we address the following research questions:

(i) How usable would GeoPass be when users would have to remember multiple location passwords?
(ii) How prominent will the interference effects be for multiple location passwords?
(iii) If an interference effect is found in GeoPass, why does interference occur in this scheme and how could we reduce interference effects and improve the memorability for multiple location passwords?

In this section, we report the results of our first interference study, where we explore the login performances and the corresponding interference effects when users have to remember multiple location passwords in GeoPass (Thorpe *et al*., 2013). Based on our findings in this study, we propose a solution to the interference effect and examine its efficacy through the second interference study. We report the results of our second interference study in Section 6.

5.1. Study design

We conducted the study in a course on computing basics intended for students from a broad range of majors. Out of 60 students in this class, 18 students (11 women and 7 men) participated in our study. Their mean age was 23. The subjects were compensated with extra credit in a class assignment for participating in this study, and an alternative assignment was offered for those who did not want to participate.

Haque *et al*., (2013) classify websites into four categories: (i) financial (e.g. WellsFargo.com), (ii) identity (e.g. Gmail), (iii) content (e.g. Netflix, Weather.com) and (iv) sketchy (unfamiliar sites offering coupons and often attracting transient user relationships). We built one website from each of the above categories and refer to them in this article as *bank*, *email*, *movie* and *deals*, respectively. Each site was equipped with GeoPass for user authentication. The sites were designed to have the images and layouts from familiar commercial sites, with the exception of the deals site, which was designed to look less professional.

5.1.1. Procedure

The 3-week-long laboratory-based interference study included three sessions, which we will call *sittings* to distinguish from login sessions. The sittings were held 1 week apart. In the first sitting (*registration*), we gave the participants an overview of GeoPass and asked them to create a location password for each account. To best study interference effects, the participants were asked to create a distinct geopass for each account. In the second and third sittings, users were asked to log into their four accounts from the laboratory computers. We refer to these latter sittings as *Login 1* and *Login 2*, respectively. The participants could log into the sites in any order.

If a participant failed to log into an account after five attempts, she was shown a button that she could use to view her geopass. She was also allowed to make more attempts without viewing her location password. Once the button was clicked to view the geopass, however, the participant was no longer able to attempt to log into that account for that sitting. During registration, the participants were asked not to write down their location passwords.

We note that we found no statistically significant differences between *Login 1* and *Login 2* for number of login attempts, login time or interference effects.

5.2. Login performance

Each of the 18 participants logged into four accounts in both *Login 1* and *Login 2*, making a total of 72 login sessions in each sitting. The results for login performances are shown in Table 6. In the rest of this article, *number of attempts* and *login time*, respectively, refer to the required attempts and time for successful logins only, unless otherwise specified.

The results for login performances are shown in Table 6. The overall login success rates were 58% in *Login 1* and 67% in *Login 2*. These rates are likely unacceptable for any real-world application. The mean number of attempts for successful logins were 2.8 in *Login 1* and 3.1 in *Login 2*, while the median was 1 in both sittings. The mean times for successful logins were 59 seconds in *Login 1* and 45 seconds in *Login 2*, while the medians were 43 seconds in *Login 1* and 35 seconds in *Login 2*.

5.3. Interference effect

We now explain how we measure the interference effect and describe the corresponding results. In each sitting, every participant was asked to complete four login sessions, each for one account. We refer to the account corresponding to current login session as the *visible account* and refer to the other three accounts as *invisible accounts*. For example, when a participant attempts to log into the bank account, the bank account is visible, while email, movie and deals accounts are considered invisible. Thus, a successful login requires the user to select the geopass of the visible account. Because of interference effects, a user may mistakenly click on the geopass of an invisible account. Table 7 shows the summarized results for interference effects in our study.

5.3.1. *Method of computation*

We did not restrict the number of attempts a participant could make for a successful login, and clicking at a location other than the geopass of the visible account results in an unsuccessful attempt. We figure out the impact of interference on the failure of an attempt in the following way: for each unsuccessful attempt, we measure the distances (in kilometers) between the clicked location and her geopasses for each of the four accounts. In this way, we find the account whose geopass is closest to the clicked location. If the closest account is the visible account, we assume that interference did not impact the failed attempt, and we show this as *non-interference* in Table 7. If the closest account is an invisible account, we say that the attempt fails because of the interference effect. In this case, if the clicked location is a correct geopass for the invisible account, we classify it as *accurate interference*, and otherwise we call it *non-accurate interference*.

5.3.2. *Results*

Our results (see Table 7) show that 14.9% of 282 attempts succeeded in *Login 1*, while 44.8% attempts failed because of

Table 6. *Interference Study I*: login performance of the participants (SD).

Sitting	Success Rate (%)	Number of attempts			Login time		
		Mean	Median	SD	Mean	Median	SD
Login 1	58	2.8	1	4.2	59	43	47
Login 2	67	3.1	1	3.7	45	35	39

Table 7. *Interference Study I*: summary of the interference effect (TA, total attempts; SA, successful attempts; Acc., accurate).

Sitting	TA	SA (%)	Failed attempts (%)		
			Interference		
			Acc.	Non-acc.	Non-interference
Login 1	282	14.9	5.0	39.8	40.4
Login 2	309	15.5	1.9	36.3	46.3

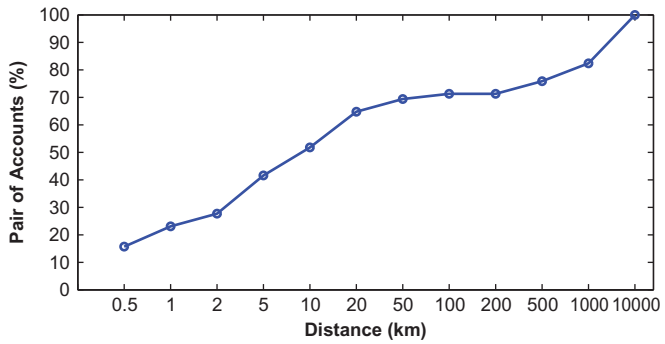

Figure 9. *Interference Study:* distances between geopasses for each pair of accounts.

interference effects (considering both accurate and non-accurate interferences). In *Login 2*, 15.5% of the 309 attempts were successful, and 38.2% attempts failed because of interference effects.

5.4. The causes of interference

It is possible that a user who selects two geopasses near each other may confuse them, leading to interference. We thus seek to determine whether the interference effect between a pair of accounts had any correlation with the distance between corresponding pairs of geopasses. Figure 9 shows an ECDF for the distances between geopasses of each pair of accounts, where over one-quarter (27.8%) of geopass pairs were within 2.0 km.

To measure the interference effect between two accounts for a given participant, such as bank and email, we counted the number of failed attempts T_1 for which bank was the visible account and email was the invisible account with the closest geopass (compared to the geopasses of the movie and deals accounts) to the clicked location. We then calculated T_2 representing the number of unsuccessful attempts for which email was the visible account and bank was the invisible account with the closest geopass to the clicked location. Here, $T = T_1 + T_2$ indicates the interference effect between bank and email accounts for this participant. We also measured the distance D (in kilometers) between the location passwords for the bank and email accounts of the given participant. In this way, we measured D and T for each pair of accounts of every participant and then calculated the correlations between D and T. However, we did not find any strong correlations in this respect either in *Login 1* ($r = 0.12$) or in *Login 2* ($r = 0.02$). Since the participants did not seem confused by geopasses that were geographically close, we speculate that they failed to associate their location passwords with the corresponding accounts. We propose a novel mental story approach to address this interference issue in GeoPass (see Section 6.1).

5.5. Summary

In this interference study, we found that users remembered location passwords in <70% of login sessions, where 41.5% of login attempts failed due to interference effects. For deeper understanding, we investigated both accurate and non-accurate interference, and our analysis shows that the interference effect between a pair of location passwords had no correlation with the geographic distance between them. In other words, participants were not especially confused by location passwords that were geographically near each other. Rather, they failed to associate their location passwords with the corresponding accounts and thus could not log in successfully. The findings from this study motivated us to design a mental story approach to address interference issue in GeoPass. We evaluate the efficacy of this approach in *Interference Study II*, described in the next section.

6. INTERFERENCE STUDY II

In this section, we describe the second interference study, which was aimed at testing the effectiveness of our mental story approach to address the interference issue in GeoPass.

6.1. Mental story

Based on our findings in *Interference Study I* (see Section 5.4), we speculate that the participants failed to associate their location passwords with the corresponding accounts, which contributed to the interference effect and thus to a large number of login failures. Thus, we propose an approach to better link location passwords with their corresponding accounts and thereby address this interference issue.

During registration, users would be asked to build a mental story to make a meaningful association between their location password and the account it is created for. For example, in *Interference Study II* (see Section 6) that we conducted to test the efficacy of our approach, one participant chose a location for the bank account and built a story: 'I had an accident here. The accident could interrupt the financial security of a family.' Another participant chose for the deals account a location in Old Trafford, UK, home to Manchester United (the famous football club), about which she said, '[They] make good deals to get skilled players in the club.'

In our approach, users are asked to type the story in a textbox that is provided along with the Google Maps interface at registration. Users have the flexibility to build and type the story either before or after choosing the location password. This story is not shown at login and thus is not required to be stored by the system. For the purpose of our analysis, however, we did retain the stories.

The mental story approach offers theoretical benefits to memorization. First, the mental story works as a cue to

remember new information, while an elaborative encoding (for new information) from short-term memory to long-term memory takes place when the information can be associated with something meaningful, such as cues (Atkinson and Shiffrin 1968). This encoding helps people to remember and retrieve the processed information efficiently over an extended period of time (Atkinson and Shiffrin 1968). Second, the mental story requires deeper processing of both the place selected and the relevant account type together. The depth of processing effect says that processing the meaning of the information, rather than thinking about it at a shallower level, increases the ability of the user to retain the information (Craik and Tulving 1975). Typing the story may help to engage the user's mind further compared to an entirely mental-only approach.

6.2. Study design

The procedure for *Interference Study II* was same as that for the first interference study, except that participants had to come to the laboratory only for registration in the first sitting and logged in from home for both the second sitting (1 week after registration) and the third sitting (2 weeks after registration). *Interference Study II* is thus, according to the terminology of Biddle *et al.*, (2012), a hybrid study. To get participants to log in from home, we sent emails to the participants with links that would redirect them to our server for logins. They had to complete the logins within 24 hours of getting the email. During registration, the participants were asked to not write down their location passwords.

We note that, as in *Interference Study I*, we found no statistically significant differences between *Login 1* and *Login 2* for number of login attempts, login time or interference effects.

The participants in *Interference Study II* were recruited from a different class than that of the first interference study. No student participated in both interference studies. Out of 60 students in this class, 38 students (mean age: 23) participated in our study. The compensation for participants was same as that in *Interference Study I*. In both studies, participants were notified that their performances in the study would not affect the compensation. They had not taken any courses on usable security or human–computer interaction, nor had they participated in any password-related user study.

6.3. Login performance

Each of the 38 participants logged into four accounts in both *Login 1* and *Login 2*, making a total of 152 login sessions in each sitting. The results for login performances are shown in Table 8.

The overall login success rates were 98% in *Login 1* and 99% in *Login 2*. This performance is particularly impressive given that, unlike in the field study, participants had to recall four different geopasses learned all at once.

The mean number of attempts for successful logins were 1.6 in *Login 1* and 1.8 in *Login 2*, while the median was 1 in both sittings. The mean times for successful logins were found to be 43 seconds in *Login 1* and 39 seconds in *Login 2*, while the medians were 27 seconds in *Login 1* and 25 seconds in *Login 2* (see Table 8).

6.4. Interference effect

We measured the interference effect in the same way as described in Section 5.3.1. Our results (see Table 9) show that 63% of 237 attempts succeeded in *Login 1*, while 4.2% attempts failed because of interference effects (considering

Table 8. *Interference Study II:* login performance of the participants (SD).

		Number of attempts			Login time		
Sitting	Success Rate (%)	Mean	Median	SD	Mean	Median	SD
Login 1	98.0	1.6	0.1	1.0	43	27	2.0
Login 2	99.3	1.8	0.1	1.0	39	25	3.0

Table 9. *Interference Study II:* summary of the interference effect (TA, total attempts; SA, successful attempts; Acc., accurate).

			Failed attempts (%)		
			Interference		
Sitting	TA	SA (%)	Acc.	Non-acc.	Non-interference
Login 1	237	62.9	2.1	2.1	32.9
Login 2	268	56.3	0.0	2.6	41.1

both accurate and non-accurate interference). In *Login 2*, 56% of 268 attempts were successful, and 2.6% attempts failed because of interference effects. In summary, interference effects were much lower with the mental story approach.

6.5. Comparison with Interference Study I

In *Study I*, users remembered location passwords in <70% of login sessions (see Section 5.2 for details). We use a chi-squared test (appropriate for unpaired results) to compare login success rates between *Study I* and *Study II*. Our results show that the login success rate of the participants in *Study II* was significantly higher than that in *Study I*, in both *Login 1*, $\mathcal{X}^2(1, N = 224) = 58.2$, $P < 0.001$, and *Login 2*, $\mathcal{X}^2(1, N = 224) = 49.4$, $P < 0.001$.

The results of Wilcoxon–Mann–Whitney tests (appropriate for unpaired results) show that the number of attempts for successful logins were significantly less in *Study II* than that in *Study I*, in both *Login 1* ($W = 2483.5$, $P < 0.01$) and *Login 2* ($W = 2487.5$, $P < 0.01$). We also found that participants required significantly less time for login in *Study II* in comparison to *Study I*, in both *Login 1* ($W = 2393.5$, $P < 0.05$) and *Login 2* ($W = 3029$, $P < 0.05$).

So, we infer that the login performances in *Study II* were significantly better than that in *Study I*, in terms of success rate, login time and number of attempts for successful logins.

6.6. Summary

Based on our analysis in *Interference Study I*, we hypothesized that interference effects could be reduced if participants would be asked during registration to make a mental story to create a meaningful association between their location password and the corresponding account. In this way, participants could better link the location password to the account for better recall later. In *Interference Study II*, we found a 98% login success rate 1 week after registration and a 99% success rate after 2 weeks, where just 3.4% of login attempts failed because of interference effects. Thus, the mental story approach played an important role to address the interference issue in remembering multiple location passwords.

7. SHOULDER-SURFING STUDY

In this section, we discuss the procedure and results of our study of shoulder surfing in GeoPass. The study of Tari *et al.* (2006) reveals the vulnerability of graphical passwords with mouse input to shoulder-surfing attack. Since GeoPass also requires users to use mouse input, it is important to study the vulnerability of this scheme against shoulder surfing.

In the prior shoulder-surfing study (Tari *et al.*, 2006), participants were given pen and paper to take note while observing the user's authentication secret. GeoPass uses Google Maps interface, and thus it would be interesting to investigate if using online map interface eases the task of shoulder surfer to record users' location passwords.

Unlike textual passwords where users use keyboard input for entering the password, or existing graphical password schemes where users select an image from a set of decoys (Authentication 2004) or choose a set of points in a given image (Wiedenbeck *et al.*, 2005b), in GeoPass, users first navigate to a certain location (e.g. city or town) in an online map, and then select a particular place from that location as her authentication secret. So, for a successful shoulder-surfing attack on geographic location password, the attacker requires to carefully observe the navigation strategy of users, which is not usually required in a shoulder-surfing attack on existing textual or graphical password schemes. Thus, navigation strategies are important to be considered while studying the resilience of GeoPass against shoulder-surfing attack.

Considering the above issues, we address the following research questions in our shoulder-surfing study on GeoPass:

- What would be the success rate of the shoulder surfers in gaining users' credentials?
- Do shoulder surfers have a higher chance to succeed if they use an online map interface instead of just pen and paper to record users' location passwords?
- Does the method of navigation of a user to her location password contribute to the success rate of shoulder surfers?

7.1. Study design

In this study, we recruited 30 participants (7 women, 23 men) from the *Computer Security Club*, a *Computer Security Class* and a *Computer Security Research Laboratory* at our university. Their mean age was 23. The participants were divided into two groups (described below), with 15 participants in each group. The ith and $(i + 1)$th participants were assigned to different groups.

Our observations during laboratory-based interference studies demonstrate that the participants either used the search bar to directly navigate to the location password or typed the name of city/town and then used panning for navigation to the location password.[5] So, we have used these navigation strategies in our shoulder-surfing study.

Procedure: In our study, the experimenter had a one-on-one session with each user. Written consent was obtained from each participant before the beginning of study. We showed and explained GeoPass to the participants and gave an overview of the shoulder-surfing study. In this study, a participant played the role of a *hacker* and the experimenter

[5]In our field study, participants could log in from any computer at any time. So, we were not able to observe the navigation strategies of users during authentication.

played the role of a *victim*, where the experimenter logged in as a user using a desktop computer with a 15-inch monitor.

The participants were divided into two groups according to the method of navigation used by the experimenter to find the location passwords: the *Panning Group* and the *Typing Group*. In the *Panning Group*, the experimenter typed the name of the city in the search bar and then used panning for navigation to the location password. For the *Typing Group*, the experimenter typed the full address of the geopass in the search bar for direct navigation to that location without using panning.

In both groups (*Panning* and *Typing*), every participant was asked to perform shoulder surfing twice. For one shoulder-surfing attempt, the participant was given a pen and paper to take notes, and for the other shoulder-surfing attempt, she was provided with a 7-inch *A13 Google Android tablet* with the Google map interface (at http://www.maps.google.com) open for them to mimic the actions of the experimenter on Google map to gain his location password. We attempted to randomize the order of pen and paper versus tablet, but due to experimental artifacts, the majority of participants used the tablet first before using pen and paper.

Thus, we get four conditions: the *Panning Group* using pen and paper (*Pan-P&P*), the *Panning Group* using the tablet (*Pan-Tab*), the *Typing Group* using pen and paper (*Typ-P&P*) and the *Typing Group* using the tablet (*Typ-Tab*).

For each participant, we used two different geopasses: one in Sydney, Australia (for pen and paper) and another one in Cape Town, South Africa (for the tablet). We expect that most participants would not be familiar with either location.

The experimenter clicked geopasses at zoom level 16 in all logins and maintained an average login time of 35 seconds for the first group and 28 seconds for the second group. Note that the mean login time for GeoPass was 32 seconds in the field study (see Table 2). We let the participants get comfortable using Google maps on the tablet before the start of the study. We hypothesized that the tablet would be easier to use to get precisely the right location.

During shoulder surfing, the participants could stand behind the experimenter, move to any side, or sit next to him to gain the geopass. Then, they were asked to log in using the information they had gained through shoulder surfing. For each participant, the second shoulder-surfing attempt started after she had completed login attempts with the credentials gained through the first shoulder-surfing attempt. Each participant was allowed to make a maximum 10 login attempts. In the post-experiment anonymous paper-based survey, they were asked about the ease of shoulder surfing on GeoPass. They were compensated with a five dollar gift card for participating in this study.

7.2. Login performance

We consider a shoulder-surfing attempt to be successful when the participant is able to log in successfully with the location password that she gains through shoulder surfing. Figures 10–12 illustrate the results of login performance of the participants.

Success rate: The participants in the Typ-P&P condition were the most successful with a 67% login success rate, while the participants in the Pan-P&P condition attained a 60% success rate. In both the Pan-Tab and Typ-Tab conditions, the

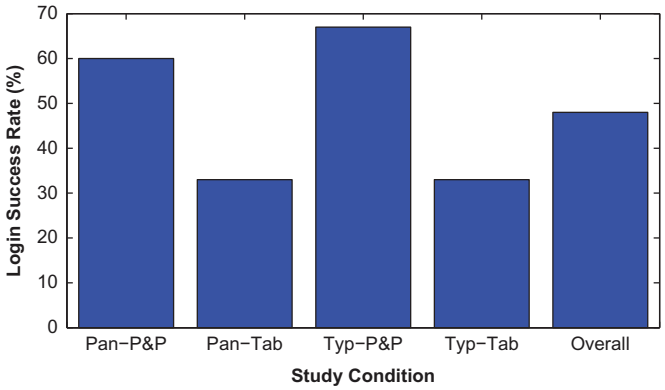

Figure 10. *Shoulder-surfing study*: login success rate.

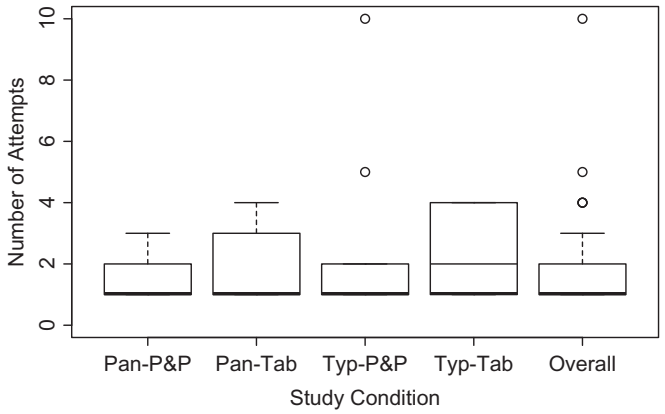

Figure 11. *Shoulder-surfing study*: number of attempts.

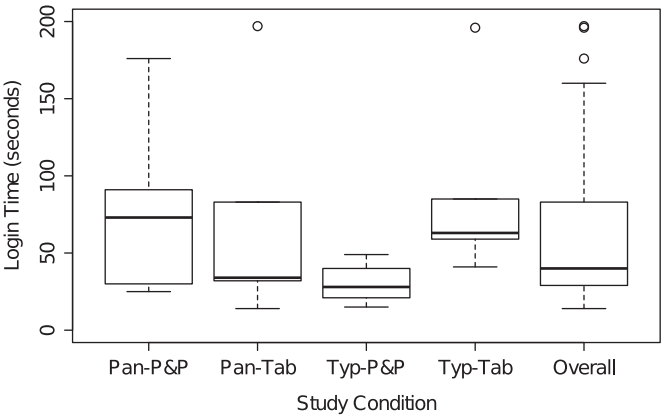

Figure 12. *Shoulder-surfing study*: login time.

login success rate was 33%. The overall success rate (considering all conditions) of the shoulder surfers was 48%.

Login success rates are a binary measure. We divide the pairs into a group for which we have paired results (Pan-P&P: Pan-Tab and Typ-P&P:Typ-Tab. and a group for which results are not paired (Pan-P&P:Typ-P&P, Pan-P&P:Typ-Tab. Pan-Tab:Typ-Tab and Typ-P&P:Pan-Tab). For the paired results, we perform McNemar's tests, while for the unpaired results, we use chi-squared tests. We found no significant differences between any pair of conditions (see Table 10). How a user navigated to the location password, by panning or typing, did not impact the success rate of the attack.

Considering that the success rate was higher for pen-and-paper conditions, though not statistically significant, we conclude that our hypothesis that the tablet would be easier to use for shoulder surfing is not supported by the participants' performances.

Number of attempts: In our shoulder-surfing study, the mean number of attempts required for a successful login was no >3 in any condition, and the median was 1 in all conditions except for Typ-Tab, in which case it was 2. In the Typ-P&P condition, one participant required 10 attempts to log in successfully. No other participant in any condition required >4 attempts for a successful login.

We do not get matched pairs of subjects while comparing login times or numbers of attempts for successful logins in the two study conditions, as a participant might succeed in one condition and fail in another. Thus, we used a Wilcoxon–Mann–Whitney test and found no significant differences between any pair of conditions (see Table 10).

Login time: The mean and median times for successful logins were no <70 seconds and 28 seconds, respectively, in any condition. One participant required 987 seconds, which was the maximum login time in any condition, while the minimum login time for any participant was 14 seconds.

For each pair of conditions, we used a Wilcoxon–Mann–Whitney test to evaluate the differences in time required for successful logins. We found that login time for the Typ-P&P condition was significantly less than that for the Typ-Tab condition ($W = 6$, $P < 0.05$), though again we note the limitation of ordering between the tablet and pen and paper conditions. No significant difference in login time was found between any other pair of conditions (see Table 10).

7.3. Min-distances for unsuccessful logins

In our study, each participant who failed to log in successfully made 10 attempts, which was also the maximum number of attempts allowed. For each unsuccessful participant, we measured the Euclidean distances (in kilometers) between the actual geopass and her selected locations. Then, we selected the closest of these 10 attempts to the actual geopass and recorded this minimum distance (*min-distance*).

Table 10. *Shoulder-surfing study*: significance tests between each pair of conditions. For success rate: McNemar's tests for matched pairs of subjects and chi-squared tests for unpaired results. For number of attempts, login time and min-distance: Wilcoxon signed-rank tests for matched pairs of subjects and Wilcoxon–Mann–Whitney tests for unpaired results. We consider results comparing two conditions to be significantly different when we find $P < 0.05$.

Pair of Conditions	Pan-P&P & Pan-Tab	Pan-P&P & Typ-P&P	Pan-P&P & Typ-Tab	Pan-Tab & Typ-P&P	Pan-Tab & Typ-Tab	Typ-P&P & Typ-Tab
Success rate	$\mathcal{X}^2(1, N = 15) = 1.5$, $P = 0.22$	$\mathcal{X}^2(1, N = 15) = 0$, $P = 1$	$\mathcal{X}^2(1, N = 15) = 1.21$, $P = 0.27$	$\mathcal{X}^2(1, N = 15) = 2.13$, $P = 0.14$	$\mathcal{X}^2(1, N = 15) = 0$, $P = 1$	$\mathcal{X}^2(1, N = 15) = 2.28$, $P = 0.13$
Number of attempts	$W = 18.5$, $P = 0.59$	$W = 44$, $P = 0.96$	$W = 14$, $P = 0.23$	$W = 23.5$, $P = 0.88$	$W = 10$, $P = 0.65$	$W = 19.5$, $P = 0.49$
Login time	$W = 22.5$, $P = 1$	$W = 67$, $P = 0.08$	$W = 18$, $P = 0.59$	$W = 19$, $P = 0.5$	$W = 8$, $P = 0.42$	$\boldsymbol{W = 6, P < 0.05}$
Min-distance	$W = 32$, $P = 0.87$	$W = 21$, $P = 0.3$	$W = 27$, $P = 0.79$	$W = 34$, $P = 0.28$	$W = 47.5$, $P = 0.88$	$W = 11$, $P = 0.1$

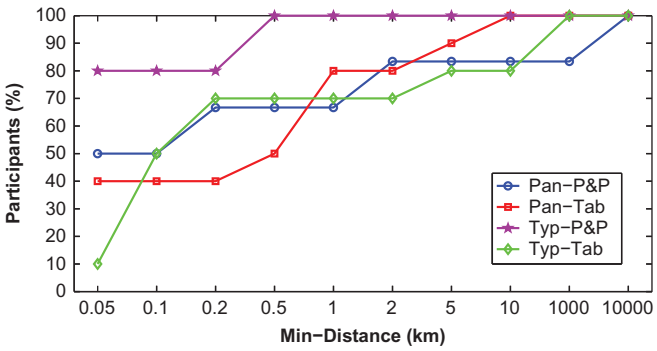

Figure 13. *Shoulder-surfing study*: minimum distances between guessed locations and actual geopasses for failed logins.

Figure 13 shows an ECDF of min-distances, where we find that the min-distance was no >0.05 km for 50% of participants in Pan-P&P, 40% in Pan-Tab, 80% in Typ-P&P and 10% in Typ-Tab. Overall, 48% of participants had a min-distance of no >0.05 km. Note that here we consider those participants only who failed to log in successfully after the maximum number of allowed attempts.

We do not get matched pairs of subjects while evaluating the difference in min-distances for failed logins of two study conditions, so we conducted a Wilcoxon–Mann–Whitney test. We did not find any significant difference between any pair of conditions for min-distances (see Table 10).

7.4. Summary

Our shoulder-surfing study demonstrates that GeoPass is vulnerable to this attack. Participants playing the role of attackers had success rates of 48% overall and 60–67% when using pen and paper (Section 7.2). It is clear from these findings that more attention must be paid from the research community to improve the resilience of geographical location-password schemes like GeoPass against shoulder surfing. Our analysis found that how a user navigated to the location password, either by panning or typing the full address in the search bar, did not significantly affect the success rate of shoulder surfing.

8. DISCUSSION

We reported four separate studies in this article, where three of our major contributions include: (i) understanding the training effect in improving the login performances for geographic location passwords in a real-world context, (ii) the efficacy of the mental story approach to reduce interference effects and thereby improve the memorability for multiple location passwords and (iii) evaluating the resilience of location-password schemes to shoulder-surfing attack.

In light of our findings, we believe that further study is needed to identify the applications of GeoPass. In this section, we identify several aspects that need to be addressed in the future work on location-password schemes. In particular, GeoPass is vulnerable to shoulder-surfing attacks, and its login time is higher than for traditional textual passwords, and these issues may prevent widespread deployment.

8.1. Field study

As noted by Biddle *et al.* (2012), a field study offers a good measure of login performance in a realistic setting. In particular, field studies help ensure that logging in becomes a secondary task for users as they are primarily seeking to use the system, plus it provides a realistic motivation to memorize the authentication secret. Our field study shows a satisfactory login success rate for GeoPass when users have to remember a single location password.

8.1.1. *Training effect*

Since the prior field studies on graphical passwords did not present a detailed analysis of the training effect, it remains unknown to the research community how login performances change over login sessions in a long-term field study. Our close examination on the training effect found an overall improvement in login performance with more login sessions, with some modest fluctuations. While Thorpe *et al.* (2013) reported concerns about the login time of GeoPass in their short-term laboratory study, our field study found a 59% reduction in median login time to just 17 seconds by the 31st login session. So, it is clear that field studies like ours provide a deeper understanding on the change in login performance with more login sessions.

Chiasson *et al.* (2007) argued that the login success rate in a laboratory study would be higher than that in a field study, since the participants in a laboratory setting are primarily focused on their login attempts, while in a field study, authentication is a secondary task. Indeed, we found that the login success rate was 84% in the first login session, which is less than the study by Thorpe *et al.* (2013).[6] The training effect, however, compensated for this, leading to a 94% success rate in the 7th login session and a 100% success rate in the 13th login session.

In summary, with enough login sessions to induce training effects, the usability and memorability of GeoPass appears to be even better than reported from the laboratory studies of Thorpe *et al.* (2013).

8.1.2. *Lockout rules*

Lockout rules are implemented in many systems to protect against online guessing attacks (Florencio *et al.*, 2007). To implement a lockout rule that is both secure and convenient

[6]We note that direct comparison between different studies should be taken with caution.

for legitimate users, it is important to figure out the number of attempts an actual user would usually require to log in successfully. Our field study gives insight to this issue, as we found that 90% of participants made at most two attempts on average to authenticate successfully, and 100% of participants logged in successfully within five attempts on average. Thus, GeoPass is amenable to reasonable lockout rules.

8.2. Interference studies

In our interference studies, we consider geographic distance to understand the interference effects on GeoPass. We distinguish the login attempts that failed because of confusion with other passwords (i.e. likely interference) from attempts that failed due to simply forgetting the desired password (i.e. non-interference). We investigate both accurate and non-accurate interferences for an in-depth understanding of interference effects. As noted by Biddle *et al.* (2012), how to best evaluate multiple-password interference still remains an open issue; our methodology for analyzing the interference effect should make an important contribution in this regard. In future evaluations, we would further improve our interference model by considering the interference due to confusion between two similar types of locations, such as two small islands in the ocean.

8.2.1. Mental story

Our studies suggest that having users create a mental story contributes to reducing interference effects and ensuring high memorability when users have to remember multiple location passwords. In *Interference Study I*, users remembered location passwords in $<70\%$ of login sessions, with substantial interference between different passwords. In *Interference Study II*, the success rates were at least 98%, with very low rates of interference. While it would be inappropriate to compare the two studies statistically due to the different study populations, the success of *Interference Study II* suggests that the mental story approach is promising and deserves further refinement and investigation.

Previously, Davis *et al.* (2004) implemented the concept of mental story to design a recognition-based graphical password. Their purpose and approach of leveraging mental stories were different from ours. Their study tested the memorability for a single graphical password, where users were asked to build a story to remember a set of images in correct order. The study (Davis *et al.*, 2004) found a 85% login success rate over the span of 1 week, while the registration and login times were not reported. In our approach, we aim to leverage the mental story approach to reduce multiple-password interference by asking users to create a meaningful association between their location password and the account it is created for.

We note that the mental story approach requires further investigation to understand its full potential in improving password memorability. First, it is not clear what the effect is on security for users to pick locations related to their account in some way. We found some imaginative ways to incorporate the account information, but one might imagine many people picking Fort Knox for a banking site and the Facebook company headquarters for their Facebook passwords. In general, Geopass could have problems with common locations; blocking the most common ones could be effective much like proactive dictionary checks on textual passwords can improve password strength. To improve the security of geographic location password, it would be also interesting to leverage the findings of Chiasson *et al.* (2008) that addressed the 'hotspot' issue in graphical password by nudging users to choose less obvious image points. Second, we cannot be sure that users will follow the requested steps in real-world deployments. Automated evaluation of the story quality, e.g. checking for minimum length and the presence of real words instead of gibberish, could help to ensure that users are taking the technique seriously.

While we attempted to categorize the mental stories collected from our study, they seem quite random at this point. We believe that we need to collect more mental stories for their generic categorization and understanding their impact on the effective password space of location-password schemes. In this case, the high login success rate with mental story approach indicates that this method worked well for users in our study irrespective of the type of mental stories they created. In future work, it would be important to run a large study to collect enough mental stories so as to categorize them and understand the impact of different types of mental stories on the usability and security (e.g. impact on effective password space) of location passwords.

We evaluated the efficacy of mental stories for multiple accounts of different types (e.g. bank, email, etc.). In future work, we would examine the effectiveness of mental stories for multiple accounts of the same type (e.g. several banks). In addition, we would focus on measuring the cognitive effort of building mental stories by people from different age groups and backgrounds.

8.3. Shoulder-surfing attack and other security issues

While prior work shows that location passwords created with GeoPass can have reasonable security against online guessing attacks, even when accounting for social engineering attacks (Hang *et al.*, 2015; Thorpe *et al.*, 2013), we demonstrate that it requires more attention from the research community to improve the resilience of location passwords to shoulder surfing, phishing and malware attacks.

Tari *et al.* (2006) performed a shoulder-surfing study similar to ours, where the university students played the role of attackers and the experimenter played the role of a victim. They found that neither traditional textual passwords nor graphical

passwords (with mouse input) provide sufficient resilience against shoulder surfing. Our results for shoulder-surfing study showed a 48% overall success rate for participants playing the role of attackers and an experimenter playing the role of a victim. However, because of the accuracy required to click on the location password, GeoPass might provide higher resilience to shoulder surfing in comparison to recognition-based graphical passwords with mouse inputs; in the study by Tari *et al.* (2006), participants playing the role of shoulder surfers had a success rate of 62% in the shoulder surfing on commercially available Passfaces scheme that uses mouse inputs (Authentication 2004).[3]

Harbach *et al.* (2014) studied the perceived security of shoulder-surfing attack, where users were found to consider this attack less of a risk in real life. In a discussion on their findings, the authors note that just because users do not perceive shoulder surfing a threat as serious does not mean that it is not. Rather, it indicates that the additional effort a user is willing to invest to protect from this attack should be carefully assessed.

Our shoulder-surfing study on GeoPass calls for further attention from the research community for better defenses. To address this challenge, we would examine the efficacy of multiple technologies in our future research that include but are not limited to: (i) using LCD screens with concurrent dual views, which show different images at various viewing angles, e.g. when not positioned directly in front of the screen (Kim *et al.*, 2012), and (ii) interacting with the system through eye-gaze input (Forget *et al.*, 2010). Forget *et al.* (2010) showed the efficacy of eye-gaze entry to gain resilience against shoulder surfing.

In addition to shoulder surfing and online guessing, phishing and malware attacks could pose security threats for GeoPass, which, like most password schemes, provides no additional resistance to these attacks. It may seem that GeoPass would present a challenge to malware that attempted to gain a user's password in the way that keystroke loggers can steal typed passwords. However, malware could use a combination of a keystroke logger to capture typed addresses in the search bar, a screen scraper to capture the map as shown to the user, and a mouse logger to capture the point on the screen selected as the location password.

Another security issue for a password system is secure storage at the server. Textual passwords, for example, should not be stored in plaintext, but rather scrambled using a slow, salted cryptographic hash function like bcrypt (Provos and Mazieres 1999) or PBKDF2 (Kaliski 2000). The prior work on location-password schemes offers no guidance on how to securely store location passwords. We note that one approach would be to round the selected coordinates to a level of precision that minimizes the number of coordinate pairs that match the user's location password. Then these few coordinate pairs could be salted and hashed. At login, if the hash of the coordinates entered by the user matches any of the stored hash values, the user would be authenticated. Selecting the appropriate level of precision may be challenging when balancing security with providing a tolerance for user error, which is an important research issue requiring more study in future work.

8.4. Limitations and ecological validity

8.4.1. *Field study*

In our field study, the participants were generally young and university educated, and our findings thus may not generalize to the entire population of Web users. Given that the participants, however, were performing a real-life task that mattered in their specific situation, the task likely had reasonable ecological validity. It is a common approach in password studies to examine the usability of a scheme through analyzing the login performance of university students (Al-Ameen *et al.*, 2015, 2015a; Chiasson *et al.*, 2007, 2008, 2007; Everitt *et al.*, 2009; Hlywa *et al.*, 2011; Wiedenbeck *et al.*, 2005a; Wright *et al.*, 2012). We note that it is important to do further studies for people from different age groups and backgrounds. Now that our results from a field study show promise, we would examine the usability of GeoPass for senior users and people with cognitive limitations to have a greater understanding of its usability.

We did not have a control condition in our field study, and thus, we do not make a direct comparison between GeoPass and traditional textual passwords in terms of login performance. A comprehensive survey on 25 different password schemes shows that only 6 of these schemes were evaluated through field studies (Biddle *et al.*, 2012), and none of these six field studies included a control condition. In future field studies, given enough participants, having a control condition would offer the opportunity to make direction comparisons between two schemes.

In our field study, participants could log in from any computer at any time. So, unlike laboratory-based studies, the experimenters were not able to observe the navigation strategies of users during authentication. For future studies, we would improve our system so that it could keep track of the navigation strategies of users, such as how many users have used search bar and how much panning has been used.

8.4.2. *Interference studies*

As with the field study, participants in our interference studies were young and university educated, which may not generalize to the entire population. While laboratory studies are preferred to examine the primary usability issues and set performance bounds for an authentication scheme, a hybrid study could provide higher ecological validity when login sessions are performed online (Biddle *et al.*, 2012). So, we conducted the first interference study in a laboratory setting to understand the causes and effects of interference, and then designed a hybrid study for *Interference Study II* to test the

efficacy of our proposed mental story approach. It is possible that the hybrid setting in *Interference Study II* helped us recruit more participants than *Interference Study I*, since participants did not need to come to laboratory for the login sessions. We had a 1-week interval between each session, since the 1-week delay is larger than the maximum average interval for a user between her subsequent logins to any of her important accounts (Hayashi and Hong 2011).

In real life, users have to remember >4 passwords. Prior multiple-password studies (Everitt *et al*., 2009; Hlywa *et al*., 2011; Wright *et al*., 2012) asked users to remember either three or four passwords. Being consistent with the prior work, we asked the participants to remember four location passwords in our study, in which all the passwords were created in the same sitting. This registration process is in agreement with prior work (Hlywa *et al*., 2011; Wright *et al*., 2012), while in real life, the geopasses would likely be created over time and possibly in different contexts, e.g. in different rooms or with different computers.

8.4.3. Shoulder-surfing study

Since shoulder surfing is performed by people who intend to steal authentication secrets, possibly professional hackers, recruiting a participant group representing the true population is very difficult. However, our participants satisfied the requirements of Tari *et al*. (2006), who stated that students with reasonable background on computer security and authentication systems can be representative of 'potential shoulder surfers' in a laboratory study. We note that for comparisons between pen and paper and tablet conditions that the majority of participants used the tablet first.

9. CONCLUSION

As GeoPass showed promise in the primary investigation of its guessing resilience and memorability (Thorpe *et al*., 2013), we explored deeper into its potential through four separate studies and identify the promising aspects for future research. Our field study represented the first long-term real-world study of any geographic location-password scheme, and we found that the login success rate was satisfactory and quantified the degree to which overall login performance improved with more login sessions. In a multiple-password interference study, we found the empirical evidence that interference from having to remember multiple location passwords is problematic. Based on our findings, we propose to leverage mental stories in reducing interference effects on GeoPass and found a high login success rate with this approach. Our shoulder-surfing study is the first ever experiment measuring the resilience of location passwords to such attacks, and our findings call for further attention from the research community for better defenses.

REFERENCES

Al-Ameen, M.N., Fatema, K., Wright, M. and Scielzo, S. (2015a) The Impact of Cues and User Interaction on the Memorability of System-Assigned Recognition-Based Graphical Passwords. In Symposium on Usable Privacy and Security (SOUPS). Ottawa, Canada.

Al-Ameen, M.N., Fatema, K., Wright, M. and Scielzo, S. (2015b) Leveraging Real-Life Facts to Make Random Passwords More Memorable. In ESORICS. Vienna, Austria.

Al-Ameen, M.N., Wright, M. and Scielzo, S. (2015) Towards Making Random Passwords Memorable: Leveraging Users' Cognitive Ability Through Multiple Cues. In ACM Conf. Human Factors in Comput. Syst. (CHI). ACM, Seoul, Korea.

Atkinson, R.C. and Shiffrin, R.M. (1968) Human memory: a proposed system and its control processes. Psychol. Learn. Motiv., 2, 89–195.

Authentication, Real User Personal. (2004) The Science Behind Passfaces. White Paper, June.

Bianchi, A., Oakley, I. and Kim, H. (2016) PassBYOP: bring your own picture for securing graphical passwords. IEEE Trans. Human-Mach. Syst, 46, 380–389.

Biddle, R., Chiasson, S. and Van Oorschot, P.C. (2012) Graphical passwords: learning from the first twelve years. ACM Comput. Sur. (CSUR), 44, 19.

Bonneau, J., Herley C., van Oorschot P.C. and Stajano F. (2012) The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In IEEE Symposium on Security and Privacy. San Francisco, USA.

Campbell, J., Ma, W. and Kleeman, D. (2011) Impact of restrictive composition policy on user password choices. Behaviour & Info. Technol., 30, 379–388.

Chiasson, S., Biddle, R. and van Oorschot, P.C. (2007) A second look at the usability of click-based graphical passwords. In Proc. 3rd symposium on Usable Privacy and Security, ACM. pp. 1–12.

Chiasson, S., Forget, A., Biddle, R. and van Oorschot, P.C. (2008) Influencing Users Towards Better Passwords: Persuasive Cued Click-points. In Proc.22nd British HCI Group Ann. Conf. People and Computers: Culture, Creativity, Interaction-Volume 1. British Computer Society, pp 121–130.

Chiasson, S., Forget, A., Biddle, R. and van Oorschot, P.C. (2009a) User interface design affects security: patterns in click-based graphical passwords. Int. J. Info. Security, 8, 387–398.

Chiasson, S., Forget, A., Stobert, E., van Oorschot, P.C. and Biddle, R. (2009b) Multiple password interference in text and click-based graphical passwords. In CCS, ACM. pp. 500–511.

Chiasson, S., Stobert, E., Forget, A., Biddle, R. and Van Oorschot, P.C. (2012) Persuasive cued click-points: design, implementation and evaluation of a knowledge-based authentication mechanism. IEEE Trans. Dependable Secure Comput., 9, 222–235.

Chiasson, S., van Oorschot, P.C. and Biddle, R. (2007) Graphical Password Authentication Using Cued Click Points. In ESORICS. Pittsburgh, USA.

Craik, F.I.M. and Tulving, E. (1975) Depth of processing and the retention of words in episodic memory. J. Experimental Psychol.: general, 104, 268.

Das, A., Bonneau, J., Caesar, M., Borisov, N. and Wang, X.F. (2014) The Tangled Web of Password Reuse. In Symposium on Network and Distributed System Security (NDSS), Internet Society. San Diego, USA.

Davis, D., Monrose, F. and Reiter, M.K. (2004) On User Choice in Graphical Password Schemes. In Proc. USENIX Security Symposium, Vol. 13, 11. San Diego, USA.

Dirik, A.E., Memon, N. and Birget, J.-C. (2007) Modeling User Choice in the PassPoints Graphical Password Scheme. In Proc. 3rd Symposium on Usable Privacy and Security, ACM. pp. 20–28. Pittsburgh, USA.

Dunphy, P. and Yan, J. (2007) Do Background Images Improve Draw a Secret Graphical Passwords?. In CCS.

Everitt, K.M., Bragin, T., Fogarty, J. and Kohno, T. (2009) A Comprehensive Study of Frequency, Interference and Training of Multiple Graphical Passwords. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM. pp. 889–898.

Florencio, D., Herley, C. and Coskun, B. (2007) Do strong web passwords accomplish anything? In Proceedings of HotSec, Vol. 76.

Forget, A. (2012) A world with many authentication schemes. Ph.D. Thesis, Carleton University.

Forget, A., Chiasson, S. and Biddle, R. (2010) Shoulder-Surfing Resistance with Eye-Gaze Entry in Cued-Recall Graphical Passwords. In CHI.

Forget, A., Chiasson, S., van Oorschot, P.C. and Biddle, R. (2008) Improving Text Passwords Through Persuasion. *Proceedings of the 4th Symposium on Usable Privacy and Security*, ACM. pp. 1–12. Pittsburgh, USA.

Hang, A., De Luca, A., Smith, M., Richter, M. and Hussmann, H. (2015) Where Have You Been? Using Location-Based Security Questions for Fallback Authentication. In SOUPS.

Haque, S.M., Wright, M. and Scielzo, S. (2013) A Study of User Password Strategy for Multiple Accounts. In Proc. 3rd ACM Conf. Data and application security and privacy, ACM. pp. 173–176.

Harbach, M., von Zezschwitz, E., Fichtner, A., De Luca, A. and Smith, M. (2014) It'sa Hard Lock Life: A Field Study of Smartphone (un) Locking Behavior and Risk Perception. In SOUPS.

Hayashi, E. and Hong, J. (2011) A Diary Study of Password Usage in Daily Life. In Proc. SIGCHI Conf. Human Factors in Computing Systems, ACM. pp. 2627–2630.

Hlywa, M., Biddle, R. and Patrick, A.S. (2011) Facing the Facts About Image Type in Recognition-Based Graphical Passwords. In Proc. 27th Annual Comput. Security Appl. Conf., ACM.

Jermyn, I., Mayer, A.J., Monrose, F., Reiter, M.K., Rubin, A.D., *et al.* (1999) The Design and Analysis of Graphical Passwords.Proc. Usenix Security Symposium, Vol. 8. Washington, DC, USA.

Kaliski, B. (2000) RFC 2898: PKCS #5: password-based cryptography specification version 2.0. https://www.ietf.org/rfc/rfc2898.txt .

Kim, S., Cao, X., Zhang, H. and Tan, D. (2012) Enabling Concurrent Dual Views on Common LCD Screens. In CHI.

Luca, A.D., von Zezschwitz, E., Pichler, L. and Hussmann, H. (2013) Using Fake Cursors to Secure On-Screen Password Entry. In CHI.

Nali, D. and Thorpe, J. (2004) *Analyzing User Choice in Graphical Passwords*. Technical Report.

Provos, N. and Mazieres, D. (1999) A Future-Adaptable Password Scheme. In USENIX ATC.

Shay, R., Bauer, L., Christin, N., Cranor, L.F., Forget, A., Komanduri, S., Mazurek, M.L., Melicher, W., Segreti, S.M. and Ur, B. (2015) A Spoonful of Sugar? The Impact of Guidance and Feedback on Password-Creation Behavior. In CHI.

Shay, R., Kelley, P.G., Komanduri, S., Mazurek, M.L., Ur, B., Vidas, T., Bauer, L., Christin, N. and Cranor, L.F. (2012) Correct Horse Battery Staple: Exploring the Usability of System-Assigned Passphrases. In Proc. 8th Symposium on Usable Privacy and Security, 7. ACM. Washington, DC, USA.

Shay, R., Komanduri, S., Durity, A.L., Huh, P.(.S.)., Mazurek, M.L., Segreti, S.M., Ur, B., Bauer, L., Christin, N. and Cranor, L.F. (2014) Can Long Passwords Be Secure and Usable? In CHI.

Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N. and Cranor, L.F. (2010) Encountering Stronger Password Requirements: User Attitudes and Behaviors. In Proc. 6th Symposium on Usable Privacy and Security, ACM. 2. Redmond, USA.

Spitzer, J., Singh, C. and Schweitzer, D. (2010) A security class project in graphical passwords. J. Comput. Sci. College, 26, 7–13.

Sun, H., Chen, Y., Fang, C. and Chang, S. (2012) A Map Based Graphical-Password Authentication Scheme. In ASIACCS, ACM.

Tari, F., Ozok, A. and Holden, S. (2006) A Comparison of Perceived and Real Shoulder-Surfing Risks Between Alphanumeric and Graphical Passwords. In SOUPS.

Thorpe, J., MacRae, B. and Salehi-Abari, A. (2013) Usability and Security Evaluation of GeoPass: A Geographic Location-Password Scheme. In Proc. 9th Symposium on Usable Privacy and Security, 14. ACM. Newcastle, UK.

Ur, B., Bees, J., Segreti, S., Bauer, L., Christin, N., Cranor, L. and Deepak, A. (2016) Do Users' Perceptions of Password Security Match Reality?. In CHI.

Ur, B., Noma, F., Bees, J., Segreti, S.M., Shay, R., Bauer, L., Christin, N. and Cranor, L.F. (2015) I Added '!' at the End to Make It Secure: Observing Password Creation in the Lab. In SOUPS.

Valentine, T. (1998) An Evaluation of the Passface Personal Authentication System. Technical Report.

von Zezschwitz, E., De Luca, A., Brunkow, B. and Hussmann, H. (2015a) SwiPIN: Fast and Secure Pin-Entry on Smartphones. In CHI.

von Zezschwitz, E., De Luca, A., Janssen, P. and Hussmann, H. (2015b) Easy to draw, but hard to trace? On the observability of grid-based (un)lock patterns.CHI. ACM.

Wiedenbeck, S., Waters, J., Birget, J., Brodskiy, A. and Memon, N. (2005a) PassPoints: design and longitudinal evaluation of a

graphical password system. Int. J. Human-Comput. Studies, 63, 102–127.

Wiedenbeck, S., Waters, J., Birget, J.-C., Brodskiy, A. and Memon, N. (2005b) Authentication Using Graphical Passwords: Effects of Tolerance and Image Choice. In Proceedings of the Symposium on Usable Privacy and Security, ACM. pp. 1–12. Pittsburgh, USA.

Wright, N., Patrick, A.S. and Biddle, R. (2012) Do You See Your Password? Applying Recognition to Textual Passwords. In Proceedings of the Symposium on Usable Privacy and Security, ACM. p. 8. Washington, DC, USA.

Zakaria, N.H., Griffiths, D., Brostoff, S. and Yan, J. (2011) Shoulder Surfing Defence for Recall-based Graphical Passwords. In SOUPS.