Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Int. J. Human-Computer Studies 63 (2005) 102–127

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

PassPoints: Design and longitudinal evaluation of a graphical password system

Susan Wiedenbeck[a,*], Jim Waters[a], Jean-Camille Birget[b], Alex Brodskiy[c], Nasir Memon[c]

[a] *College of Information Science & Technology, Drexel University, Philadelphia, PA 19104, USA*
[b] *Department of Computer Science, Rutgers, The State University of New Jersey, Camden, NJ 08102, USA*
[c] *Department of Computer Science, Polytechnic University, Brooklyn, NY 11201, USA*

Abstract

Computer security depends largely on passwords to authenticate human users. However, users have difficulty remembering passwords over time if they choose a secure password, i.e. a password that is long and random. Therefore, they tend to choose short and insecure passwords. Graphical passwords, which consist of clicking on images rather than typing alphanumeric strings, may help to overcome the problem of creating secure and memorable passwords. In this paper we describe PassPoints, a new and more secure graphical password system. We report an empirical study comparing the use of PassPoints to alphanumeric passwords. Participants created and practiced either an alphanumeric or graphical password. The participants subsequently carried out three longitudinal trials to input their password over the course of 6 weeks. The results show that the graphical password users created a valid password with fewer difficulties than the alphanumeric users. However, the graphical users took longer and made more invalid password inputs than the alphanumeric users while practicing their passwords. In the longitudinal trials the two groups performed similarly on memory of their password, but the graphical group took more time to input a password.

Keywords: Graphical password; Alphanumeric password; PassPoints; Authentication; Password security; Usable security

[*] Corresponding author. Tel.: +1 215 895 2490; fax: +1 215 895 2494.
E-mail address: Susan.Wiedenbeck@cis.drexel.edu (S. Wiedenbeck).

1071-5819/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2005.04.010

1. Introduction

Today's world of networked computing can be a frightening and dangerous place with attackers, hackers, crackers, scammers, and spammers at work. Computer security, which was relatively simple in prenetwork days, is now a major and expensive problem for organizations and individuals. Constant attention to security is needed to protect against damage or theft of one's electronic assets. A home computer user installing high-speed internet service and a wireless network cannot even begin without a strong firewall, up-to-date virus protection, and 128-bit encryption. As Edward Tenner writes in his book *Why Things Bite Back*, "The price of protection is chronic vigilance" (1977, p. 243).

The community of security researchers and practitioners has evolved rapidly in response to threats, on the one hand increasing vigilance in practice and, on the other hand, driving research innovation. Until recently the security problem has been formulated as a technical problem. However, it is now becoming widely recognized that security is also fundamentally a human–computer interaction (HCI) problem (Patrick et al., 2003; Dourish, 2004). Most security mechanisms cannot be effective without taking into account the user. HCI matters in two ways: the usability of the security mechanisms themselves and the interaction of the security mechanisms with user practices and motivations.

One of the key areas in security research and practice is authentication, the determination of whether a user should be allowed access to a given system or resource. Traditionally, alphanumeric passwords have been used for authentication. Today other methods, including biometrics and smart cards (Scholtz and Johnson, 2002; Coventry et al., 2003), are possible alternatives. However, passwords are likely to remain dominant for some time because of drawbacks of reliability, security, or cost of other technologies (Brostoff and Sasse, 2000). In particular, smart cards also need PINs and passwords, while biometrics raises privacy concerns. Passwords also have drawbacks, most notably in terms of memorability and security. This has led to innovations to improve passwords. One such innovation is graphical passwords, i.e. passwords that are based on images rather than alphanumeric strings. The underlying idea is that using images will lead to greater memorability and decrease the tendency to choose insecure passwords, which will in turn increase overall password security. Several graphical password systems, described in the next section, have been developed and some HCI evaluation has been done.

In this paper, we describe a new, more flexible, and more secure graphical password system that we have designed and implemented. We discuss the security properties of the system compared to alphanumeric passwords and some other graphical password systems. Then we report on an empirical study comparing the use of our graphical passwords to alphanumeric passwords in a longitudinal study, focusing on user performance and perceptions.

Our two principal research questions are stated below:

RQ1: Are graphical passwords a viable alternative to alphanumeric passwords in terms of security, as well as password creation, learning, performance, and retention?

RQ2: Are users' perceptions of graphical passwords different from those of alphanumeric passwords?

2. Background to the research

In this section, we first describe problems associated with the use of traditional alphanumeric passwords. Following that we introduce several existing graphical password systems and discuss their strengths and weaknesses.

2.1. The password problem

We refer to the security and usability problems associated with alphanumeric passwords as "the password problem." The problem arises because passwords are expected to comply with two conflicting requirements, namely:

(1) Passwords should be easy to remember, and the user authentication protocol should be executable quickly and easily by humans.
(2) Passwords should be secure, i.e. they should look random and should be hard to guess; they should be changed frequently, and should be different on different accounts of the same user; they should not be written down or stored in plain text.

Meeting both of these requirements is almost impossible for users. The problem is well known in the security community. Classical studies going back over 25 years (Morris and Thompson, 1979; Feldmeier and Karn, 1990; Klein, 1990) have shown that, as a result, human users tend to choose and handle alphanumeric passwords very insecurely. More recent studies confirm these results (Sasse et al., 2001; Brown et al., 2004).

The password problem arises primarily from fundamental limitations of human long-term memory (LTM). Once a password has been chosen and learned the user must be able to recall it to log in. However, people regularly forget their passwords. The Power Law of Forgetting describes rapid forgetting soon after learning, followed by very slow drop-off thereafter (Bahrick, 1984; Wixted and Ebbesen, 1991). Psychological theories have attributed forgetting to decay through the passage of time and to interference, in which new items in memory disrupt existing ones (retroactive interference) or, conversely, are disrupted by existing ones (proactive interference). A recent review emphasizes the importance of retroactive interference in everyday forgetting (Wixted, 2004).

Decay and interference help to explain why people forget passwords. Users are expected to learn a password and remember it over time. However, other items in memory compete with the password and can prevent its accurate recall. If a password is not used frequently it will be especially susceptible to forgetting. Research has shown that when users fail to recall a password, they often are able to recall parts of it correctly (Sasse et al., 2001). However, the use of passwords in authentication is predicated on completely accurate recall, so a partially correct password has no value.

Furthermore, today users have many passwords for computers, networks, web sites, and more. In addition, some computer systems require frequent password changes, in a probably misguided effort to increase security. This proliferation of passwords increases potential interference and is likely to lead either to forgetting passwords or forgetting which system a password is associated with.

What is a user to do? Most often a user will decrease the memory burden at the expense of security. Perhaps most commonly, the user will write down passwords and keep them in a convenient place, raising the potential of compromise of the passwords. Available evidence indicates that writing down passwords is a common practice (Adams and Sasse, 1999). In the case of multiple systems, users may decide to use one password for all systems, reducing security and putting the password owner at risk of widespread damage if the password file of one of the systems is breached (Ives et al., 2004). Alternatively, users may make their own rules to generate multiple passwords that are linked by a common element (e.g. adding a digit to a base word for each new password), also an unsafe practice. Secure single sign-on schemes are commercially available and can simplify password handling for users, but this solution is not feasible in every situation.

From a security viewpoint, the ideal password is a string of eight or more random characters, including digits, letters with a mixture of upper and lower case, and special characters. A random password lacks meaningful content, context, and familiarity. It can only be learned by rote. However, since rote repetition is a weak way of remembering (Rundus, 1971), users tend to skirt or ignore entirely the recommendations for pseudo-random passwords. Surveys show that frequent passwords are the word "password," personal names of family members, names of pets, and dictionary words (Sasse et al., 2001; Brown et al., 2004). Passwords also tend to be short. Users view authentication as an enabling task (Adams and Sasse, 1999). They want to input the password quickly and get on to real work. This, too, can lead to short, simple passwords. Weak passwords are susceptible to dictionary attacks or attacks based on knowledge about the password owner. However, to users the most salient consideration is having a password that they can remember and input quickly. The use of weak passwords may be promoted by naiveté of users, who do not realize the power of a dictionary attack, e.g. believing that spelling a dictionary word backwards will foil an attacker. Also naively, users may believe that they have little to lose if their computer is broken into.

One approach to increasing password security is education of the user (Adams and Sasse, 1999; Sasse et al., 2001). However, given the mismatch between passwords and human capabilities, the likelihood of fundamental change is not great (Nielsen, 2004). A better way to overcome the password problem is to develop password systems that reduce fundamental memory problems while preserving security.

3. The case for graphical passwords

The first idea for graphical passwords was described by Blonder (1996). His approach was to let the user click, with a mouse or stylus, on a few chosen regions in an image that appeared on the screen. If the correct regions were clicked in, the user was authenticated, otherwise the user was rejected.

3.1. Creation and learning

From a human viewpoint, the problem of creating a password is making it memorable so that the user can retrieve it later. In a graphical password system, a user choosing click locations in an image needs to choose memorable locations. There are two issues in memorability: the nature of the image itself and the sequence of click locations. In terms of the choice of image, studies of perception indicate that in a jumbled image people will be slow to recognize individual objects (Biederman et al., 1973). Therefore, at the very least images should be coherent. For high memorability of click locations, images should have semantically meaningful content, since meaning for arbitrary things is poor (Norman, 1988). This suggests that abstract images (e.g. swirls of color) may be less memorable than concrete scenes. LTM does not store a replica of the image itself, but rather a meaningful interpretation of the image that captures the meaning of the image but not unimportant visual details (Mandler and Ritchey, 1977). Thus, a user will be dependent on the encoding used while learning to retrieve the locations. A simple way to store information in LTM is rehearsal (Rundus, 1971), or repetition, of the material, e.g. repeatedly visualizing the click locations via mental imagery, or actually logging into a system with the password. However, mere rehearsal by rote tends not to be successful for long-term retention. Instead, to store information in LTM the information must be processed in a deep and meaningful way that creates a strong memory (Craik and Lockhart, 1972). A poor encoding will hurt retrieval (e.g. failing to distinguish similar objects). Imagine an image showing five mountain peaks. If the user chooses a click location on one peak there is the possibility of confusing it at retrieval time with one of the other peaks, unless the encoding at storage time is strong.

3.2. Memory

Most existing graphical password systems can be classified as being based on either recognition or cued recall. Recognition involves identifying whether one has encountered an item before. In a graphical password system based on recognition, the user has to be able only to recognize previously seen images, not generate them unaided from memory. By contrast, pure recall is retrieval without external cues to aid memory, e.g. remembering a textual password that one has not written down. Pure recall is a harder memory task than recognition (Norman, 1988). However, there is an intermediary form of recollection between pure recall and pure recognition: cued recall. An example of cued recall is scanning an image to find previously chosen locations in it. Viewing the image cues the user about the locations. This is easier than having to recall something entirely from memory, but harder than simply recognizing whether a particular image has been seen before or not.

Decay and interference can disrupt memory. Decay is loss of the ability to retrieve items from memory over time. As discussed previously, rapid decay of memory occurs soon after learning, but the rate of loss declines over time and in the long run may become almost flat (Bahrick, 1984; Wixted and Ebbesen, 1991). The speed and amount of loss depends on the material being remembered, how well it was encoded, how much it was practiced, and whether it was retrieved and used after training (and therefore reinforced). With recognition memory, psychological studies show that images are recognized with extremely high accuracy, up to 98.5% after a 1 h delay, which is significantly higher than the accuracy of recognition of words or sentences (Shepard, 1967). Indeed, it has been found that participants made only 17% recognition errors after briefly viewing 10,000 successive images (Standing, 1973). With recall memory, studies also show that images are recalled better than words (Paivio et al., 1976). This well-known result is referred to as the "picture superiority effect" (Nelson et al., 1977). Studies of detection of changes in images also suggest that people retain relatively detailed visual information in memory for natural scenes (Hollingworth and Henderson, 2002). The content, organization, and affect of images also have been shown to have an effect on memorability (Mandler and Ritchey, 1977; Bradley et al., 1992).

3.3. *Efficiency*

Efficiency, and perception of efficiency, are important in password systems because users want quick access to systems. Time to input a highly practiced graphical password can be predicted by Fitts' Law, which states that the time to point to a target depends on the distance and size of the target (Fitts, 1954). Greater distance and smaller targets lead to slower performance. Existing evidence suggests that alphanumeric passwords may be faster to input than graphical passwords (Dhamija and Perrig, 2000). However, input time also needs to be considered in context. A user's subjective perception of time to complete a task may be more important than the actual time. A password input mechanism that users perceive as slow to use may lead to insecure password behavior. On the other hand, a password input mechanism that is perceived as engaging may be accepted by users even if it is rather slow to use.

4. Design of a graphical password system

4.1. *Existing graphical password schemes*

We categorize graphical password systems as based either on (1) recognition, (2) cued recall, or (3) pure recall. Here we focus on recognition and cued recall of images, rather than graphical systems based purely on unaided recall, such as "Draw-A-Secret" (Jermyn et al., 1999). In the first category, a few schemes have been devised, but they have strong limitations. Several systems based on image recognition, are Passfaces (Bensinger, undated; Brostoff and Sasse, 2000; Real User

Corporation, 2001), Déjà Vu (Dhamija, 2000; Dhamija and Perrig, 2000), and an image scheme devised by Weinshall and Kirkpatrick (2004). A graphical PIN approach has also been developed (De Angeli et al., 2002). To explain how recognition-based graphical passwords work, we describe Passfaces as it was used in Brostoff and Sasse's (2000) empirical study. To create a password, the user chooses four images of human faces from a fixed portfolio of faces. Those four faces become the user's password. In authentication, the user sees a grid of nine faces, consisting of one face previously chosen by the user and eight decoy faces. The user recognizes and clicks anywhere on the known face. This procedure is repeated with different target and decoy faces, for a total of four rounds. If the user chooses the correct four faces, he or she will be authenticated. Empirical evidence shows that Passfaces may be more memorable than alphanumeric passwords. Déjà Vu also works through recognizing several images in repeated rounds of selection (five rounds in this case), but uses abstract images and photographs rather than faces. The results of a small study of Déjà Vu suggest that choosing the images that will make up the password takes longer than choosing PINs and alphanumeric passwords, but the images are easier to remember over time. However, passwords based on image (e.g. face) recognition have a significant disadvantage. On each screen, only a small number of faces can be displayed, e.g. nine, one of which is a chosen "passface." An attacker has then a 1-in-9 chance of guessing this passface. To reduce the chance of an attacker guessing the password, the login process requires several rounds of face recognition. If four rounds are used the chance of guessing the password, i.e. all the passfaces, is $(1/9)^4 = 1.5 \times 10^{-4}$. Thus, in a few thousand random guesses the attacker would be likely to find the password. To obtain security similar to that of an eight-character alphanumeric password (over an alphabet of 64 characters), 15 or 16 rounds with nine faces each would be required. This would make the login slow and tedious, and most likely, the login would also be perceived by the user as slow and tedious. Research has also shown that recognition-based passwords consisting of faces may yield passwords with very low entropy because of predictable ways that users choose the faces that make up their password (Davis et al., 2004). Finally, image recognition schemes will not work well on small screens (e.g. palmtops) because of the requirement to simultaneously show multiple images.

Blonder's original idea for a graphical password fits into the second category of graphical password schemes, namely those based on cued recall. In this scheme, a user clicks on several previously chosen locations in an image to authenticate him/herself. The company Passlogix (Boroditsky, 2002) later developed an implementation of Blonder's scheme. In the Passlogix implementation an image has predefined click objects, outlined with thick boundaries. The user chooses several objects as the password. For example, in a cartoon-like image of a kitchen with 20 vegetables and 20 pots, the user might drag some vegetables to certain pots to create a password; the password thus consists of a sequence of several vegetable-pot pairs. In authentication the user has to repeat this process. Another example is images (again, cartoon-like, with predefined click regions or click objects with thick boundaries), in which the user clicks on a few chosen objects; to log in, the user has to click again on the same objects. The drawback of this scheme is that the number of available click regions is small, perhaps only dozens in an image; thus a password has to be rather long to be secure. (For example, with 20 vegetables and 20 pots, a secure password would need at least six vegetable-pot pairs, i.e. 12 clicks would be needed.) Another drawback of this scheme is the need for predefined click regions that are readily identifiable; this requires artificial, cartoon-like images rather than complex, real-world scenes.

It should also be noted that all graphical password schemes, including our scheme introduced below, have some inherent weaknesses. First, people with poor vision will have difficulties using graphical passwords. Second, people who have poor motor control and experience difficulties using pointing devices may not be able to use graphical passwords effectively. Third, people with various kinds of color blindness will see color differently. This may or may not affect their ability to use graphical passwords.

4.2. Passpoints: a new graphical password scheme

We developed a graphical password scheme that is similar to Blonder's scheme but that overcomes some of its main limitations. Our scheme is flexible because it allows any image to be used, e.g. natural images, paintings, etc. The images could be provided by the system or chosen by the user. The only practical requirement is that the image be intricate and rich enough so that many possible click points are available. Another source of flexibility is that we do not need artificial predefined click regions with well-marked boundaries. A user's password consists of any arbitrarily chosen sequence of points in the image. Since an intricate image easily has hundreds of memorable points, not many click points are needed to make a password hard to guess. For example, with five or six click points one can make more passwords than 8-character Unix-style alphanumeric passwords over a standard 64-character alphabet. In order to log in, the user has to click close to the chosen click points, within some (adjustable) tolerance distance, e.g. within .25 cm from the actual click point.

For security reasons, the system should not store passwords explicitly, "in the clear," but in hashed form.[1] At first, it seems that our graphical passwords cannot be hashed; indeed, when users log in they click close to their chosen click points, but not exactly on the chosen points. So, at the pixel level, the password that is entered changes all the time. One reason why Blonder's scheme had well delimited, predefined click regions was to avoid this variability. Moreover, hashing does not allow approximation: two passwords that are almost (but not entirely) identical will be hashed very differently. A first step towards solving the hashing problem is to discretize the image into squares that are large enough so that we can expect users to always hit the same square if they want to. However, this still leaves the possibility that a user may choose a click point that happens to be close to an edge of a

[1] The system stores a transformed ("hashed," "encrypted") version of the password in the password file. A crucial property of the hash transformation is that it is not feasible to reconstruct the actual password from the hashed version. When a user logs in, the system hashes the password entered by the user, and compares this with the stored version.

discretization square (i.e. the tolerance around the click point). We do not want to display the discretization grid to the user (because that would be ugly and would limit the user's freedom of choice for click points). The problem of the edges can be solved by using three discretization grids simultaneously. The details of this method are explained in Birget et al. (2003). In short, every possible click point is within a safe distance from the edges of at least one of the three discretization grids; at password creation, for each click point the system remembers the safe grid (one out of three) for that click, and it hashes the location of the click square relative to that grid. Finding a discretization of images that allows hashing of the graphical passwords, without inconveniencing the user, is one of the main achievements of our design.

In Table 1 we compare the password spaces of PassPoints with that of alphanumeric passwords, for various parameter settings. The "password space" is the set of all passwords that are possible for a given password scheme and for a given setting of parameters. For example, for alphanumeric passwords of length eight over a 64-character alphabet, the number of possible passwords is $64^8 = 2.8 \times 10^{14}$. In our graphical password scheme, as we tested it in our empirical study, with image size 451×331 and grid square size 20×20 (all measured in pixels) there are about $451 \times 331/20 \times 20 = 373$ grid squares; hence, for passwords consisting of five click points, the password space has size $373^5 = 7.2 \times 10^{12}$. With the same settings, but with six click points, the password space has size $373^6 = 2.69 \times 10^{15}$. If in our graphical password scheme the image size is 1024×752 (roughly the full screen), with grid square size still 20×20 (all measured in pixels), and with passwords consisting of five clicks, the password space will have size 2.6×10^{16}. If we assume that in the latter setting only half of the area of the image is used (because the other half of the area has no memorable features to click on), the password space will have size 3.2×10^{15}. Obviously, the password space should be large for a password scheme to be secure. However, the size of the password space is not the only thing that matters. The usability and memorability of passwords are just as important.

We see from these comparisons that for just five clicks, and for reasonably sized grid squares, graphical passwords have a larger password space than alphanumeric

Table 1
Comparison of password space for alphanumeric passwords and PassPoints with different parameters

	Image size	Grid square size (pixels)	Alphabet size/ No. squares	Length/No. click points	Password space size
Alphanumeric	N/A	N/A	64	8	2.8×10^{14}
Alphanumeric	N/A	N/A	72	8	7.2×10^{14}
Alphanumeric	N/A	N/A	96	8	7.2×10^{15}
Graphical	451×331	20×20	373	5	7.2×10^{12}
Graphical	1024×752	20×20	1925	5	2.6×10^{16}
Graphical	1024×752	14×14	3928	5	9.3×10^{17}
Graphical (1/2 screen used)	1024×752	14×14	1964	5	2.9×10^{16}

passwords. Moreover, one could easily increase the number of click points to six or decrease the grid square (e.g. from 20×20 to 14×14). Indeed, with around 10 click points our password space is comparable in size to a cryptographic keyspace. Thus, another contribution of this design is a large password space with a small number of clicks.

For our experiment we used a smallish image of 451×331 pixels, because we needed room for buttons and text on the side of the image; thus we had to restrict the size of our password space, for the sake of the experiment. Even in that limited setting our password space is close in size to the space of random alphanumeric passwords of length eight, over a standard 64-character alphabet.[2]

4.3. Methodology

4.3.1. Experimental design

The longitudinal experiment involved a between-subjects design with two conditions and multiple sessions covering creating, learning and retaining passwords across six weeks. The between-subjects factor was the kind of password, with two conditions: alphanumeric and graphical. There were three within-subjects sessions in the experiment. The sessions occurred in week 1, week 2, and week 6. In week 1 the participants first created a password, then learned the password by entering it multiple times. In the password creation phase we measured the number of attempts and amount of time to create a valid password. In the learning phase we measured the number of attempts and amount of time to meet a fixed learning criterion. After the learning phase, the participants filled out a questionnaire about their perceptions of the password system they used. Finally, at the end of the first session, participants carried out their first retention trial. The second session took place in Week 2; this was the participants' second retention trail. The third session in Week 6 was the final retention trial. In the retention trials we measured the number of attempts and amount of time for participants to enter a valid password.

4.3.2. Participants

For our password study we targeted a population of experienced computer users. The participants were 40 members of a university community, including staff, students, and faculty. The highest educational level achieved varied from some post-secondary education to a Ph.D. Most of the participants had a Bachelor's degree. The mean age of the participants was 32.9 years ($SD = 10.9$), and the range was from 20 to 55 years. The sample included 23 males and 17 females. All of the participants used PCs regularly.

[2] The 64 characters are the 10 digits, 26 lower-case letters, 26 upper-case letters, underscore, and dot. For portability reasons we leave out other punctuation characters and symbols. Table 1 considers also a 96-character alphabet with added symbols and punctuation characters. In any case, we do not consider characters that are not on the keyboard.

4.3.3. Materials

The system included both graphical and alphanumeric password interfaces. An implementation of the PassPoints system was developed that was instrumented for collecting data. The graphical password interface included the picture used for testing and several buttons (Fig. 1). A single picture was used in this experiment, a scene of a hotel swimming pool with people moving around it. The picture was colorful with many elements that could serve as memorable click points, including people, objects (trees, tables, chairs, canopies, bins, clothing, hats, bags) and architectural elements (columns, beams, windows). The size of the picture was 451 × 331 pixels. The grid square, or tolerance, around a click point was set to 20 × 20 pixels; on the PC monitor this equated to a square of area .53 cm^2. The picture occupied about half of the full screen. The rest of the screen consisted of a background with five buttons and with empty space to present instructions in certain phases of the experiment. The "Submit" button was used to submit the password when the user had entered it. The "Undo" and "Clear" buttons were used to correct a password before it was submitted; the "Clear" button erased all password points input so far; the "Undo" button erased only the user's most recent point. The "See My Password" button allowed the user to view his or her password during the learning phase and under certain circumstances in the following retention phase. The "Quit" button allowed a user to quit the experiment. This button was placed to the side to avoid inadvertent quitting. All instructions for the participants were given on the screen and feedback on correctness of a password input was given on screen when the user clicked the "Submit" button. The on-line testing system also included a questionnaire that asked the user's perceptions of the password system. The questions were answered on a 7-point Likert-type scale ranging from strongly agree (1) to strongly disagree (7).

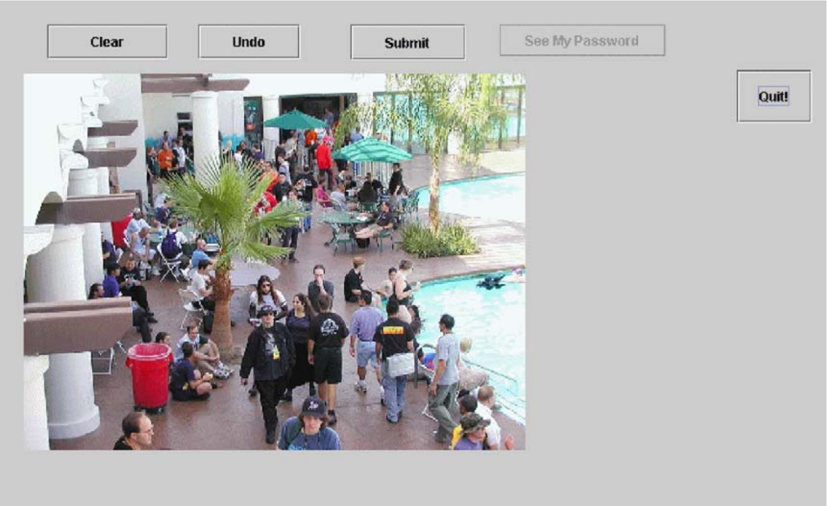

Fig. 1. Graphical password interface used in experiment.

In the alphanumeric interface a typical password input field was shown in which the user typed a password. The buttons were "Clear," "Submit," "Show My Password," and "Quit." The "Undo" button was omitted because users deleted characters using the backspace. The questionnaire was the same as the graphical password questionnaire except for small wording changes in a few questions.

4.3.4. Procedure

A single computer, a PentiumIV PC with a high-resolution 19-inch monitor, was used in the experiment. Testing was done individually in a quiet room. Participants were randomly assigned to the graphical or alphanumeric condition. Each individual participated in three sessions. The first session lasted about 35 min. The participants were introduced to the purposes and procedures of the experiment by viewing a 5 min PowerPoint presentation. On entering their assigned password system, the participant entered demographic data.

In the password creation phase the participant was given instructions on the screen to create a password. Graphical password users had to select and enter five distinct points on the picture with no point within the tolerance around any other chosen point. A graphical password of five points was used based on the analysis above, which shows that in terms of security five click points provide a password space about as large as or larger than an alphanumeric password of eight characters. Participants were told that they would have to remember the points and the order in which they were input. When the participant chose a point the tolerance around it was outlined on the screen with a number from 1 to 5 to indicate which ordered click it was; this showed the participant its size and helped the individual avoid choosing the next point too close to the previous one or choosing a password with too few or too many clicks. When the participant clicked the "Submit" button, the system gave feedback on whether the password was valid. If it was invalid, the participant was instructed to try to create a password again. When the user had created a valid password, the entire password was displayed as feedback to the user before going on to the learning phase (Fig. 2). The image was shown to the user outlining the tolerance of 20×20 pixels around each point chosen. This was done to give the users feedback about the location and tolerance around the click points. The points were also numbered 1–5 to indicate their order of input. Note that the feedback mechanisms used in the experiment would be feasible in a real system only if the user created the password in a private place. Otherwise, attackers might be able to steal the user's password.

To create a password, alphanumeric users had to enter eight characters including at least one upper-case letter and one numeric character (a 62-character alphabet since no special characters were used). They were also told not to choose a password they had already used or a variation of a password they had used. As the participant typed each character a dot was shown to help the individual avoid entering too few or too many characters. The system enforced that the user re-enter the password until the conditions for a valid password were met. The 8-character password was displayed to the user in clear text as feedback before going on to the learning phase.

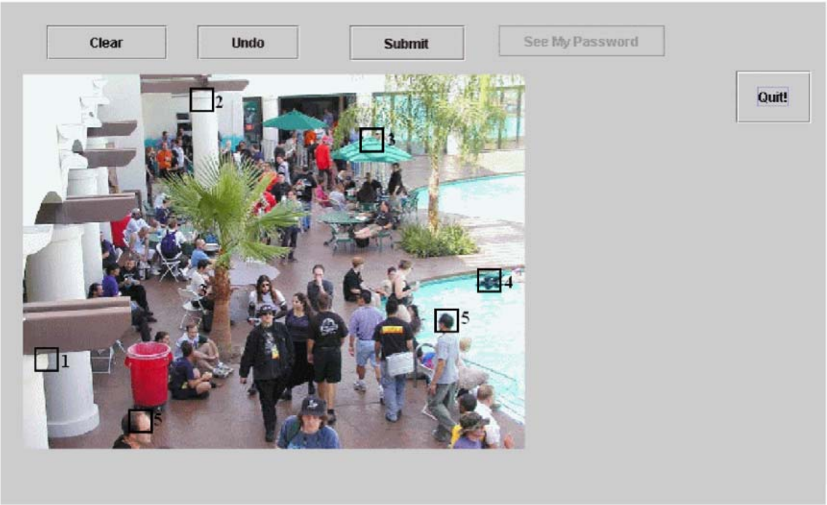

Fig. 2. Example of participant password with tolerance and click order displayed.

As with the graphical password, this was for the purpose of learning in the experimental setting.

When the user had created a valid password, the learning phase began. To reinforce the password the user entered the password repeatedly until he or she achieved ten correct password inputs. Users received binary feedback on the correctness of each password input and could see an on-screen count of how many correct and incorrect entries they had made. If the user was not able to remember the password, he or she could click on “Show My Password,” which displayed the password (either the alphanumeric string or the image with the password points shown as in Fig. 2). After the learning phase, the participant filled out the questionnaire on-line. This was designed to gather user perceptions and act as a distractor between the learning phase and the first retention trial.

In the retention phase, password retention was measured longitudinally three times. At the end of the first session, the user had to enter the password correctly one time. This served as the first retention trial, R1. The following retention trials, R2 and R3, took place in week 2 and week 6, respectively. The retention trials took about 5 min. The participant was finished as soon as the correct password was entered. If the participant entered an incorrect password, the system gave feedback that the password was wrong, and the user was instructed to re-enter the password. If the user failed to input the password correctly after five attempts, the “Show My Password” button was enabled and the user could view the password, if desired, then make another attempt to input it. After the last retention trial, R3, the user again filled out a questionnaire as in the first session and wrote answers to five open-ended questions.

5. Results

The results are reported below by experimental phases: password creation, learning, and retention. The relevant questionnaire results are also reported for each phase. Results of the open-ended questions are not reported here.

5.1. Creation phase

In the creation phase participants repeatedly created and submitted passwords until they succeeded in creating a valid password. The total number of attempts required to create a valid password was measured. The amount of time to create a valid password (including all attempts) was also measured. The means and standard deviations of these two measures are shown in Table 2.

Both measures were analysed using a *t*-test. For total attempts to create a valid password the results showed that the graphical group took significantly fewer attempts: $t(38) = 3.13, p < .005$. In the alphanumeric group 11 of 20 participants took two or more attempts to create a valid password. By contrast, in the graphical group 19 of 20 participants created a valid password on their first attempt. The graphical group took less total time to create a valid password than did the alphanumeric group, but the difference did not reach significance: $t(38) = 1.79, p < .083$.

In the questionnaire participants were asked two questions about the password creation, as shown in Table 3. Recall that the scale ran from 1 to 7 *with lower numbers indicating stronger agreement*. The questions were analysed using the nonparametric Mann–Whitney *U* test. For the first question there was a significant difference ($U = 127.00, p < .043$), with the graphical subjects agreeing more strongly that they did not have trouble choosing a password. For the second question there was no significant difference.

5.2. Learning phase

In the learning phase participants practiced inputting their password in order to remember it. They input their password repeatedly until they had achieved 10 correct

Table 2
Means (standard deviations) of total attempts to create a valid password and total time to create a valid password (alphanumeric $N = 20$/graphical $N = 20$)

	Mode	Mean (SD)
Total attempts to create	Alphanumeric	1.70 (.18)
	Graphical	1.10 (.07)
Total time to create (seconds)	Alphanumeric	81.10 (36.50)
	Graphical	64.03 (21.93)

Table 3
Means (standard deviations) of questions about the creation phase (alphanumeric $N = 20$/graphical $N = 20$)

	Mode	Mean (SD)
I did not have much trouble thinking up a password	Alphanumeric	3.30 (1.59)
	Graphical	2.35 (1.57)
It did not take me long to think up a password	Alphanumeric	3.15 (1.63)
	Graphical	2.60 (1.42)

Table 4
Means (standard deviations) of number of incorrect submissions and total practice time in the learning phase (alphanumeric $N = 20$/graphical $N = 20$)

	Mode	Mean (SD)
Number of incorrect submissions	Alphanumeric	.40 (.68)
	Graphical	4.80 (7.16)
Total practice time (seconds)	Alphanumeric	66.08 (4.92)
	Graphical	171.89 (24.46)

logins. The number of incorrect password submissions was measured and the total time spent on practicing, including correct and incorrect inputs. Table 4 shows the means and standard deviations.

The two measures were analysed using *t*-tests. There were significant differences in favor of the alphanumeric group in both cases: number of incorrect inputs $t(38) = -.2.73, p < .013$; total practice time $t(38) = -4.24, p < .0001$. Graphical participants had much more variability in their practice (Fig. 3). Of the 20 alphanumeric participants 14 had no incorrect inputs, four had one incorrect input, and two had two incorrect inputs. In the graphical group eight had no incorrect inputs, four had one incorrect input, and one had two incorrect inputs. The remaining seven graphical participants had more than three incorrect inputs. Four of these participants were extreme with 17–20 incorrect inputs. We reanalysed the date removing these four outliers and found that the difference in the number of incorrect inputs was not significant ($t(34) = -1.85, p < .113$), but the total practice time remained significantly longer for the graphical group ($t(34) = -4.58, p < .0001$).

In the questionnaire, participants were asked five questions relevant to password learning (Table 5). There was a significant difference on the first question (Mann–Whitney $U = 98.50, p < .004$), with the alphanumeric subjects agreeing more strongly that it did not take them long to complete the password inputs. There were no significant differences in the other questions.

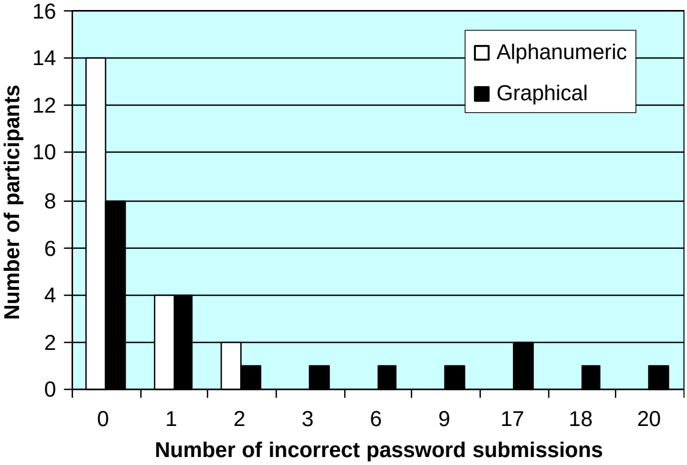

Fig. 3. Incorrect password submissions by participant in the learning phase.

Table 5
Means (standard deviations) of questions about the learning phase (alphanumeric $N = 20$/graphical $N = 20$)

	Mode	Mean (SD)
It did not take me long to input my password 10 times	Alphanumeric	1.65 (.67)
	Graphical	3.40 (2.14)
Once I created my password I was able to input it correctly	Alphanumeric	1.65 (.79)
	Graphical	2.60 (1.82)
My password input got better with practice	Alphanumeric	1.20 (.52)
	Graphical	1.15 (.50)
Inputting my password was easy	Alphanumeric	1.90 (1.02)
	Graphical	2.70 (2.18)
Inputting my password was fast	Alphanumerical	2.35 (1.14)
	Graphical	3.05 (1.73)

5.3. Retention phase

In the retention phase participants input the password longitudinally three times: at the end of the first session in week 1, in week 2, and in week 6. Recall that in each retention session the participant had only to enter his or her password correctly one time. The number of incorrect password submissions and time for the correct submission were measured. See Table 6 for the means and standard deviations.

A two-way, mixed mode ANOVA was used for the analysis of the two measures in Table 6. In each case the within-subjects factor was retention trial (R1/R2/R3) and

Table 6
Means (standard deviations) of number of incorrect password submissions and time for correct submission (s) in the retention phase (alphanumeric $N = 20$/graphical $N = 20$)

	Mode	Mean R1 (SD)	Mean R2 (SD)	Mean R3 (SD)
Number of incorrect submissions	Alphanumeric	.25 (.79)	2.20 (2.73)	1.75 (2.47)
	Graphical	1.55 (1.57)	2.75 (3.88)	1.50 (2.80)
Time for correct submission (s)	Alphanumeric	5.23 (1.66)	9.42 (3.70)	9.24 (3.72)
	Graphical	8.78 (4.40)	24.25 (15.21)	19.38 (17.57)

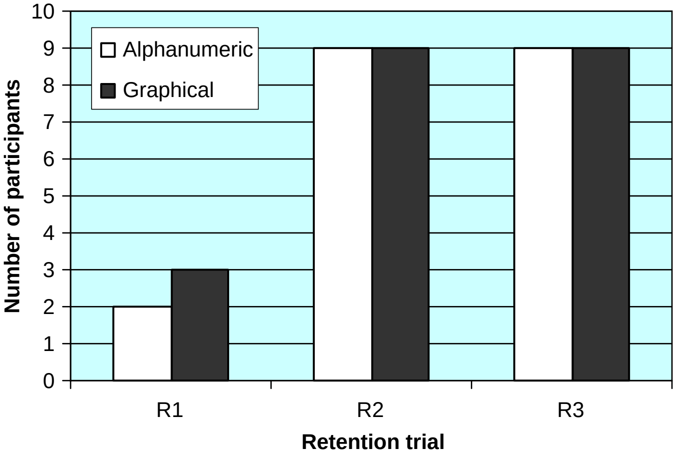

Fig. 4. Number of participants who failed to submit a valid password on the first attempt.

the between-subjects factor was mode (alphanumeric/graphical). Tukey's HSD was used for post hoc tests. For the number of incorrect submissions, the effect of retention trial was significant ($F(2, 76) = 6.907, p < .002$). The effect of mode was not significant. The interaction between retention trial and mode also was not significant. A post hoc test for differences between the retention trials showed that the only significant difference was between R1 and R2 ($p < .001$), with R1 taking fewer trials than R2. Fig. 4 shows the number of participants who failed to submit a valid password on their first attempt for each retention trial.

We also show in Fig. 5 the number of participants who failed to "log in" by the criterion of allowing a maximum of five attempts. Many password systems limit the number of attempts to some similar small number. We enabled the "See My Password" button after five failed attempts; all participants who had not succeeded after five attempts did ultimately view their password via the "See My Password" button.

For the total time for the correct submission, the effect of retention trial was significant ($F(2, 76) = 21.67, p < .0001$). The effect of mode was also significant

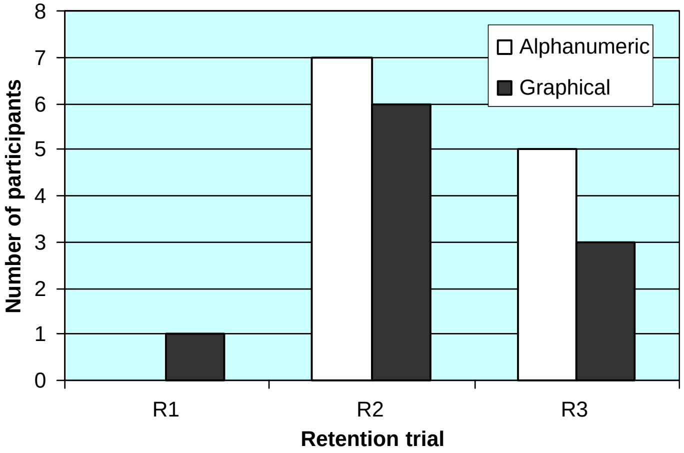

Fig. 5. Number of participants in each retention trial who failed to submit valid password after five attempts. Note that these participants subsequently viewed their password.

Table 7
Means (standard deviations) of questions about the retention phase (alphanumeric $N = 20$/graphical $N = 20$)

	Mode	Mean (SD)
Inputting my password was easy	Alphanumeric	1.90 (1.02)
	Graphical	2.70 (2.18)
Inputting my password was fast	Alphanumeric	2.35 (1.14)
	Graphical	3.05 (1.73)
I think the password system was pleasant to use	Alphanumeric	3.00 (1.34)
	Graphical	2.40 (1.57)
I think that the rules [about password creation] make it easy to remember the password	Alphanumeric	5.25 (1.71)
	Graphical	3.55 (2.09)

(1,38) = 39.24, $p < .04001$) with the graphical group taking longer times. There was a significant interaction between retention trial and mode ($F(2, 76) = 6.67, p < .004$). Tukey's HSD showed that there were significant time differences between R1 and R2 and between R1 and R3, but no significant difference between R2 and R3. A post hoc Newman–Keuls test for the interaction indicated that in the R2 trial the graphical time was highly elevated with respect to the other conditions ($p < .05$).

Results from the questionnaire at the end of the final retention trial are shown in Table 7. There were no significant differences between the alphanumeric and graphical groups on perceived ease of inputting the password, speed of entering the password, and pleasantness of the password system. Graphical password users agreed more strongly than alphanumeric users that the rules for creating a password made it easy to remember the password.

6. Discussion

PassPoints has the security advantage of a large password space over alphanumeric passwords, as shown in Table 1. It also has an advantage in password space over Blonder-style graphical passwords and recognition-based graphical password, such as Passfaces. However, usability is also a critical consideration. Our results of testing the prototype PassPoints system in comparison to alphanumeric passwords yielded mixed results.

The graphical group found creating a password easy. Users created passwords quickly (an average of 64 s) and only one out of 20 participants had to repeat the password creation because of failure to create a valid password in their initial attempt. Alphanumeric password users had more difficulty. They took somewhat longer (an average of 81 s) than the graphical users. Eleven of the 20 alphanumeric users were not successful in their first attempt to create a valid password. Users' perceptions were consistent with these results: the graphical group reported significantly stronger agreement that it was easy to create a password. The rules for creation of a graphical password were clear and simple for users. The only way they could go wrong was by choosing too many or too few points or by choosing points within the tolerance around another point. Alphanumeric subjects had three rules that were also straight-forward and consistent with typical recommendations for password choice: the password had to consist of eight characters, with at least one uppercase character and one numeric character. Although the alphanumeric rules were not highly restrictive and were displayed while the participants were creating their passwords, over half failed initially. Furthermore, even in week 6 the alphanumeric group still reported that the rules about password creation made it hard to remember the password. This suggests that the alphanumeric participants were not accustomed to selecting passwords with a view toward security.

In the learning phase the alphanumeric group took fewer trials to achieve 10 correct password inputs than did the graphical group. This is also reflected in significantly longer total times to input the graphical passwords. Seventy percent of the alphanumeric participants input the password 10 times without any errors, and all alphanumeric participants were able to achieve the criterion with a maximum of two incorrect password inputs. The graphical group took more trials and had more variability. Forty percent of graphical participants achieved input of the password 10 times without any errors, and 70% achieved the criterion with a maximum of three incorrect password inputs. However, strikingly, the least successful 20% of the group made between 17 and 20 incorrect password inputs.

It is not surprising that the alphanumeric group achieved the practice more easily than the graphical group. Although the alphanumeric group had some difficulty in creating a valid password, inputting an alphanumeric password was a highly familiar process. The time between viewing the display of the valid password at the end of the creation phase and inputting the first password input in the learning phase was a matter of less than 30 s. Most participants were able to retain the password and input it correctly. The few users who forgot the password were able to see it by clicking "Show my Password." In sum, the password learning phase for the alphanumeric users was familiar and easy to carry out. It involved repetitive inputs to reinforce the password, without any other new concepts or procedures to learn.

For the graphical group the learning was much harder, probably because graphical passwords were completely new both conceptually and procedurally. There were several sources of error. Occasionally participants skipped one password point, inputting four rather than five points. Also, they occasionally input points in the wrong order. However, this was fairly rare because most participants created their password to follow a certain trajectory of points across the screen (e.g. an arc) and thereby were able to remember the order of input. The most common problem in graphical password input was clicking close to, but outside, the tolerance around the click point. The 20 × 20 pixel tolerance yielded a square of .53 cm^2. Although at the end of the creation phase participants had viewed their click points with the grid boxes representing the tolerance outlined on the image, they still had difficulty being as precise as required. They input their points slowly, so this was not a phenomenon of working carelessly. Rather, it appears that in password creation participants focused on the general neighborhood of their click points rather than the close surrounds. For example, an individual might have remembered that his or her click point was on a certain hat, rather then on the crown of the hat. During the learning trials participants had to learn their click points in a precise way. Most were able to do this by a few extra trials, sometimes clicking on "Show my Password" once or twice. Those who took the highest number of trials tended to have interacting errors, e.g. not clicking precisely enough, then in the next trial correcting the precision problem but, in their concentration on the precision, omitting a point. As in any real password system, the feedback was whether the entire password input was correct or not. The feedback given did not pinpoint in what way(s) the password was wrong. Another complication for participants who had serious difficulties was that, in the concerted effort to input the password accurately, some participants seemed to forget that they could use the "Show my Password" button for help. Basically, a large majority of graphical users had little difficulty in the learning phase, but 20% had substantial problems.

In spite of the difficulties of some in the graphical group, most of the perceptions, reported after the learning phase, were similar across the two groups. The graphical group did agree significantly more than the alphanumeric group that the 10 password inputs took a long time. However, the groups were similar in their perceptions that they were able to enter the password correctly and that they had improved with practice. Interestingly, the two groups did not differ significantly on questions about how easy and how fast it was to input a password. This reflects that only a minority of graphical participants had serious difficulties in learning.

In the retention phase, the correctness of password inputs differed strongly by trial for both groups. Since the R1 trial took place in the same session as the creation and learning phases, there were relatively few bad inputs, given the short time the participants had to retain the password. However, in the R2 trial 1 week later participants had more difficulty recalling their passwords. Interestingly, in the final R3 trial at the end of the experiment there appears to have been some consolidation of the passwords in memory because the incorrect inputs were lower than in R2, though not significantly, in spite of the long time lapse. The lack of significant differences between the alphanumeric and graphical modes on the correctness of password inputs and lack of interactions between mode and trial, indicate that the main factor in correctness was password memory for both groups.

The time for the correct input of the password showed that the alphanumeric group was faster across the board. A longer input time is expected for the graphical group based on the time for mouse movement and selection of the target (Fitts, 1954), but the large difference between the two groups suggests strongly that the time difference was not due primarily to the mechanics of movement and selection, but to the think time to locate the correct click region and determine precisely where to click. Among the three sessions, the fastest input was in R1, a result that was expected and corresponded to the high correctness of inputs in R1. The increased time for the correct password input in R2 and R3 was substantially elevated over R1 for both alphanumeric and graphical groups. Like the correctness of inputs, this shows the effect of intermittent use on memory for passwords. In addition, the graphical group in R2 was by far the slowest, indicating that participants proceeded slowly and carefully to ensure correct recall and input. Interestingly, in the graphical group the R3 trial at the end of six weeks was a bit faster (though not significantly) than the corresponding R2 trial. This suggests that the graphical participants were becoming familiar enough with their graphical passwords and input procedures to work more quickly.

The general perceptions of the alphanumeric and graphical groups were similar on the ease and speed of entering the password and on the pleasantness of using the system. This is a bit surprising with respect to the speed issue because, in fact, the speed for correct inputs was substantially slower in the graphical group, particularly in the R2 retention trial. A possible explanation is that the graphical group was intrigued enough with the system to lessen their perception of the lapse of time. In fact, during our debriefing most graphical participants expressed enthusiasm for the graphical password system.

7. Conclusion

In conclusion, the empirical testing of PassPoints indicates strengths and weaknesses. Graphical password users were able to easily and quickly create a valid password, but they had more difficulty learning their passwords than alphanumeric users, taking more trials and more time to complete the practice. Graphical users' retention of their password over the course of six weeks was similar to alphanumeric users, but the graphical users continued to take more time to input their passwords than alphanumeric users. Graphical users had similar perceptions to alphanumeric users in terms of ease and speed of input and pleasantness of their password system.

Our testing was affected by the fact that we were comparing a totally new graphical password concept and procedure to the extremely well known and practiced alphanumeric password. Almost by definition the playing field was not entirely level. Even with instructions, one-fifth of graphical participants faced difficulties inputting their password in the learning phase. The main difficulty was in understanding the precision requirement and attending to the tolerance around the click point during password creation. Users who did this had little difficulty in entering their password correctly. For security reasons it is probably not feasible to increase the size of the tolerance around click points, so emphasis on precision is important in training, as is practice inputting the password. Preliminary results of a study we are currently carrying out suggest that users can retain the location of click points with even smaller tolerances than used here, if they attend precisely to the tolerance around their click point during training. New users of graphical passwords probably will benefit from a self-training module for password creation and practice. A simple training module could be something like our procedure of creating a password and inputting it multiple times to reinforce it, with the ability to change the password if the user had difficulty accurately recognizing his or her click points. To reinforce precise memory of the click points it might be useful during practice to show users how closely their clicks were located with respect to their actual password points.

An encouraging outcome is that, after the learning phase, the graphical users were as correct in using their password as the alphanumeric group. This result is notable because it was achieved in very intermittent use and with very little experience with graphical passwords. We expect that in practice users of graphical passwords will equal or exceed alphanumeric password users, given more experience with graphical passwords and the opportunity to use their graphical password regularly for some period of time.

A clear problem is that the graphical group had quite elevated input times in the week 2 and week 6 retention trials. The alphanumeric group also had much higher times in these last two retention trials compared to their first retention trial. For the alphanumeric group we assume that their extra time is explained entirely by think time to remember their password, since they were highly experienced in the procedure for typing in and submitting a password. We believe that the slow input in the graphical password group actually has two sources: the think time to scan the picture and recover the password *and* their low experience actually inputting graphical passwords. More work is needed to determine the contribution of these two sources.

In addition to the current work mentioned above, we are exploring other HCI aspects of our graphical passwords. First, we believe that it is important to study user interactions with PassPoints in everyday use over time, in order to get a realistic sense of how our graphical password scheme works for people. The current study focused on initial learning and retention given gaps in use, but did not investigate regular use in a normal environment. Such a field study is needed to understand the learning curve over time, as well as the perceptions of people who use graphical passwords regularly. A field study might be carried out over several months in a normal computer laboratory where users have to enter graphical passwords to log in to various systems or to unlock a screen saver.

A second area of interest is using PassPoints on small screens. In principle, this could be an ideal solution for small-screen devices because users can easily tap on the image with a stylus, as opposed to the difficulty of typing a long, secure alphanumeric password on a very tiny keyboard. The challenge is to provide a large enough password space on a small image. However, this might be handled by magnification of an area of the image when the user moves the stylus to a given area of the screen. We have an early prototype of such a graphical password solution on a PDA.

Third, an important issue to explore is the retention of graphical and alphanumeric passwords when users have multiple passwords. Psychological research indicates that interference can cause significant memory problems (Wixted, 2004). Security research (Adams and Sasse, 1999; Ives et al., 2004) confirms that users have difficulty remembering multiple passwords and develop unsafe practices to overcome the problem (writing passwords down, etc.). A question to investigate about graphical passwords is whether interference occurs as much with graphical passwords as with alphanumeric passwords, or indeed whether it occurs more. Multiple graphical passwords could be created using one image or multiple images. We hypothesize that using the same image for multiple passwords (e.g. two different sets of password points on one image for logging into two different systems) would be likely to lead to interference, because it would be difficult to associate the correct set of points with a specific system. It is also reasonable to speculate that the content of the image itself may have an important effect on interference, i.e. if there are similar objects in an image they may become confusable when used in different passwords. Using a different image for each password might lead to less interference, but it raises another problem of remembering which image corresponds to which system. As discussed previously, password memory problems happen most when the use is intermittent, and that applies to the problem of interference also. We are currently carrying out an experiment that compares interference longitudinally in alphanumeric passwords, graphical passwords using the same image, and graphical passwords using different images.

Fourth, in graphical passwords there are important questions about whether the full range of pixels in an image will be used. If the areas in which users realistically click are limited, this reduces the entropy of passwords and makes it easier for an attacker to guess a password. (The same is true in alphanumeric passwords because users do not choose all possible passwords in their password space.) In PassPoints we do not assume that the full range of points in an image will be used. In most images there are undifferentiated areas that are not good targets for a memorable password point, e.g. an area of an image showing dense foliage. To be conservative in our password space calculations we considered the case that only half the pixels are likely to be reasonable click points (Table 1, last row). To increase the memorable click points, care must be taken in choosing password images to eliminate those that have large uniform areas. However, the issue of users' distribution of password points does not end here. Davis et al. (2004) report that when human faces are used as password images the entropy is low. Essentially, users tend to choose faces that are like themselves, e.g. faces of people of the same ethnicity. Our password images are not limited to faces or humans; they can be a very wide range of scenes. However, if there are "hot spots" in the image that attract many users, the entropy problem can occur. Logically, it seems that many users may be attracted to incongruous or unexpected elements in an image. Indeed, this is confirmed by a study of Hollingsworth et al. (2001), in which an image of a chemistry laboratory was shown with a teddy bear sitting on a laboratory bench. Participants attended to the teddy bear more than to expected objects, such as a microscope. In addition to incongruity, graphical password users may also be attracted to other kinds of salient objects in an image; salience could be the result of size, placement, or color. The image used in this study did not have any obvious incongruities, but it appears that there were a few areas that attracted people to click, the large red bin in the foreground (gray in this paper, Fig. 1) being the most notable with 10 clicks over the whole bin. (But note that with our robust discretization technique the bin was discretized into several grid squares and only a few clicks would have been considered the "same" point.) With only 20 participants we cannot make any strong conclusions, but it seems important to further study this issue. We are planning to collect a large number of password click points on multiple images to learn more about click point distributions.

In conclusion, the contributions of this work fall in two areas: security and usability. With respect to security, stringent rules for alphanumeric passwords lead to poor password practices and lower overall security. However, PassPoints seems to hold out the prospect of a much more secure system. First of all, it is easy to obtain large passwords spaces, based on intricate, natural images with hundreds of potential click points. Second, we developed a "robust discretization" which enables the system to cryptographically hash PassPoints passwords; this allows safe storage and protection during file back-up. Third, it appears from the small sample in our experiment that users did not too often chose points that were within a grid square chosen by another individual. Moreover, the discretization into small grid squares reduces the possibility of users having the same click points even if they are attracted to a salient area of the image. This reduces the chance of an attacker being able to guess passwords. With respect to usability, which is the main focus of this paper, our results indicate first that users can quickly and easily create graphical password. Second, most users can quickly learn to input graphical passwords without error, and even those who do make repeated errors can input them successfully given enough practice. Third, users remember graphical passwords as well as alphanumeric passwords over weeks without use. These observations about security and usability motivate further study of the "usable security" of PassPoints and of graphical passwords in general.

Acknowledgments

This work was supported in part by NSF grants CCR-0310490, CCR-0310793, and CCR-0310159.

References

Adams, A., Sasse, M.A., 1999. Users are not the enemy. Communications of the ACM 42 (12), 41–46.

Bahrick, H.P., 1984. Semantic memory content in permastore: fifty years of memory for Spanish learned in school. Journal of Verbal Learning and Verbal Behavior 14, 1–24.

Bensinger, D., undated. Human memory and the graphical password. http://www.activetechs.com/solutions/security/sso/bensinger.pdf. Accessed January 14, 2005.

Biederman, I., Glass, A.L., Stacy, E.W., 1973. Searching for objects in real world scenes. Journal of Experimental Psychology 97, 22–27.

Birget, J.C., Hong, D., Memon, N., 2003. Robust discretization, with an application to graphical passwords. Cryptology ePrint Archive http://eprint.iacr.org/2003/168. Accessed January 17, 2005.

Blonder, G.E., 1996. Graphical passwords. United States Patent 5559961.

Boroditsky, M., 2002. Passlogix password schemes. http://www.passlogix.com. Accessed December 2, 2002.

Bradley, M.M., Grenwald, M.K., Petry, M.C., Lang, P.J., 1992. Remembering pictures: pleasure and arousal in memory. Journal of Experimental Psychology 81 (2), 379–390.

Brostoff, S., Sasse, M.A., 2000. Are Passfaces more usable than passwords: a field trial investigation. In: McDonald, S., et al. (Eds.), People and Computers XIV—Usability or Else, Proceedings of HCI 2000. Springer, Berlin, pp. 405–424.

Brown, A.S., Bracken, E., Zoccoli, S., Douglas, K., 2004. Generating and remembering passwords. Applied Cognitive Psychology 18, 641–651.

Coventry, L., De Angeli, A., Johnson, G., 2003. Usability and biometric verification at the ATM interface. In: Proceedings CHI 2003. ACM Press, New York, pp. 153–160.

Craik, F.I.M., Lockhart, R.S., 1972. Levels of processing: a framework for memory research. Journal of Verbal Learning and Verbal Behavior 11, 671–684.

Davis, D., Monrose, F., Reiter, M.K., 2004. On user choice in graphical password schemes. In: Proceedings of the 13th USENIX Security Symposium, San Diego, August 2004.

De Angeli, A., Coventry, L., Cameron, D., Johnson, G.I., Fischer, M., 2002. VIP: a visual approach to user authentication. In: Proceedings of the Working Conference on Advanced Visual Interfaces AVI2002. ACM Press, New York, pp. 316–323.

Dhamija, R., 2000. Hash visualization in user authentication. In: Proceedings of CHI 2000. ACM Press, New York, pp. 279–280.

Dhamija, R., Perrig, A., 2000. Déjà Vu: User study using images for authentication. In: Ninth Usenix Security Symposium.

Dourish, P., 2004. Security as experience and practice: supporting everyday security. Talk given at the DIMACS Workshop on Usable Privacy and Security Software, July 7, 2004.

Feldmeier, D.C., Karn, P.R., 1990. UNIX password security—ten years later. In: Advances in Cryptology—CRYPTO'89. Lecture Notes in Computer Science 435. Springer, Berlin, pp. 44–63.

Fitts, P.M., 1954. The information capacity of the human motor system in controlling amplitude of movement. Journal of Experimental Psychology 47, 381–391.

Hollingworth, A., Henderson, J.S., 2002. Accurate visual memory for previously attended objects in natural scenes. Journal of Experimental Psychology—Human Perception and Performance 28, 113–136.

Hollingsworth, A., Williams, C.W., Henderson, J.M., 2001. To see and remember: visually specific information is retained in memory from previously attended objects in natural scenes. Psychometric Bulletin and Review 8 (4), 761–768.

Ives, B., Walsh, K.R., Schneider, H., 2004. The domino effect of password reuse. Communications of the ACM 47 (4), 76–78.

Jermyn, I., Mayer, A., Monrose, F., Reiter, M.K., Rubin, A.D., 1999. The design and analysis of graphical passwords. In: Proceedings of the Eighth USENIX Security Symposium, pp. 1–14.

Klein, D., 1990. A survey of, and improvement to, password security. In: UNIX Security Workshop II.

Mandler, J.M., Ritchey, G.H., 1977. Long-term memory for pictures. Journal of Experimental Psychology: Human Learning and Memory 3, 386–396.

Morris, R., Thompson, K., 1979. Password security: a case study. Communications of the ACM 22, 594–597.

Nelson, D.L., Reed, U.S., Walling, J.R., 1977. Picture superiority effect. Journal of Experimental Psychology: Human Learning and Memory 3, 485–497.

Nielsen, J., 2004. User education is not the answer to security problems. http://www.useit.com/alertbox/20041025.html. Accessed January 26, 2005.

Norman, D.A., 1988. The Design of Everyday Things. Basic Books, New York.

Paivio, A., Rogers, T.B., Smythe, P.C., 1976. Why are pictures easier to recall then words? Psychonomic Science 11 (4), 137–138.

Patrick, A.S., Long, A.C., Flinn, S., 2003. HCI and security systems. In: Proceedings of the CHI 2004. ACM Press, New York, pp. 1056–1057.

Real User Corporation, 2001. The science behind Passfaces. http://www.realusers.com. Accessed December 2, 2002.

Rundus, D.J., 1971. Analysis of rehearsal processes in free recall. Journal of Experimental Psychology 89, 63–77.

Sasse, M.A., Brostoff, S., Weirich, D., 2001. Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. BT Technical Journal 19, 122–131.

Scholtz, J., Johnson, J., 2002. Interacting with identification technology: can it make us more secure? In: Proceedings of the CHI 2002 Extended Abstracts. ACM Press, New York, pp. 564–565.

Shepard, R.N., 1967. Recognition memory for words, sentences, and pictures. Journal of Verbal Learning and Verbal Behavior 6, 156–163.

Standing, L.P., 1973. Learning 10,000 pictures. Quarterly Journal of Experimental Psychology 25, 207–222.

Tenner, E., 1977. Why Things Bite Back: Technology and the Revenge of Unintended Consequences. Vintage Books, New York.

Weinshall, D., Kirkpatrick, S., 2004. Passwords you'll never forget, but can't recall. In: Proceedings of CHI 2004. ACM Press, New York, pp. 1399–1402.

Wixted, J.T., 2004. The psychology and neuroscience of forgetting. Annual Review of Psychology 55, 235–269.

Wixted, T.J., Ebbesen, E.B., 1991. On the form of forgetting. Psychological Science 2, 409–415.