Design and Evaluation of a Shoulder-Surfing Resistant Graphical Password Scheme

Susan Wiedenbeck and Jim Waters

College of IST

Drexel University

Philadelphia, PA 19104 USA

{sw53, jw65}@drexel.edu

Leonardo Sobrado and Jean-Camille Birget

Computer Science Department

Rutgers University at Camden

Camden, NJ 08102 USA

{lsobrado, birget}@camden.rutgers.edu

ABSTRACT

When users input their passwords in a public place, they may be at risk of attackers stealing their password. An attacker can capture a password by direct observation or by recording the individual's authentication session. This is referred to as shoulder-surfing and is a known risk, of special concern when authenticating in public places. Until recently, the only defense against shoulder-surfing has been vigilance on the part of the user. This paper reports on the design and evaluation of a game-like graphical method of authentication that is resistant to shoulder-surfing. The Convex Hull Click (CHC) scheme allows a user to prove knowledge of the graphical password safely in an insecure location because users never have to click directly on their password images. Usability testing of the CHC scheme showed that novice users were able to enter their graphical password accurately and to remember it over time. However, the protection against shoulder-surfing comes at the price of longer time to carry out the authentication.

Author Keywords

Shoulder-surfing, Convex Hull Click scheme, graphical passwords, authentication, password security, usable security.

ACM Classification Keywords

H.5.2 [**Interfaces and Representation**]: User Interfaces – *Graphical user interfaces;* K.6.5 [**Computing Milieux**]: Security and Protection – *Authentication.*

INTRODUCTION

Traditionally, alphanumeric passwords have been used for user authentication. While today other methods including biometrics and smart cards are possible alternatives, passwords are likely to remain dominant for some time because of concerns about reliability, privacy, security, and ease of use of other technologies [2]. However, in the use of passwords dilemmas often arise in the tradeoff between security and usability. The dilemma, as formulated by Birget in [21, p. 104], arises because passwords are expected to comply with two fundamentally conflicting requirements:

(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.

Because it is difficult for humans to remember random strings, users tend to ignore requirements for secure passwords. This leads to poor password practices, including short, simple passwords that are easy to break either by a dictionary attack or personal knowledge of the password owner, use of the same password over months or years, reuse of identical or nearly identical passwords on multiple systems, and propensity to write down passwords and store them insecurely, e.g., a text file containing the user's passwords stored on insecure computers or PDAs, Post-Its notes stuck on or near the computer monitor or inside a desk drawer [1, 3, 10, 11, 13, 14].

In an effort to improve password security by making passwords easier to remember, researchers have developed graphical passwords. In a typical graphical password scheme a user chooses several images to be his or her password. When logging in, the user must click on the password images among a larger group of distractor images. If the user clicks on the correct images, he or she is authenticated. Users' memory for a graphical password may be better than for an alphanumeric password. Secure alphanumeric passwords (i.e., random strings) are based on pure recall from memory, a skill that is notoriously difficult for humans. By contrast, graphical passwords are based on recognition of previously known images, a skill at which humans are proficient. Indeed, image-based passwords have shown good memorability in user testing [2, 5, 6, 8, 9, 21].

However, the problem of shoulder-surfing is a recognized drawback of graphical passwords. Shoulder-surfing refers to someone watching over the user's shoulder as the user enters a password, thereby capturing the password. While alphanumeric passwords systems are vulnerable to shoulder-surfing if the attacker can see the keyboard, graphical password systems may be more vulnerable in certain settings. For example, clicking on images on a large, vertical display screen may make users' actions easier to capture.

In this paper we introduce a graphical password scheme based on [19] that is resistant to shoulder-surfing, whether by a human observer or by a video camera recording the login. Using concepts

from affective computing, the scheme harnesses the concepts of challenge and mastery [7, 12] to motivate high password performance in a game-like setting. A user study explores the password scheme in terms of accuracy, efficiency, memorability, and user satisfaction.

BACKGROUND TO THE RESEARCH

Graphical Password Systems

A common approach to design of graphical password systems is a challenge-response scheme. In a challenge-response scheme the user creates a password by choosing several images from a large portfolio of images. The chosen images become the user's password. To log in the user must successfully respond to a series of challenges. In a challenge the user is simultaneously shown several images on the screen, where one of the images is a password image of the user and the rest are decoy images. The user responds by clicking anywhere on the password image. In each subsequent challenge the user is shown another password image surrounded by different decoys. The user logs in successfully if all challenges are responded to correctly.

The advantage of this kind of challenge-response system is that it relies on recognition memory [14]. In each challenge the user simply views displayed images and chooses the known image. However, a possible drawback is the amount of time for carrying out a series of challenges. A larger password space, and therefore higher security, can be achieved only by a large number of decoy images in each challenge or a large number of challenges. Both of these increase the login time. Another potential drawback is that users may be strongly attracted to certain images [4, 6]. If different users tend to choose the same images for their password, the entropy of the system decreases, making it less secure.

Usability testing of three such challenge-response systems, Passfaces [2], Déjà Vu [8, 9], and VIP [5, 6] give a sense of performance in such systems. Passfaces was tested in a longitudinal field study, using four rounds of challenges to log in. In each challenge there were nine images of human faces, with one password face and eight decoy faces. The user had to click on his or her password face. If the user responded to all four rounds correctly the user was authenticated and gained access to the system, but if there was an error in any challenge the user was rejected. Passfaces and alphanumeric passwords were compared over a ten week period. The Passfaces login failure rate was 4.9 percent, while the alphanumeric password failure rate was 15.1 percent. Data on speed of login were not reported.

Déjà Vu is somewhat similar to Passfaces in terms of the login procedure, but it uses abstract images and concrete photographs instead of faces. The login is accomplished in one round where the user simultaneously sees 25 images displayed on the screen, five of which are the user's password images and the remaining 20 are decoy images. To be authenticated the user must click on all five password images and not click on any of the decoy images. A comparative usability study was carried out of Déjà Vu, six-character alphanumeric passwords, and four-digit numeric PINs. After initially choosing password images and practicing the login procedure, participants using Déjà Vu with both photographs and abstract images were able to log in successfully, whereas the failure rate was 5 percent for both the PINs and the alphanumeric passwords. In a follow-up session one week later, the Déjà Vu login failure rate on photographs and abstract images were 5 percent and 10 percent respectively. The failure rates for PINs and alphanumeric passwords were much higher in the follow-up

session: 35 percent for PINs and 30 percent for passwords. The efficiency of authentication in Déjà Vu was relatively low, taking approximately 30 seconds to log in.

VIP is a graphical PIN authentication system that is meant for use with both a PIN *and* an ATM card. Consequently, it has a smaller password space than would be required in a password system. Nevertheless, the usability testing results are highly relevant to our work. VIP uses the challenge-response approach with several rounds of challenges. On each challenge the user sees one of his or her password images along with nine distractor images (which change over challenges). A study compared memorability, accuracy, and speed of input of four-digit alphanumeric PINs and three versions of VIP. Longitudinal testing showed that the visual PIN schemes were less error prone than the alphanumeric PINs. Individuals using VIP failed in 6 percent of their authentication trials. An analysis of these errors showed that the most common error was confusing a distractor image with one of the user's password images when they were in the same image category (e.g., the "flower" category). It was also found that errors occurred less frequently over multiple logins if, in each login attempt, the user's password images always occurred in a fixed position in the ATM interface. The efficiency of authentication using VIP varied across the three versions, ranging from approximately 6 to 16 seconds, compared to 3 seconds for the numeric PIN.

The studies of the three systems suggest that passwords using a challenge-response scheme are easy for users to remember over time. This is important because, as discussed above, many of the insecure practices of users occur because of difficulties remembering passwords. This led us to use a challenge-response approach to developing a shoulder-surfing resistant password system.

Shoulder-Surfing Problem and Defenses Against It

Shoulder-surfing occurs when an attacker learns a user's password by watching the user log in. Shoulder-surfing is a well known method of stealing passwords and other sensitive information and is recognized by practitioners as a serious security threat [16, 18, 20]. It can occur in offices and public places without the user's awareness.

In the simplest version of shoulder-surfing a human attacker takes up a position where a user's login can be seen. Typically this might occur at a wireless hotspot in a busy, crowded public environment, such as a shopping mall, airport, or coffee shop [16, 20]. Using an alphanumeric password, though the user's password is not displayed on the screen, a practiced attacker can "read" the user's keystrokes as the user types the password. The user's only defense is to shield the keyboard with an object or one's body. Using a graphical password, the user would have to shield the screen. The same considerations apply to entering PINs at ATMs. High tech versions of shoulder-surfing are also a threat, although it is not known how prevalent the threat is. Technology-based attacks include using binoculars or a low power telescope to enhance vision, using video cameras, video mobile phones, keystroke logging software, or Trojan software to record a login, and listening to a user input a PIN or account number on a telephone keypad. Remote electro-magnetic sensors can also be used to capture actions without the user's knowledge.

As indicated above, both alphanumeric and graphical passwords are vulnerable to shoulder-surfing. The degree of the threat depends on the situation. For example, a graphical password might be quite vulnerable if the user enters the password by clicking on

images on a large screen in a physical environment where observation is easy. On the other hand, a graphical password entered with a stylus on a smaller screen would probably be much harder for a human attacker or a video camera to capture because the device can be held closer to the body, the user's hand tends to obstruct observation, and the user does not necessarily stay in a fixed location.

Aside from advising users to be aware of the threat and to avoid entering sensitive information in an insecure location, there is little practical help for the user against shoulder-surfing. However, recent research is beginning to target this problem. Roth and his colleagues [15] have developed a scheme to protect numeric PIN-entry systems against shoulder-surfing. The scheme involves a classical form of two-factor authentication, where the legitimate user is in possession of a token (in this case, a card with a magnetic strip), and knows a secret (in this case, a four-digit PIN). Instead of asking the user to input the explicit digits, the system verifies that the user knows the PIN by asking three or four questions about each digit of the PIN (for example, is it even or odd; is it less than 5). The questions and answers are conveyed not through words but through a visual coding of the digits on the screen that is not immediately obvious to an observer. Different questions are used at different logins so an observer cannot simply memorize the legitimate user's responses.

Roth et al.'s scheme makes PIN-entry shoulder-surfing resistant, to a limited degree. The limitation of this scheme is that an attacker can figure out the PIN given full information about the questions and answers, for example, if the attacker has an excellent memory or records the login. An attack would further be facilitated by observing multiple logins in which the attacker could accumulate more knowledge about the PIN that would eventually determine it. These limitations may be acceptable because the attacker must have both the token and the secret PIN to gain access. Note that this differs significantly from authentication by passwords, where there is no token and the secret is the only defense against unauthorized access. Consequently, passwords should be stronger (longer and from a larger set of characters or images).

CONVEX HULL CLICK SCHEME

Our shoulder-surfing resistant scheme, the Convex Hull Click Scheme (CHC), is a graphical password scheme that guards against shoulder-surfing attacks by human observation, video recording, or electronic capture. Like Passfaces, CHC is based on several rounds of challenge-response authentication. Like Roth et al.'s PIN-entry scheme, users never point directly to the items that form their passwords. In CHC the graphical elements used in authentication are icons shown in a window on the screen. In a challenge the user must recognize some minimum number of his or her password icons, or "pass-icons," out of a much larger number of randomly arranged icons. The user responds to the challenge by clicking within the convex hull of the pass-icons. Several such challenges are presented in sequence, and if the user responds correctly to every one then the user is authenticated. Using a game-like approach, CHC is designed to motivate the users to log in quickly and accurately [19]. The following paragraphs describe the design and implementation of CHC in more detail.

The system uses a large portfolio consisting of several hundred icons. In our implementation the icons used were all icons of software applications, but the portfolio of icons could be any kind of small icons, even user-provided ones. The icons are displayed using only the image without text. To create a password the user chooses several icons from the portfolio to be his or her pass-icons (Figure 1). The number of pass-icons is determined by the system administrator. The user has to remember the pass-icons he or she selected. Therefore, it is advisable for the user to commit them to memory and practice using them.

Figure 1. Five pass-icons (from left to right: Adobe Photoshop, Quark Express, Internet Explorer, Mozilla, Netscape). These pass-icons were used by participants in the usability study.

At login time a large number of icons from the portfolio are randomly arranged in the password window (Figure 2). These icons include mostly non-pass-icons along with a few pass-icons. The number of pass-icons displayed is a random number between three and the total number of pass-icons. (At least three pass-icons are guaranteed to be displayed on the window, since forming a convex hull requires at least three icons.) The login takes place in a series of challenge-response rounds. The number of rounds is controlled by the administrative setting, so this is easily changed, with more rounds providing higher security.

Figure 2. Graphical password interface used in the experiment.

When the login begins, the user must visually locate three or more of his or her pass-icons. The user's next step is to mentally create the convex hull formed by those pass-icons. A convex hull is the area encompassed by the edges joining a set of three or more points. In CHC the pass-icons serve as the points, and the edges are lines visualized in the user's mind. For illustrative purposes, Figure 3 shows a highlighted convex hull formed by three pass-icons. (Note that highlighting is not used when a user interacts with the system.) Figure 4 shows a convex hull formed by five pass-icons. To respond to the challenge, the user clicks anywhere within the convex hull. The user does *not* click on the pass-icons themselves and therefore does not give away to an attacker the identity of the pass-icons. Some convex hulls may be very narrow, as shown in the excerpt of the password window in Figure 5. This

can make clicking accurately in the convex hull difficult. However, if this occurs the implementation guarantees that there is always at least one more pass-icon in the window that can be used to form a wider convex hull.

Figure 3. Example of a convex hull with 3 pass-icons.

Figure 4. Example of a convex hull with 5 pass-icons.

Figure 5. A narrow triangle.

When the user has responded to the challenge, another challenge appears, and this continues until all challenges have been completed. The password window changes between the rounds of challenges. The non-pass-icons move to new random positions in the window. In addition, some portion of them randomly leaves the window, and a random number of new ones enter the window. Thus, the total number of icons visible in the window changes from one challenge to the next. Pass-icons likewise move to new positions. They may move out of the window and other pass-icons may enter it, with the constraint that there must always be three or more displayed in the window. The reason for moving icons into and out of the window is to make it harder for an attacker to guess the pass-icons. An individual cannot use the same pass-icons in every challenge, and therefore the attacker cannot easily determine which ones are the pass-icons. Further, guessing becomes more difficult given the constant changes of the non-pass-icons.

The rearrangement of the icons between challenges is done in a fluid, game-like animation. New icons pour into the window from the lower right corner. Simultaneously, icons leaving the window flow out in the upper left corner. Icons remaining in the window between challenges make quick movements from their old location to a new location. The rearrangement of the icons takes approximately three seconds. The movement of the icons is visually attractive and contributes to the sense of interest and speed. The smoothness and rapidity of the rearrangement is meant to increase the user's eagerness to respond to the following challenge in a like manner.

If the user responds correctly to each of the challenges he or she is authenticated, but if the user fails any challenge the login fails. The user is given feedback at the end of the login that indicates whether the logon is correct or not. As in all password systems, the user is not given specific information about the location of the error. To support the game-like design of CHC the time to complete the authentication is also displayed to motivate the user to accuracy and higher speed in the next password input.

Beyond its shoulder-surfing resistant properties, there are three main considerations about the security of CHC. First, the password space can be made very large, and therefore more secure, by increasing the number of icons, the number of pass-icons, or both. The only practical limits are the size of the window and the ability of users to locate their pass-icons among a large number of icons. Second, a brute-force attack is infeasible. An attacker could try to record all possible passwords that do not contain the click points observed by shoulder-surfing. After successive observations, the attacker could rule out more and more passwords. However, eventually the attacker would have to record a significant portion of all possible passwords, which would require far too much memory. Third, in challenge-response authentication there is always the possibility of accidental login (i.e., an attacker could click in the convex hull by luck). This is different from guessing the password. To make accidental login unlikely we do three things: (1) icons are randomly placed in the password window so that all locations except near the border of the window have about the same probability of being in the convex hull of the pass-icons, (2) large convex hulls that cover half the window or more are only rarely generated, and (3) to log in the user has to respond to multiple challenges.

Two additional security considerations about CHC are worth mentioning. First, potentially an attack against the system could be mounted using an eye-tracker. The eye-tracker could map where the user is looking while creating the convex hull and, at least in some cases, discover the pass-icons. We consider this a potential

threat, not an imminent threat, because current eye-trackers cannot be used without being detected by the user. Many eye-trackers use head mounted cameras. Recently eye-tracking cameras have been integrated into a panel attached below a monitor (e.g., the Tobii ET-1750). Nevertheless, these integrated cameras are still quite obvious and, to our knowledge, they are only integrated into stand-alone monitors. Second, a known security problem of all human usable challenge-response systems is that the system needs to know the password explicitly (in order to make challenges and to check correctness of the responses), and therefore the password cannot be encrypted. This is not so for alphanumeric password systems that can store password as encrypted strings and compare the encrypted password input of a user to the stored encrypted string.

USABILITY STUDY

Participants
Fifteen novice users, who were unfamiliar with the CHC scheme, were recruited from a university community. The participants were staff and students. Nine were female and six were male. The mean age was 37 (StdDev=13.6). The participants were experienced computer users who reported using computers for at least 5 hours per day for work and personal activities.

Materials
The system used in testing displayed up to 112 icons in a window of 800x600 pixels. The set of software application icons was used (See Figure 2 for a window containing all 112 icons). The number of icons displayed in any challenge ranged randomly from 43 to 112. The average number of icons visible in a challenge was 83. The pass-icons were five icons designated by the experimenter and used by all the participants (Figure 1). During a login three to five pass-icons randomly chosen by the system were shown in the window in each of five rounds of challenges. Note that the number of non-pass-icons, pass-icons, and challenges was smaller than desired for high security. We used a scaled down setting to determine whether, in a simple situation, novices could learn, remember, and enter passwords successfully using CHC. A computer-based tutorial was developed in Powerpoint to explain the system and train the participants to use it.

Procedure
Participants carried out the usability study individually in two sessions, an initial session and a follow-up session one week later.

In the first session the participant answered a short demographic questionnaire. Next the experimenter explained the purpose of the system and how it worked, using the tutorial materials. The experimenter explained the concept of the convex hull using the highlighting as shown in Figures 3 and 4. The participant was told to locate three or more pass-icons in the window and mentally form the convex hull of those pass-icons. The participant was specifically told not to click on the pass-icons, but within the convex hull. The experimenter also instructed the participant not to move the mouse from one pass-icon to another to "trace" the boundaries of the convex hull, since that could reveal the pass-icons to an attacker. The experimenter explained to the user that feedback on correctness would be given only at the end of the whole password input, not between each of the five challenges. The tutorial took approximately ten minutes.

At that point the participant authenticated him or herself repeatedly until ten successful logins were achieved. The system was instrumented to collect data on the number of correct and incorrect logins, the number of correct and incorrect challenges, and the total time for each correct and incorrect login and challenge. The authentication trials took about 15 minutes. Finally, the participant was interviewed for approximately ten minutes about perceptions of the system.

In the follow-up session one week later the participant was shown a list of the 112 unlabeled icons in random order and was asked to identify the five pass-icons. The session lasted about five minutes.

RESULTS
All participants achieved the criterion of ten correct logins. Eight of the participants accomplished the criterion in ten attempts with no incorrect logins. The other seven participants made from 1 to 4 incorrect logins, for a total of 15 incorrect logins across all participants. These individuals continued inputting their password until they reached ten correct logins. The mean percentage of correctness for password inputs among all participants was 90.35 percent (StdDev=10.50). An analysis of correctness of the challenges showed that the mean percentage of correctness was 97.95 (StdDev=2.20). The data further indicated that the login failures were the result of an incorrect response in a single challenge.

The mean time for correct password inputs was also analyzed. One outlier was excluded from the time analyses because of extremely slow password inputs. The mean time of this individual was more than two times higher than the next slowest participant and increased the overall mean time for all participants by 8 seconds. It was clear that this participant did not follow the study instructions, which were to balance time and accuracy. The mean time for correct password inputs (without the outlier) was 71.66 seconds (StdDev=25.17). The range was 24.08-150.42 seconds. The mean time in seconds for the challenges was 10.97 (StdDev=8.58). The range was 0.35-56.50 seconds.

Figure 6 shows the mean times for input of the ten correct passwords. The graph indicates that there was a gentle downward trend in time to input the password. A one-way within-subjects ANOVA showed a significant decrease in time to input the password over the 10 logins ($F_{(9, 117)} = 9.73$, $p < .031$).
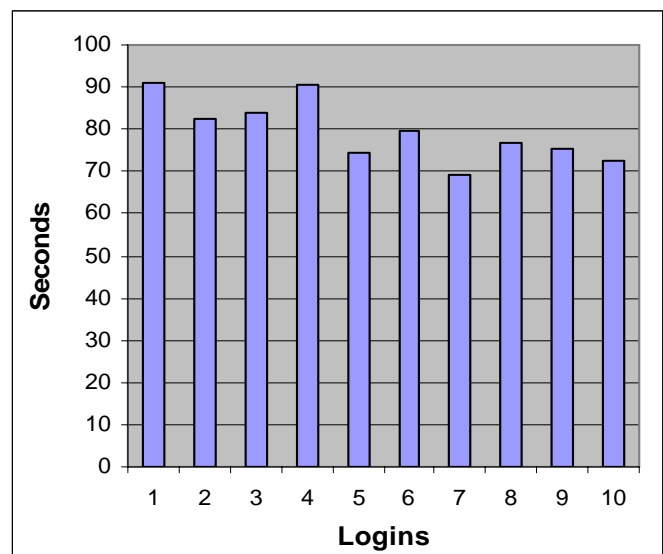

Figure 6. Mean times (seconds) for input of 10 correct logins (N=14).

In each challenge 3, 4, or 5 pass-icons were displayed in the window with the number of pass-icons randomly determined. A question was whether the number of pass-icons displayed affected the time to respond to a challenge. Table 1 shows the means and standard deviations of the time for responding to challenges displaying 3, 4, and 5 pass-icons. The mean time to respond declined when higher numbers of pass-icons were displayed. A one-way within-subjects ANOVA was carried out. The result showed that there was a significant difference in time to respond to a challenge ($F(2, 26)=5.62$, $p<.023$). A follow-up Newman-Keul's test indicated that challenges with five pass-icons were responded to significantly more quickly than challenges with 3 and 4 pass-icons ($p<.05$). There was no significant difference in the time to respond to challenges with 3 and 4 pass-icons.

Table 1. Means and standard deviations (seconds) of challenges with 3, 4, and 5 pass-icons.

	Mean	StdDev
3 pass-icons	11.72	3.77
4 pass-icons	11.55	4.74
5 pass-icons	9.71	3.82

An analysis of the relationship of the total number of icons (pass-icons plus non-pass-icons) displayed in the window in a challenge and the time to respond to the challenge was also carried out using linear regression. The result was non-significant, and the R^2 value indicated that the total number of icons displayed explained less than 1 percent of the variance in time to respond.

In session 2, one week later, 14 participants were able to correctly identify their five pass-icons from the randomly ordered printed list of 112 icons. One participant was able to identify four of the five pass-icons correctly.

The interview data were analyzed to get a better understanding of novice's perceptions of the usability of the system. CHC was designed to have a fluid, game-like feel. Ten participants reported that they did indeed find it fun. The game-like attributes cited included the challenge to oneself to enter the password as quickly as possible ("beating my best time"), the "cool" animations of the icons between the challenges, and the challenge to oneself of tracking the pass-icons as they moved. Three participants did not find CHC fun, describing the challenges as too repetitive (but this was influenced by the fact that participants input 10 passwords in a row, something a user would not normally do). Two participants did not use computer games and had no opinion.

Participants were asked about the ease of finding pass-icons in the password window. The participants agreed that in most cases they were able to pick out at least three icons quickly without much systematic scanning. Several mentioned that occasionally they had difficulty finding the minimum of three icons. They attributed this to certain specific pass-icons that they found did not stand out sufficiently, particularly the Adobe Photoshop icon. Two participants reported that it took more time to find pass-icons

when a larger number of icons was displayed in the window (a perception that was not supported by our analysis). Two participants gave a different explanation, saying that it took longer to find pass-icons when they were not close to each other, requiring them to scan a larger portion of the window. Indeed, it may be the case that in this study users' input time was affected more by the location of the pass-icons than the total number of icons displayed.

When participants located their pass-icons they had to mentally form the convex hull in order to click in it. Participants agreed strongly that mentally forming the convex hull was easy by simply looking from one pass-icon to the next. One participant wanted to trace an imaginary line between pass-icons by moving the mouse but did not do so because he understood that this would reveal the pass-icons to an attacker. A problem described by six participants was the difficulty of forming a convex hull and clicking in it when the pass-icons defined a narrow triangle, as in Figure 5. Participants were not sure whether they were clicking inside or outside the triangle. Two people who discussed this problem said that they looked for additional pass-icons in this case. However, other users simply tried to make their best effort to click within their estimated boundaries of the triangle, a risky strategy because the user must slow down to click accurately and may still fail to click within the convex hull.

To find pass-icons quickly, one can track the position of the pass-icons as they move during the rearrangement phase between challenges. Icon tracking was not mentioned during training. Nevertheless, seven participants reported tracking their pass-icons and found it very helpful in speeding up the following challenge. Most agreed that it was feasible to track only one or two pass-icons at a time. Four participants reported that they tried to track pass-icons, but they gave up this strategy because they were frustrated when the icons they were tracking moved out of the window. Finally, four participants did not consider tracking the pass-icons.

Participants were asked about their perceptions of how long it took to correctly input a password. Ten of the 15 participants thought that the time was acceptable, while five thought that it was too slow for practical use. Two participants who considered the input time acceptable explained that they would use this scheme for situations that required high security but did not involve frequent logins. One individual quantified this by saying that logging in with this system once or twice a day would be acceptable.

On the question of perceived ease of use, 13 individuals reported that the system was easy to use. One participant characterized it as "interesting, attractive, not boring." Others, while considering it easy to use in general, mentioned the problem of narrow triangles and the cognitive load of remembering and locating pass-icons. When participants were asked for open-ended comments about the system, they were largely positive, using words such as "pretty easy," "didn't frustrate," "interesting, fun," "novel," "I would use it in an internet café." Feedback on improvement centered on making the icons more distinct by changing the shape and color and removing icons that were difficult to spot.

DISCUSSION
In summary, the CHC scheme was easy to learn and remember. After a short training session, more than half the participants made no errors in entering their password consecutively ten times. Furthermore, participants who entered incorrect passwords were "nearly" correct, having only one incorrect click in the five

challenges. With one exception, participants who entered an incorrect password succeeded in the next attempt. Memorability was also high, with only one participant failing to remember one pass-icon a week later.

The high accuracy and memorability are similar to other challenge-response graphical password systems. For example, in Passfaces [2] the password failure rate was 4.9-5.3 percent in a longitudinal field trial. In Déjà Vu [8, 9] the error rate was 0 percent in a test immediately after training and 5-10 percent one week later. The slightly higher percentage of incorrect password inputs in CHC is understandable in terms of the cognitive effort involved. While the systems mentioned are similar to CHC in the use of a challenge-response approach, they were not designed to guard against shoulder-surfing. Guarding against shoulder-surfing makes CHC more complex. For example, Passfaces required the user to recognize one face out of nine in each challenge. In a challenge in CHC users had to locate three to five small pass-icons randomly interspersed among many non-pass-icons. Finding the pass-icons in the window led to more scanning than in Passfaces, and the small icons required more attention, as some icons looked similar enough to be confusable. Once the user found the pass-icons it was not merely a question of clicking on them, as in Passfaces. Instead, the user had to mentally form the convex hull before clicking. These factors increase the cognitive load.

From a usability viewpoint, the weak aspect of this system is the time to input a password. First, it should be noted that in general challenge-response password systems share this drawback to some extent, given multiple rounds of challenges as in Passfaces or multiple images to recognize as in Déjà Vu. In CHC a login with five challenges took novices a mean of 72 seconds, while Déjà Vu reported times of 27-32 seconds to login. CHC takes longer because of scanning the window for pass-icons and forming the convex hull. In addition, the time to rearrange the icons between challenges takes a mean of 3.4 seconds, which adds a total of 17 seconds per password. This overhead does not exist in Passfaces or Déjà Vu.

One way that CHC password inputs can be speeded up is for users to become more skillful. The novices in our study increased their speed moderately over ten trials, and we expect that they would improve more with extended use. Since CHC is a research prototype, we do not have an experienced user base for comparison. However, a rough benchmark calculated from the input speed of two individuals in our lab, who used the system daily, suggests that an experienced user might be able to input a password in about 34 seconds. This is much closer to Déjà Vu input times.

In addition to simply using CHC more, users will increase their speed as they learn a few tricks to facilitate fast input, for example, tracking pass-icons as they move. A second way to speed up the input is to carry out the icon rearrangement faster between challenges. If we reduced the rearrangement time by half this would reduce the total rearrangement time per password by 8.5 seconds. This modification would require another usability study. The shorter rearrangement time could change the game-like feel of the system and make it harder for users to track icons. A third way to increase speed would be to show more pass-icons in the password window, since our study found faster performance with more pass-icons visible. The user is likely to find the minimum of three pass-icons more quickly when there is a larger number of pass-icons in the window. A fourth approach is to improve elements that slowed the users. For example, the narrow triangles slowed the users, who had to pay careful attention to

where they clicked to ensure that the click was within the convex hull. Likewise, certain icons were difficult to notice in scanning. Finally, we should put the input time in the context of use. Most of our participants thought that the time, while substantially longer than alphanumeric passwords, was not prohibitive. In particular, users said that they were willing to trade-off time for higher security. If the user does not have to login repeatedly during the day, then the extra time becomes more acceptable.

We introduced novices to CHC in a scaled-down environment. Moving from this prototype to real usage, the security settings would have to be more stringent. In CHC security depends on a combination of total icons, pass-icons, and number of rounds of challenges. These three factors can be manipulated to achieve the level of security desired. A reasonable target for many purposes would be 500 hundred total icons in the portfolio, with about 200 hundred displayed in a challenge (in a larger window up to double the size of our current 800x600 window). A larger number of pass-icons, for example 10 to 12, would make the password less guessable to an observer because the user would have a wider variety of subsets of three or more icons in forming their convex hull. Finally, increasing the number of rounds to about 10 would decrease accidental login and planned attacks. We think that a highly practiced user could login using these more secure settings in about one minute. Usability testing with experienced users would have to be conducted to evaluate how the higher security affects users' performance and willingness to adopt CHC.

Given stronger security settings, other risks of CHC come largely from insecure user behavior. First, the users tend not to work with the entire convex hull (of all the pass-icons shown). Instead, they click in a triangle of pass-icons, and a shoulder-surfer does not know which triangle of the convex hull is used. However, if the triangle being used is very narrow, there is a tendency to click very near a pass-icon; this is an unsafe behavior. Second, moving the mouse to trace the edge between two pass-icons is highly insecure. If an observer is watching, it will reveal the pass-icons. Participants were trained to avoid this in our study, but it could be a greater problem with self-taught users. Third, if users choose their own pass-icons (which was not the case in this study), they may tend to choose certain salient icons more than others, reducing the entropy of the system.

Attention to redesign of icon sets and individual icons may substantially improve the overall usability of CHC. Users found that the icons failed to stand out visually because of their uniform shape and size. While the icons must be fairly small to display them in the password window, the icons can be made more distinct by varying the size, shape, and color. Furthermore, the icons do not have to be software icons. The icon sets can be user chosen and involve more personally memorable images.

CONCLUSION

The Convex Hull Click Scheme is an effort to develop security innovations with people in mind. As such, it is an example of "usable security," an approach to design of security systems that is gaining increasing attention. The contribution of this paper is the design of a graphical password system that extends the challenge-response paradigm to resist shoulder-surfing. In doing so it aims to motivate the user with a fun, game-like visual environment designed to develop positive user affect and counterbalance the drawback of the longer time to input the password. The user study and interviews support the overall concept but identify areas of improvement needed to enhance usability and reduce risks.

Future work should target increasing the speed of input of the password. There is no single solution to this problem. Instead, several incremental changes, human, technical, visual, and contextual, will improve the system. Humans can speed up with practice, the system can be tweaked to improve efficiency, and the icons can be improved. Contextual changes have to do with how the user thinks about the system. Most of our novice users felt the time was acceptable, although it was objectively long compared to a traditional alphanumeric password. Factors that potentially increase its acceptability to users are multiple: high security which warrants taking more time to login, use of CHC in contexts that do not entail logging in at frequent intervals, ease of remembering the pass-icons and inputting the password accurately, and importantly the "fun factor" of a game-like environment.

Further directions for CHC are to improve the current icons, create additional icon sets, make the security settings more fully realistic, and then test it in a longitudinal study of everyday use. This longitudinal study could be carried out in a research or teaching lab where users log in to computers daily. We also plan to investigate the entropy issue of pass-icons and to study in more depth the motivational aspects of the game-like approach.

ACKNOWLEDGMENTS

REFERENCES

1. Adams, A. and Sasse, M.A. Users are not the enemy. *CACM* 42, 12 (1999), 41-46.

2. Brostoff, S. and Sasse, M.A. Are Passfaces more usable than passwords: A field trial investigation. In McDonald S., et al. (Eds.), *People and Computers XIV - Usability or Else, Proc. of HCI 2000*, Springer, 2000, 405-424.

3. Brown, A.S., Bracken, E., Zoccoli, S. and Douglas, K. Generating and remembering passwords. *Applied Cognitive Psychology*, 18, (2004), 641-651.

4. Davis, D., Monrose, F., and Reiter, M.K. On user choice in graphical password schemes. In *Proc. of the 13th USENIX Security Symposium*, San Diego, 2004.

5. De Angeli, A., Coutts, M., Coventry, L., Cameron, D., Johnson, G.I., and Fischer, M. VIP: A visual approach to user authentication. In *Proc. of AVI 2002*, ACM Press, NY, 2002, 316-323.

6. De Angeli, A., Coventry, L., Johnson, G., and Renaud, K. Is a picture worth a thousand words? Exploring the feasibility of graphical authentication systems. *International Journal of Human-Computer Studies*, 63, 1-2 (2005), 128-152.

7. Deci, E.L. *Intrinsic Motivation*, Plenum, New York, 1975.

8. Dhamija, R. Hash visualization in user authentication. In *Proc. of CHI 2000*, ACM Press, NY, 2002, 279-280.

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

10. Feldmeier, D. C. and Karn, P. R. UNIX password security – ten years later. In *Advances in Cryptograph y – CRYPTO'89*, Lecture Notes in Computer Science 435, Springer-Verlag 1990, 44-63.

11. Ives, B., Walsh, K. R., and Schneider, H. 2004. The domino effect of password reuse. *CACM*, 47, 4 (2004), 76-78.

12. Lepper, M.R. and Malone, T.W. Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow and M. J. Farr (Eds.), *Aptitude, Learning, and Instruction,* Lawrence Erlbaum, Hillsdale, NJ, 1987, 255-286.

13. Morris, R. and Thompson, K. Password security: A case study. *CACM*, 22, (1979), 594-597.

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

15. Roth, V., Richter, K., and Freidinger, R. A PIN-entry method resilient against shoulder-surfing. *Proc. of the 11th ACM Conference on Computer and Communications Security*, 2004, 236-245.

16. Giblin, P. Identities snatched in blink of eye. http://www.sachitechcops.org/news012604.htm. Accessed December 9, 2005.

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

18. Shoulder-surfing gets secret numbers on tape. http://www.wftv.com/money/3964515/detail.html. Accessed December 9, 2005.

19. Sobrado, L. and Birget, J.C. Graphical passwords. The Rutgers Scholar, 4, (Sept. 2002). http:// RutgersScholar.rutgers.edu/volume04/sobrbirg/sobrbirg.htm.

20. Wagstaff, J. Shoulder-surfing: the old new phishing. http://loosewire.typepad.com/blog/2005/04/shoulder_surfin.html. Accessed December 9, 2005.

21. Wiedenbeck, S., Waters, J., Birget, J.C., Brodskiy, A., and Memon, N. PassPoints: design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, 63, (2005), 102-127