Graphical Passwords: Learning from the First Twelve Years

Robert Biddle, Sonia Chiasson, P.C. van Oorschot
School of Computer Science
Carleton University, Ottawa, Canada
robert_biddle@carleton.ca, chiasson@scs.carleton.ca, paulv@scs.carleton.ca

ABSTRACT

Starting around 1999, a great many graphical password schemes have been proposed as alternatives to text-based password authentication. We provide a comprehensive overview of published research in the area, covering both usability and security aspects, as well as system evaluation. The paper first catalogues existing approaches, highlighting novel features of selected schemes and identifying key usability or security advantages. We then review usability requirements for knowledge-based authentication as they apply to graphical passwords, identify security threats that such systems must address and review known attacks, discuss methodological issues related to empirical evaluation, and identify areas for further research and improved methodology.

Categories and Subject Descriptors

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

General Terms

Human factors, Security

Keywords

Authentication, graphical passwords, usable security

1. INTRODUCTION

Beginning around 1999, a multitude of graphical password schemes have been proposed, motivated by the promise of improved password memorability and thus usability, while at the same time improving strength against guessing attacks. Like text passwords, graphical passwords are knowledge-based authentication mechanisms where users enter a shared secret as evidence of their identity. However, where text passwords involve alphanumeric and/or special keyboard characters, the idea behind graphical passwords is to leverage human memory for visual information, with the shared secret being related to or composed of images or sketches.

Despite the large number of options for authentication, text passwords remain the most common choice for many reasons [55, 92]. They are easy and inexpensive to implement; are familiar to essentially all users; allow users to authenticate themselves while avoiding privacy issues that have been raised about biometrics; and have the advantage of portability without, for example, having to carry physical tokens. However, text passwords also suffer from both security and usability disadvantages — for example, passwords are typically difficult to remember, and are predictable if user-choice is allowed [10, 65, 73, 101, 144].

One proposal to reduce problems related to text passwords is to use password managers. These typically require that users remember only a master password. They store (or re-generate) and send on behalf of the user the appropriate passwords to web sites hosting user accounts. Ideally the latter are generated by the manager itself and are stronger than user-chosen passwords. However, implementations of password managers introduce their own usability issues [26] that can exacerbate security problems, and their centralized architecture introduces a single point of failure and attractive target: attacker access to the master password provides control over all of the user's managed accounts.

When text password users resort to unsafe coping strategies, such as reusing passwords across accounts to help with memorability, the decrease in security cannot be addressed by simply strengthening, in isolation, the underlying technical security of a system. Usability issues often significantly impact its real-world security. User interface design decisions may unintentionally sway user behaviour towards less secure behaviour. Successful authentication solutions must thus also include improved usability design based on appropriate research taking into account the abilities and limitations of the target users. In graphical passwords, human memory for visual information is leveraged in hope of a reduced memory burden that will facilitate the selection and use of more secure or less predictable passwords, dissuading users from unsafe coping practices.

Early surveys of graphical passwords are available [72, 113]. More recent papers briefly summarize and categorize 12 schemes [53], and review numerous graphical password systems while offering usability guidelines for their design [94]. In this paper we provide a comprehensive review of the first twelve years of published research on graphical passwords, and reflect on it. It is now clear that the graphical nature of schemes does not by itself avoid the problems

Version: January 4, 2011. Technical Report TR-11-01, School of Computer Science, Carleton University, 2011 This paper updates and obsoletes the earlier technical report TR-09-09. A version of this paper will appear in ACM Computing Surveys.

typical of text password systems. However, while proposals in this first period of research exhibit some familiar problems, we see signs that an emerging second generation of research will build on this knowledge and leverage graphical elements in new ways to avoid the old problems.

Our motivation is multi-fold. Knowledge-based authentication has increasing impact on society as its use expands from login to a single computer, to large numbers of remote computers hosting personal and corporate information, to authorizing online financial transactions via mobile devices. Some schemes such as PassFaces [84] and grIDsure [52] have commercial interests. Because of the difficulty of typing on mobile devices, authentication schemes using alternatives to keyboard entry are receiving increased attention. This magnifies the importance of understanding usability and security implications of such schemes. But, for example, as PIN-level graphical schemes are used for unlocking Android smart phones, we see little discussion of the security difference between PIN-level (with password spaces of 10,000 elements or 12–15 bits) and password-level schemes (with spaces of 30–60 bits). Besides providing specific authentication alternatives, we find research into graphical passwords allows for better understanding of knowledge-based authentication in general by looking at issues such as user choice in password selection, memory interference, and the role of cueing in password memorability.

We classify schemes into three main categories based on recall, recognition, and cued-recall, beginning discussion of each with a primary exemplar. We discuss further schemes and extensions offering interesting additional characteristics and improvements, or where significant usability studies or security analysis has allowed better understanding. Summary tables comparing schemes are given in Section 7. We review usability requirements and features for comparative analysis, highlight specialized analysis techniques, consider threat models, catalogue known attack strategies, and discuss the suitability of different schemes for various environments. We consider methodological issues for evaluation of proposals, discuss challenges related to empirical evaluation, and include an evaluation checklist. Throughout, we also extract lessons that can be learned from the research to date.

2. MEMORABILITY

For over a century, psychology studies have recognized the human brain's apparently superior memory for recognizing and recalling visual information as opposed to verbal or textual information [62, 68, 83, 106]. The most widely accepted theory is the *dual-coding theory* [82], suggesting that verbal and non-verbal memory (respectively, word-based and image-based) are processed and represented differently in the mind. Images are mentally represented in a way that retains the perceptual features being observed and are assigned perceived meaning based on what is being directly observed. Text is represented symbolically, where symbols are given a meaning cognitively associated with the text, as opposed to a perceived meaning based on the form of the text. For example, 'X' may represent the roman numeral 10 or the multiplication symbol; the exact meaning is associated in relation to some deeper concept. This additional processing required for verbal memory renders this a more difficult cognitive task.

Tasks involving visual memory can also vary in difficulty due to the particular characteristics of the retrieval process. Graphical passwords can be broadly categorized according to the memory task involved in remembering and entering the password: *recall*, *recognition*, and *cued-recall* [90].

Recall requires that a person remember information without cueing. With recognition, a person is provided with the information and has to decide if this matches the information previously memorized. Several theories exist to explain the difference between recognition and recall memory, based on whether these are two unique processes or are similar and differ only in their retrieval difficulty [4]. It is generally accepted, however, that recognition is an easier memory task than recall [61,123]. In cued-recall, an external cue is provided to help remember information. Tulving and Pearlstone [122] explain that items in human memory may be available but not accessible for retrieval and show that previously inaccessible information in a pure recall situation can be retrieved with the aid of a retrieval cue.

3. SECURITY

An authentication system must provide adequate security for its intended environment, otherwise it fails to meet its primary goal. A proposed system should at minimum be evaluated against common attacks to determine if it satisfies security requirements. A brief introduction is provided here and a more detailed discussion of security follows in Section 9.

We classify the types of attacks on knowledge-based authentication into two general categories: guessing and capture attacks. In successful *guessing attacks*, attackers are able to either exhaustively search through the entire theoretical password space, or predict higher probability passwords (i.e., create a dictionary of likely passwords) so as to obtain an acceptable success rate within a manageable number of guesses. Guessing attacks may be conducted *online* (as defined in Section 9.1) through the intended login interface, or *offline* if some verifiable text [50] (e.g., hashes) can be used to assess the correctness of guesses. Authentication systems with small theoretical password spaces or with identifiable patterns in user choice of passwords are especially vulnerable to guessing attacks.

Password *capture attacks* involve directly obtaining the password, or part thereof, by capturing login credentials when entered by the user, or by tricking the user into divulging their password. Shoulder-surfing, phishing, and some kinds of malware are common forms of capture attacks. In shoulder-surfing, credentials are captured by direct observation of the login process or through some external recording device such as a video camera. Phishing is a type of social engineering where users are tricked into entering their credentials at a fraudulent website recording user input. Malware uses unauthorized software on client computers or servers to capture keyboard, mouse, or screen output, which is then parsed to find login credentials.

As will be seen, early graphical password systems tended to focus on one particular strength, for example being resistant to shoulder-surfing, but testing and analysis showed that they were vulnerable to one or more other types of attacks. Except in very specific environments, these would not provide adequate security.

Often playing an important role related to security is the particular process of encoding or discretization used — transforming the user input into discrete units that can be identified by the system and used for comparison during

password re-entry. As will be seen, some schemes require that the system retains knowledge of the exact secret (or portion thereof), either to display the correct set of images to the user or to verify password entries. In other cases, encoded or discretized passwords may be hashed, using a one-way cryptographic hash, to provide additional security in case the password file is compromised.

4. RECALL-BASED SYSTEMS

Recall-based graphical password systems are occasionally referred to as *drawmetric systems* [31] because users recall and reproduce a secret drawing. In these systems, users typically draw their password either on a blank canvas or on a grid. Recall is a difficult memory task [29] because retrieval is done without memory prompts or cues. Users sometimes devise ways of using the interface as a cue even though it is not intended as such, transforming the task into one of cued-recall, albeit one where the same cue is available to all users and to attackers. Text passwords can also be categorized as using recall. With text passwords, evidence shows that users may use the name of the system as a memory cue and include it within their passwords [24, 131] . Although there is currently no evidence of this happening with graphical passwords, it remains a plausible coping strategy if users can devise a way of relating a recall-based graphical password to a corresponding account name.

A number of security vulnerabilities are common to most recall-based systems, as these systems share similar features. We briefly discuss some attacks here; see Section 9 for background and additional details. These systems are generally susceptible to shoulder-surfing to the extent that in many cases, the entire drawing is visible on the screen as it is being entered, and thus an attacker need accurately observe or record only one login for the entire password to be revealed.

Social engineering attacks remain a concern in cases where users can describe their password by, for example, verbalizing a path through grid squares, or by showing a sketch of the password. Phishing attacks are easily mounted. A phishing website can copy the login page from a legitimate site, including the area for drawing the graphical password (see Figure 1). Once users enter their username and password, this information can be used by attackers at the legitimate site. The recall-based schemes discussed below are also vulnerable to malware attacks based on screen scrapers, and mouse-loggers if an attacker can identify the position of the password entry grid on the screen through other means.

In typical recall-based systems, users choose their own passwords. Thus a personalized attack may be more successful than a general attack—someone familiar with the user may have a higher probability of guessing the user's password. For example, some users might choose to draw the initials of their name. While successful personalized attacks have yet to be reported in the literature for recall-based graphical systems, experimental results have been reported for password recovery mechanisms such as personal verification questions [102].

The following subsections offer an overview of recall-based graphical password schemes, based on Draw-A-Secret [58].

4.1 Canonical Example: Draw-A-Secret

Draw-A-Secret (DAS) [58] was the first recall-based graphical password system proposed. Users draw their password on a 2D grid using a stylus or mouse (see Figure 1). A drawing can consist of one continuous pen stroke or preferably several strokes separated by "pen-ups" that restart the next stroke in a different cell. To log in, users repeat the same path through the grid cells. The system encodes the user-drawn password as the sequence of coordinates of the grid cells passed through in the drawing, yielding an *encoded* DAS password. Its length is the number of coordinate pairs summing across all strokes.

There is little information on either the usability or the practical security of the original DAS system, as to date it has only been user tested through paper prototypes (but see also the related Pass-Go system, below). Nali and Thorpe [75] asked 16 participants to draw 6 "doodles" and 6 "logos" on 6×6 grids. These drawings were visually inspected for symmetry and number of pen strokes. They found that participants tended to draw symmetric images with few pen strokes (1-3), and to place their drawing approximately in the center of the grid. Study limitations included: users were not told that their drawings were "passwords", users did not have to later reproduce their drawings, and data was collected on paper (rather than users drawing using a computer). No usability data (login times, success rates, etc.) was collected. Dunphy and Yan [40] compared DAS to their BDAS scheme in two paper prototype studies (see below). For DAS, they found success rates ranging from 57-80%.

The size of the *theoretical password space*, that is, the number of all possible passwords regardless of how small their probabilities in actual practice, is related to the coarseness of the underlying 2D grid, and the maximum password length. For a 5×5 grid and maximum length 12, the theoretical password space of DAS has cardinality 2^{58} [58]. This is often stated as 58 bits for brevity, but should not be misinterpreted as 58 bits of entropy, since passwords are far from equi-probable. To allow verification, the system must store the encoded DAS passwords. To avoid storing them cleartext, a one-way function of the password, or cryptographic *hash*, may be stored, as is done with text passwords (see Section 9). Note that there is a many-to-one mapping from user-drawn passwords to encoded DAS passwords; for example, all doodles drawn entirely within one grid square are equivalent to a dot.

In summary, DAS does offer a theoretical space comparable with text passwords, but the possibility that users will prefer predictable passwords such as symmetric passwords with few strokes [126] suggests that, as with text passwords, the effective space will be considerably smaller. Without an implementation and user studies, we can tell little more. Similarly, while a key motivation for DAS was the superior memorability associated with images, the lack of suitable user studies leaves as an open question how effectively this can be leveraged in graphical authentication.

4.2 Other recall-based schemes

BDAS, proposed by Dunphy and Yan [40], added background images to DAS to encourage users to create more complex passwords. In a comparison of BDAS to DAS using paper prototypes, they reported that the background image reduced the amount of symmetry within password images, and led users to choose longer passwords that were similarly memorable to the weaker DAS passwords. It is not known whether the background images introduced other types of predictable behaviour such as targeting similar areas of the images or image-specific patterns. Gao et al. [46] proposed

Figure 1: Draw-A-Secret [58]

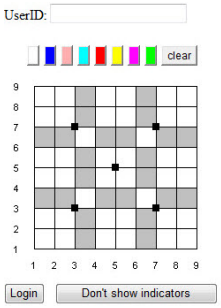

Figure 2: Pass-Go [116]

YAGP (Yet Another Graphical Password), a modification to DAS where approximately correct drawings can be accepted, based on Levenshtein distance string matching and "trend quadrants" looking at the direction of pen strokes. As consequences of this approximation algorithm, a finer grid may be used, but the original password must be stored in a system-accessible manner (rather than hashed) to allow for comparison with the user's input.

Passdoodle [47, 129] is similar to DAS, allowing users to create a freehand drawing as a password, but uses more complex matching process without a visible grid. The use of additional characteristics such as pen colour, number of pen strokes, and drawing speed were suggested to add variability to the doodles. Later, Govindarajulu and Madhvanath [51] separately proposed a web-based password manager using a "master doodle" instead of a master password.

The three Passdoodle studies focus on users' ability to recall and reproduce their doodles, and on the matching algorithms used to identify similar entries. While usability metrics such as login times or success rates are not reported, the scheme would likely require training of the recognition algorithm during password creation, to build an accurate model of the password. Passdoodle passwords (the drawings themselves or a characterization thereof) must apparently be stored in a manner accessible to the system, as opposed to hashed, since the recognition algorithm requires both original and entered doodles to test if they are sufficiently similar.

Weiss and De Luca [134] proposed PassShapes, a similar system. Passwords are translated into alphanumeric characters based on 8 stroke directions, recognized at 45° intervals. During login, PassShapes can be drawn in a different size or location on the screen and still be translated into correct output provided the stroke direction is accurate. The password space is reduced since only 8 possible choices can be made with each stroke, giving a theoretical password space of size similar to PINs if the number of strokes is similar to the number of digits in a PIN. Lab-based studies show that memorability and login times for system-assigned 7 stroke passwords are acceptable according to the authors, but no security analysis has been reported.

The Pass-Go scheme (see Figure 2) designed by Tao and Adams [116] was motivated by an expected DAS usability issue: the difficulty of accurately duplicating sketches whose lines cross near grid lines or grid line intersections. It is named for the ancient board game Go, which involves strategically placing tokens on the intersection points of a grid. In Pass-Go, users draw their password using grid intersection points (instead of grid cells in DAS). The user's movements are snapped to grid-lines and intersections, eliminating the impact of small variations in the trace. Surprisingly, Pass-Go is the only recall-based graphical password system to date for which testing in a field study has been reported. Results of the 167 participant study showed that login success rates were acceptable (as judged by the study's authors) at 78%; no login times were reported. The theoretical password space of Pass-Go is larger than for DAS, due to a finer grid (more squares); allowing diagonal movements (DAS encodes only horizontal and vertical movements); and pen colour as an additional parameter. The designers suggest using a finer grid to further increase the theoretical password space. Users selected longer passwords and used colour, both resulting in greater password complexity than in DAS. Thus in Pass-Go, some dictionary attacks (as explained in Section 9) may be less effective but attacks which exploit patterns [23, 126], for example, remain a concern.

A similar scheme was proposed by Orozco et al. [81], using a haptic input device that measures pen pressure while users draw their password. Although intended to help protect against shoulder-surfing (an observer would have difficulty distinguishing variances in pen pressure), their user study showed that users applied very little pen pressure and hardly lifted the pen while drawing. The differences were so small that the use of haptics did not increase the difficulty of guessing passwords. Por et al. [88] proposed modifying Pass-Go to include background images to aid memorability, optionally highlighting the user's input to facilitate password entry at times when shoulder-surfing is not a threat, and adding decoy input traces to confuse an observer.

GrIDsure [52], a commercial product, displays digits in a 5×5 grid. Users select and memorize a pattern consisting of an ordered subset of the 25 grid squares, and enter the corresponding digits therein using a keyboard. On subsequent logins, digits are randomly displayed within the grid cells and users enter the new sequence of digits found within the cells of their memorized pattern. The system must store the user's pattern itself in a recoverable manner (i.e., storing it as the equivalent of a password, rather than a hashed password) to allow verification of the user's input, which will vary across logins. GrIDsure was user-tested on PDAs brought to participants' home or work locations [19]. With passwords of length 4, users achieved a login success rate of 87% on first attempt. Of the subset of participants taking part in two studies, two years apart, 12% were able to recall their password on the first attempt. Initial security analysis by Weber [132] reported grIDsure passwords as much more secure than traditional PINs, especially against shoulder-

surfing. Independent analysis by Bond [18] notes several weaknesses.

A grid-based system resembling a mini Pass-Go has also been deployed commercially for screen-unlock on Google Android cell phones. PatternLock [114], a similar system, is available for the Blackberry. Rather than entering a 4-digit PIN, users touch-draw their password on a 3×3 grid. The Android screen-unlock scheme has been shown to be susceptible to "smudge attacks" [7], where attackers can determine a user's password through the finger smudges left on the smart phone's surface.

These later recall schemes offer design and understanding beyond DAS. In particular, BDAS suggests that it might be possible to influence the user to select stronger passwords than they might otherwise. Also, the Pass-Go variant was implemented and tested in user studies, with results supporting its usability in practice; a comparison with the memorability of text passwords remains to be done.

5. RECOGNITION-BASED SYSTEMS

Recognition-based systems, also known as *cognometric systems* [31] or *searchmetric systems* [94], generally ask users to memorize a portfolio of images during password creation, and then recognize their images from among decoys to log in. Humans have exceptional ability to recognize images previously seen, even those viewed very briefly [77, 109]. From a security perspective, such systems are not suitable replacements for text passwords, as they have password spaces comparable in cardinality to only 4 or 5 digit PINs (assuming a set of images whose cardinality remains reasonable, with respect to usability). Proposed recognition-based systems use various types of images, most notably: faces, random art, everyday objects, and icons. Renaud [94] discusses specific security and usability considerations, and offers usability design guidelines for recognition-based systems.

Phishing attacks are somewhat more difficult with recognition-based systems as a correct set of images must be presented to the user before password entry. In schemes were the image cues are identical with each login, attackers can retrieve the images beforehand. In schemes with variant responses, only a portion of the user's secret is exposed on any one login attempt, so multiple server probes are necessary. Alternatively, a man-in-the-middle (MITM) attack can be used (see Section 9), where the phishing site relays information between the legitimate site and the user in real-time; the phishing site would get the user to enter a username, pass this to the legitimate site, retrieve the panel of images and display these to the user on the phishing site, then relay the user's selections to the legitimate site. Thus the attacker gains access to the user's account. While somewhat more involved than phishing attacks on recall-based schemes, similar MITM attacks can be launched against all recognition-based schemes discussed in this section.

Shoulder-surfing is of particular concern in recognition-based systems when an attacker can record or observe the images selected by users during login. This is especially problematic for this category of schemes because there are relatively few images (indeed, the theoretical password space is small) and the images selected by users are large discrete units that may be more easily identifiable. Consequently, many recognition-based schemes have specific mechanisms to address this threat. For example, in many systems users perform some action based on the location of their portfolio images within a panel of images, without directly selecting their images. Varying the presented location of portfolio images, as determined by the system, creates a form of *challenge-response* system. In such cases, an attacker would need to observe several (ideally, many) successful logins by a user to gather enough information to correctly deduce sufficiently many portfolio images for a later fraudulent login. Screen scraping malware would similarly require multiple login observations. Shoulder-surfing resistant approaches are often more time consuming and have additional usability costs because they require more effort from users.

In some graphical password schemes, the system must retain knowledge of some details of the shared secret, i.e., user-specific profile data. In recognition schemes, the system must know which images belong to a user's portfolio in order to display them. This information must be stored such that its original form is available to the system (possibly under reversible encryption), and is thus available to anyone gaining access to the stored information. Attackers with access to system-side files may acquire user-specific images or equivalent identifying information. This is true for all recognition-based systems described in this section and may also apply to any scheme requiring that the system retains direct knowledge of the shared secret.

5.1 Canonical Example: Passfaces (and Face)

The recognition-based system studied most extensively to date is Passfaces [84]. Users pre-select a set of human faces (see Figure 3). During login, a panel of candidate faces is presented. Users must select the face belonging to their set from among decoys. Several such rounds are repeated with different panels. For successful login, each round must be executed correctly. The set of images in a panel remains constant between logins, but images are permuted within a panel, incurring some usability cost. The original test systems had $n = 4$ rounds of $M = 9$ images per panel, with one image per panel from the user portfolio. The user portfolio contains exactly 4 faces, so all portfolio images are used during each login. The theoretical password space for Passfaces has cardinality M^n, with $M = 9$, $n = 4$ yielding $6561 \approx 2^{13}$ passwords.

In a study with 77 users, Valentine [124] found that people could remember their Passfaces password over extended periods of time, with login success rates between 72% and 100% by the third attempt for various time intervals up to 5 months. The 34-user, 10-week field study of Brostoff and Sasse [20] found mixed results. While users made fewer login errors (95% success rate for Passfaces), they tended to log in less frequently than users with text passwords because the login process took too long (although no login times are reported).

Davis et al. [30] conducted a 16-week field study where students used one of two graphical password schemes to access class material: Face (their own version of Passfaces), and Story (see further below). They found that users selected predictable passwords that could be successfully guessed by attackers with little effort, as detailed in Section 9. To avoid this problem, a commercial Passfaces product [84] uses system-assigned portfolios that users memorize during an initial training process.

None of the above studies reports password creation time. The Passfaces corporate website [84] says that password creation takes 3-5 minutes for a panel of 9 faces and 5 rounds.

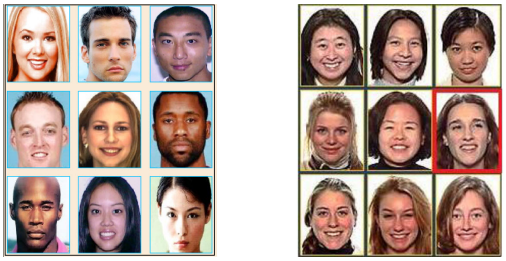

Figure 3: Passfaces system. Left: sample panel from the original system [30]. Right: panel with decoys similar to the image from the user's portfolio [39].

Dunphy et al. [39] investigated whether Passfaces could be made less vulnerable to social engineering attacks where attackers convince users to describe the images in their portfolio. They found that in 8% of 158 login attempts, participants could log in based on verbal descriptions of the images. They further found that participants were less likely (statistically significant) to correctly identify the portfolio image within a panel when decoys were strategically selected to be similar to the portfolio image. Alternatively, social engineering attacks could prompt users to take photographs or screenshots of their images for sharing, especially since all portfolio images are revealed with each login.

Comparing shoulder-surfing risks between Passfaces, text passwords, and PINs in a lab study, Tari et al. [117] found that Passfaces using keypad entry rather than a mouse was significantly less vulnerable to shoulder-surfing than even text passwords or PINs. If Passfaces uses a keyboard for password entry, then malware attacks would need both a keystroke logger and screen scraping software to gain enough knowledge for password entry; with regular mouse entry, only a screen scraper is needed. For further resistance against shoulder-surfing, Dunphy et al. [37] proposed and tested a version of Passfaces using eye-gaze as input at a simulated ATM machine. After initial "play" and "enrollment" phases, they found that participants improved in their ability to enter their passwords over time and that login took an average of 20 seconds for passwords consisting of 5 panels of 9 faces.

Everitt et al. [41] evaluated Passfaces for multiple password interference in a 5 week study where users received email prompts asking them to log on to 4 different fictitious "accounts" according to different schedules. Those who logged in more frequently and those who practiced each new password individually for several days in succession were more successful at remembering their passwords.

5.2 Other recognition-based schemes

Story (see Figure 4) was proposed by Davis, Monrose and Reiter [30] as a comparison system for Face. Users first select a sequence of images for their portfolio. To log in, users are presented with one panel of images and they must identify their portfolio images from among decoys. Story introduced a sequential component: users must select images in the correct order. To aid memorability, users were instructed to mentally construct a story to connect the everyday images in their set. In the test system, a password involved selecting a sequence of 4 images from a panel of 9 images, for a full password space of $9 \cdot 8 \cdot 7 \cdot 6 = 3024 \approx 2^{12}$ passwords.

Story was user-tested along with Face in a field study.

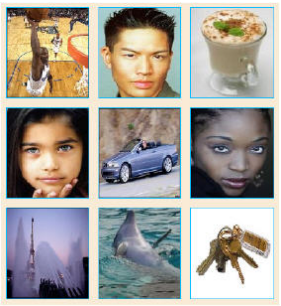

Figure 4: Story system [30].

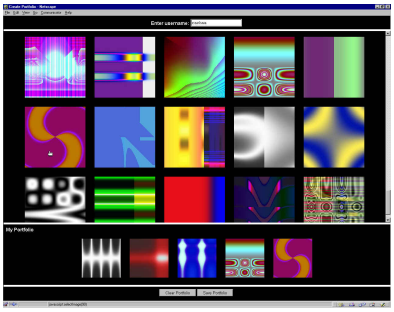

Figure 5: Déjà Vu [33].

The authors [30] found that user choices in Story were more varied but still displayed exploitable patterns, such as differences between male and female choices. Users had more difficulty remembering Story passwords ($\approx 85\%$ success rate) and most frequently made ordering errors. Surveys with participants revealed that they were unlikely to have formulated a story as a memory aid, despite the designers' intentions; this may explain the high number of ordering errors. Different instructions or more user experience might possibly result in greater usage of a story strategy.

In Déjà Vu [33] (see Figure 5), users select and memorize a subset of "random art" images from a larger sample for their portfolio. To log in, users must recognize images belonging to their pre-defined portfolio from a set of decoy images; in the test system, a panel of 25 images was displayed, 5 of which belonged to the user's portfolio. Users must identify all images from their portfolio and only one panel is displayed. Images of random art are used to make it more difficult for users to write down their password or share it with others by describing their images. The authors suggest that a fixed set of 10000 images suffices, but that "attractive" images should be hand-selected to increase the likelihood that images have similar probabilities of being selected by users.

The theoretical password space has $\binom{N}{M}$ passwords, for N images in the panel, and M portfolio images shown. For example, $\binom{25}{5} = 53130 \approx 2^{16}$. Déjà Vu was asserted [33] to be resistant to dictionary attacks because few images in the user study were selected by more than one user. This claim remains to be rigorously tested. Participants found it difficult to describe their portfolio images and those with the same image gave different descriptions from each other. This may stop social engineering attacks trying to gather enough information to log in by tricking the user to verbalize a pass-

Figure 6: Cognitive Authentication scheme [133].

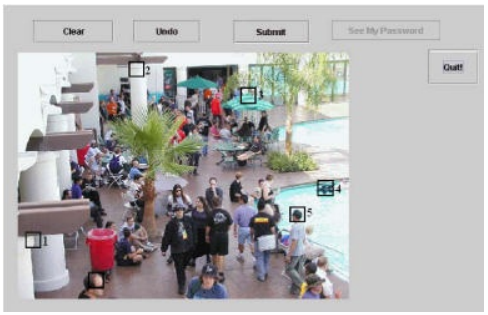

Figure 7: PassPoints password example [139]. The 5 numbered boxes (not ordinarily visible to users) illustrate the tolerance area around click-points.

word. Similarly, it would seem difficult to identify images belonging to a particular user based on knowing other information about that user; however, problems resulting from predictable user choice remain possible, such as users selecting images that include their favourite colour.

Weinshall [133] proposed the Cognitive Authentication scheme (see Figure 6) intended to be safe against spyware and shoulder-surfing. Keyboard input is used rather than a mouse and users must recognize images from their previously memorized portfolio. The login task involves computing a path through a panel of images starting from the top-left corner, based on whether particular images belong to the user's portfolio: move down if you stand on a picture from your portfolio, move right otherwise. On reaching the panel's right or bottom edge, identify the corresponding label for that row or column. A multiple-choice question is presented, which includes the label for the path's correct end-point. Users perform several such rounds, each on a different panel. After each round, the system computes the cumulative probability that the correct answer was not entered by chance. When the probability passes a certain threshold, login succeeds. This tolerates some user error. If the threshold is not passed by a certain number of rounds, the login fails.

Users receive a system-assigned portfolio containing a large number (about 100) of randomly chosen images, and extensive initial training to memorize it. No times are reported for this training phase. Average login time is 1.5 to 3 minutes. In a user study with 9 participants, a 95% login success rate is reported, with users logging in over a period of 10 weeks.

Although the main claim [133] of resisting shoulder-surfing was proven false [48] (see Section 9), the scheme offers interesting lessons. The number of different passwords possible from a user's viewpoint is $\binom{N}{M}$, based on unique collections of images. N is the number of images in a panel, M the number of portfolio images displayed; N=80, M=30 gives $\binom{80}{30} = 2^{73}$ passwords. However, the redundancy which encodes the user's portfolio images into row and column labels apparently results in a many-to-one mapping of image sets onto system passwords, reducing the password space. For example, for exactly 5 rounds and 4 different multiple choice answers, there are $4^5 = 2^{10}$ distinct system passwords. Dictionary and personalized attacks have no advantage over exhaustive attacks, due to the random assignment of images. It appears impossible to verbalize enough information to convey a password to an attacker to allow successful login, making such social engineering attacks also improbable.

Other recognition-based systems have been proposed, with similar usability and security profiles as those above. We therefore mention them only briefly. In the VIP system [31, 71], a panel of images is displayed. Users must select images from their portfolio among decoys. Different configurations allow for multiple rounds or sequencing of images. In the Photographic Authentication system [86], users initially provide their own set of digital photos and must identify these from among decoys, with panels of 4 images, and 10 rounds. The decoy images are randomly selected from the images collected from other users. Use Your Illusion [54] also requires that users select portfolio images from panels of decoys; the selected images are distorted after original selection. The idea is that the legitimate user can still recognize the images despite distortion, while the distortion creates difficulties for others. The distortion is intended to protect against social engineering and shoulder-surfing attacks. In the Convex Hull Click Scheme [140], users select and memorize a portfolio of images, and must recognize these images from among decoys displayed, over several rounds. The images are small icons and several dozen are randomly positioned on the screen. Each panel contains at least 3 of the user's icons. Users must identify their icons, visualize the triangle they form, and click anywhere within this triangle. This design is intended to protect against shoulder-surfing, but comes at a cost of longer login times. In Bicakci et al.'s [14] GPI (Graphical Password with Icons) and GPIS (Graphical Password with Icons suggested by the System) systems, users log in by selecting their 6 icons, in order, from a panel of 150 icons. The theoretical password space of these two schemes is similar to most cued-recall schemes at 2^{43} (see below). The two systems differ only in how passwords are set. GPI allows users to choose any 6 icons as their password. In GPIS, passwords are suggested by the system but users may shuffle until they find an acceptable password, reducing (but not eliminating) problems with user choice.

Renaud [95] ran a field study comparing different types of user involvement in selecting portfolio images for recognition-based schemes. Users could select images from a photo archive, take their own photos, or draw doodles that were subsequently scanned and converted to JPEG format. Results show a significant increase in login success rates when user portfolios contain self-drawn doodles rather than either type of photos. The memorability improvements, however, need to be balanced with the additional risk of personalized attacks if attackers know a user's drawing style or recognize personally-identifiable features within the doodles.

An important feature in these schemes is the challenge-response approach where users are presented with a panel of images and must respond based on knowledge of a shared secret. In the simplest case, users select their portfolio images directly, while other schemes require additional mental processing to identify the correct response. Most of these early recognition-based schemes compromise between the size of the theoretical password space and usability in terms of memorability and login time. As proposed, most schemes offer a password space comparable to a 4-digit PIN which, while useful in some environments, does not offer a substitute (with respect to security) for common text passwords. Everitt et al.'s [41] study of interference in Passfaces is a positive step in understanding multiple password interference in recognition-based schemes. Further work is needed to better understand whether exposure to multiple sets of portfolio and decoy images increases chances of memory interference over time, especially as decoys become familiar.

6. CUED-RECALL SYSTEMS

Cued-recall systems typically require that users remember and target specific locations within an image. This feature, intended to reduce the memory load on users, is an easier memory task than pure recall. Such systems are also called *locimetric* [31] as they rely on identifying specific locations. This memory task differs from simply recognizing an image as a whole. Hollingworth and Henderson [56] show that people retain accurate, detailed, visual memories of objects to which they previously attended in visual scenes; this suggests that users may be able to accurately remember specific parts of an image as their password if they initially focused on them. In an ideal design, the cue in an authentication system is helpful only to legitimate users (not to attackers trying to guess a password).

Cued-recall graphical password systems date back to Blonder's patent [17]. PassPoints, its successor, launched research in the cued-recall subclass sometimes called *click-based graphical passwords.*

The schemes discussed below share a vulnerability to shoulder-surfing and malware, and are vulnerable to MITM phishing attacks similar to recognition-based schemes. To capture a click-based graphical password using malware, a mouse-logger may suffice if the attacker can also determine the position of the image on the screen. Alternatively, a screen scraper may identify the image location, and be sufficient if the attacker can identify when the user clicked the mouse button (some users very familiar with their password may not necessarily stop moving the cursor while clicking). Shoulder-surfing may also reveal a user's password in a single login, as the entire password may be observable on the screen as the user enters it.

6.1 Canonical Example: PassPoints

The literature on cued-recall graphical password systems is dominated by PassPoints [137–139] and its variations. A password is a sequence of any $n = 5$ user-selected click-points (pixels) on a system-assigned image (see Figure 7). The user selects points by clicking on them using a mouse. During login, re-entry of the click-points must be in the correct order, and accurate within a system-specified tolerance. The image acts as a memory cue to the location of the originally chosen click-points. Note that this is not an optimal cued-recall scenario: users see only one cue, but must recall 5 pieces of information, in the correct order. The standard parameterization provides a theoretical password space of 2^{43} conceivable passwords; this increases with larger n and smaller tolerance, though usability impacts are expected.

An important implementation detail is the type of *discretization* used — this is related to how the system determines if entered click-points are acceptably close to the original points, and affects whether the system-side passwords stored for verification can be hashed. *Robust discretization* [16], *centered discretization* [25], and *optimal discretization* [13] are possible alternatives. Kirovski et al. [63] suggest how discretization could be implemented using Voronoi polygon tiling by analyzing image features and centering likely click-points within the polygons.

Wiedenbeck et al. [137–139] conducted three lab-based user studies of PassPoints. Users took 64 seconds to initially create a password, and required an additional 171 seconds of training time on average to memorize their password. Login took between 9 and 19 seconds on average. Login success rates varied from 55-90%, with users returning at different intervals to log in again. User performance was found to be similar on the four images tested, and it was recommended that tolerance areas around click-points be at least 14×14 pixels for acceptable usability. Chiasson et al. [21] conducted a lab study and a field study, finding that image choice does impact usability, that tolerance areas could be further reduced, and that memory interference from remembering multiple PassPoints passwords may be problematic. When explored further, memory interference was shown to be less problematic for PassPoints passwords than text passwords [24]. Later security analyses found it to be vulnerable to hotspots and simple geometric patterns within images [23, 36, 49, 100, 121, 127], as elaborated in Section 9. Bicakci et al. [15] conducted a lab study where PassPoints was used as the master password for a web-based password manager and concluded that it was more usable than an alphanumeric master password. Their implementation used a visible grid dividing the image into discrete sections rather than any of the aforementioned discretization methods.

A commercial version of PassPoints for the PocketPC is available from visKey [104] for screen-unlock by tapping on the correct sequence of click-points using a stylus or finger. Users may define settings such as n, the size of the tolerance regions, and which image is displayed.

6.2 Other cued-recall variants

PassPoints has received attention from others, who have proposed modifications. Suo [112] proposed a shoulder-surfing resistant version as follows. During login, the image is blurred except for a small focus area. Users enter Y (for yes) or N (for no) on the keyboard, or use the right and left mouse buttons, to indicate if their click-point is within the focused area. The process repeats for at most 10 rounds, until all 5 click-points are identified. We note as the user's click-points are guaranteed to be within the 10 focus areas, observing one login narrows the search space considerably, and observing a few logins would allow password recovery.

Cued Click-Points (CCP) [27] is a click-based scheme where users select one click-point on each of 5 images presented in sequence, one at a time; this provides *one-to-one cueing.* Each image after the first is a deterministic function of the current image, the coordinates of the user-entered click-point, and a user identifier. Users receive immediate

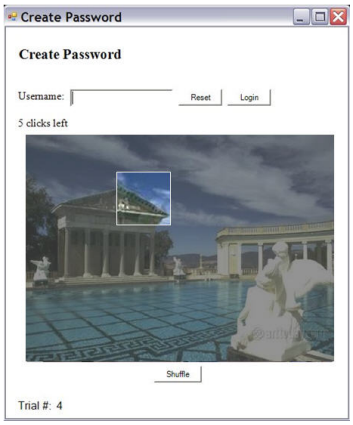

Figure 8: Persuasive Cued Click-Points. During password creation, users select a click-point from the highlighted viewport or press the shuffle button to relocate the viewport.

Figure 9: Inkblots from the Inkblot Authentication user study [111].

feedback if they enter an incorrect click-point during login, seeing an image that they do not recognize. At this point they can restart password entry to correct the error. This *implicit feedback* [27] is not helpful to an attacker not knowing the expected image sequence.

In a lab-based user study [27] of CCP, users successfully logged in on the first attempt, without errors or restarts, in 96% of trials. On average, participants took 25 seconds to create a password, and 7 seconds to login. Analysis of user choice revealed that users tended to select click-points falling within known hotspots [22], but that simple patterns of click-points were eliminated (cf. PassPoints above) [23].

Persuasive Cued Click-Points (PCCP) [22] is a variation of CCP designed to persuade users to select more random passwords. It functions like CCP, but during password creation the image is dimmed except for a small square viewport area randomly positioned on the image. Users select a click-point from within this viewport (see Figure 8), or may press a "shuffle" button to randomly reposition the viewport until a suitable location is found. On subsequent logins, images are displayed in their normal format with no dimming or viewport. Common wisdom that users choose the path-of-least-resistance here means selecting a click-point within the first or first few viewports. The design intent of the viewport is to flatten the distribution of click-points across multiple users, reducing hotspots.

In a lab study [22], login success rates were similar to CCP. Participants took 50 seconds on average to create a password (an increase mainly due to participants who shuffled repeatedly, though most shuffled relatively infrequently), and 8 seconds to log in. A later two-week study [110] comparing PCCP configured with different image sizes and numbers of click-points, found both manipulations had similar effects on usability. PCCP reportedly [23] removes major concerns related to common patterns and hotspots.

As mentioned earlier, proposed implementations of PassPoints, CCP, and PCCP use a grid-based discretization algorithm to determine whether login click-points are within tolerance. In system-side storage for verification, these passwords can be hashed; additional information such as a grid identifier (for each click-point), however, is stored in a manner accessible to the system, to allow the system to use the appropriate grid to verify login attempts. It is unclear if attackers gaining access to the server-side storage can use these grid identifiers to their advantage.

Inkblot Authentication [111] (see Figure 9) is not strictly a graphical password system, but uses images as a cue for text password entry. During password creation, users are shown a series of computer-generated "inkblots" and asked to type the first and last letter of the word/phrase that best describes the inkblot. The letter pairs form the password. The inkblots are displayed in shuffled order as cues during login, and users enter each of their 2-character responses. The same shuffled order is used for each subsequent login. It was suggested that with time, users would memorize their password and would no longer need to rely on the inkblots as cues. Twenty-five users in a lab study were presented with 10 inkblots and created a corresponding password. After one day, 80% of users entered their entire password correctly; 72% were successful after one week. With only one exception, when users made mistakes, it was on only one of their 10 character-pairs. The resulting passwords were relatively strong (20 characters long with no recognizable words; although some letters were more popular than others). It is claimed that inkblots are abstract enough that an attacker seeing the inkblots would not have an advantage in guessing a user's password.

Similarly, Jiminy [96, 97] is a graphical tool for remembering text passwords. A grid of alphanumeric characters is placed over an image and users are provided with coloured templates that contain several openings. To log in, users select the appropriate template, "anchor" it to the correct location on the image, then enter the sequence of characters visible through the openings. Instead of remembering their text password, users remember the position of the template on the image. Several users in paper-based and web-based studies selected the same anchor points, indicating that the security impact of hotspots in this scheme is in doubt.

Alsulaiman and El Saddik [2] proposed a 3D scheme where users navigate a 3D world and perform actions interpreted as their password. Much like the 2D graphical passwords above, the 3D environment acts as a cue to prompt users to perform their actions. The designers envision that users could perform various actions such as clicking on certain areas, typing or drawing on a virtual surface, supplying a biometric, or interacting with parts of the virtual world (like turning on a light switch). A prototype system implements a small portion of the scheme (users can walk through a virtual art gallery and enter text passwords at virtual computers or select pictures as part of a graphical password). Detail

about other proposed components is conceptual only. No user testing or security results are reported, making usability or security evaluations difficult.

While some analysis of the above schemes can be done using standard statistical tests, occasionally novel or specialized approaches are required. For example, in click-based graphical passwords, analysis of the two-dimensional space is desirable to identify patterns in user behaviour. Conventional statistics do not apply, but point pattern analysis [35] from spatial statistics has been used [22] to evaluate and compare clustering of click-points on images.

With click-based graphical passwords, as well as other types of authentication schemes, getting an accurate measure of the effective password space remains a challenge when user choice is involved. One alternative approach is to evaluate whether the set of passwords (or password components) selected by users is representative of the full theoretical password space T. A Monte Carlo approach can determine the likelihood that a particular set of passwords occurred by chance (and thus is similar to a random set taken from T). With Monte Carlo methods, randomly generated datasets are used to identify the range of probable analytical measures which can then be compared to those based on datasets collected from actual usage. This approach has been used to compare models of the effective password spaces for PassPoints, CCP, and PCCP [23].

In summary, early cued-recall schemes, such as PassPoints, offered promise as alternatives to text passwords due to their large theoretical password space and short login times. However, analysis revealed reduced security due to hotspots and simple patterns in user selection of click-points. Later schemes, such as PCCP, aim to explicitly address these issues without resorting to system-assigned passwords, and have introduced other features such as implicit feedback, and graphical aids for text passwords that might benefit other next generation authentication schemes as well.

7. SUMMARY OF GRAPHICAL PASSWORD SYSTEMS

Tables 1–3 summarize the 25 graphical password schemes discussed. Cells containing *i.d.* indicate insufficient detail available in the literature to evaluate the scheme on this criterion; asterisk (*) indicates our best estimate based on available information; dagger (†) represents an approximation based on reported figures; dash (–) indicates that to our knowledge, no published results are available.

We identified four measures to summarize security. First, we labeled schemes based on theoretical password space for the configurations most commonly reported or used in any user testing, according to three ranges: under 20 bits (*PIN-level*, ○), 20 to 60 bits (*password-level*, ◑), and over 60 bits (*crypto-level*, ●). It remains open to debate exactly what size of theoretical password space is sufficient in different threat environments — for example, some suggest that 2^{20} suffices for online environments with lock-out rules [45]. Next, schemes were rated on the degree to which user choice issues may weaken security, e.g., allowing more efficient dictionary attacks. Schemes known to suffer from skewed password distributions related to user choice were rated weakest (◇), and those with system-assigned passwords rated strongest (◆). Those attempting to influence user choice towards more secure options were rated ◈. In schemes allowing two deployment options (user-chosen or system-generated passwords), the table cell includes a rating for each. In the "variant response" row, 'yes' indicates a non-static response (e.g., different parts of a password portfolio are entered across login instances), and a simply recorded response replayed later will fail. In PassFaces, a constant set of images are permuted within a panel but users select the same image; we mark such schemes 'no' for variant response. Ideally in variant response schemes, shoulder-surfing or key-logging requires more than one login instance to recover a password equivalent. Finally, we examined the number of probes to the legitimate server an attacker requires to prepare a phishing attack, i.e., to acquire the images or cues needed to extract a password from a user. A probe means one login instance with the legitimate server and does not involve interaction with the user (e.g., no probes are needed for DAS as an attacker need only display a drawing grid on a phishing site; for PassPoints, one probe is needed to retrieve the image for a specified userid).

For usability, we summarize the types of user studies that have been conducted, note whether memory interference from multiple passwords has been studied ('yes' indicates at least one study examined interference), and include two of the most commonly reported usability measures: login time and success rate, tabulating the range of results across the reported studies for a given scheme. When available, login success rates on the first attempt are reported. For the user studies, we identify the number of sessions and the duration of study in weeks. For example, two lab studies have been conducted for PassShapes (Table 1)—the first consisted of a single lab session, the second of 3 sessions spread over 1.5 weeks. A row related to hybrid studies is omitted from the tables, since to our knowledge no hybrid studies have been reported in the literature for any of the schemes.

8. USABILITY ASPECTS

This section is based on an examination of the literature reporting results of usability testing of graphical password systems. As there has been essentially no coordinated work towards an accepted standard for evaluating the usability of graphical password schemes, nearly every system evaluated, has been tested (if at all) using different criteria. This makes comparison difficult at best. Even when apparently similar measures are reported, they have often been calculated using different methods and may represent completely different measures. In what follows, we provide context and offer specific recommendations intended to facilitate comparisons of such schemes in the future. Some of the observations are common knowledge to human-computer interaction (HCI) experts, but are either not widely practiced, or widely acknowledged in the graphical password literature to date.

8.1 Target Users

Characteristics of the intended users must be taken into account when designing or selecting an appropriate graphical password scheme. The expertise level of target users may dictate the acceptable complexity of the interaction, and the level of training required or expected. The frequency of use may also have a significant influence on usability. Frequently accessed systems should be quick to use, and may rely more heavily on users' memory, as frequent repetition aids memory. If passwords are used for infrequently accessed systems, they must be especially memorable since memory decays

Table 1: Recall-based systems (summary).

Scheme	Android screen unlock	GrIDsure	PassShapes	DAS	BDAS	PassGo	YAGP	Haptic password	Passdoodle
Theoretical space (bits)	○ 18	○ 18	◑ 21	◑ 58	◑ 58	● 77	● 300	*i.d.*	*i.d.*
User choice resilience	*◇	*◇	◆	*◇	*i.d.*	◇	*◇	*◇	*◇
Variant response	no	yes	no	no	no	no	no	no	no
Server probes	0	0	0	0	1	0	0	0	0
Paper study	–	–	3×/1.5wk	1× 2×/1wk	2×/1wk 2×/1wk		–	–	2×/1wk
Lab study	–	≥ 11wk ≥ 1.5wk	3×/1.5wk	–	–	–	3×/2wk	1×	1×
Field study	–	–	–	–	–	13wk	–	–	–
Web study	–	–	–	–	–	–	–	–	–
Login time	–	–	6s	–	–	–	–	–	–
Success rate	–	87%	63-100%	57-80%	50-80%	78%	87-96%	–	38-46%
Interference studied	–	–	–	–	–	–	–	–	–

Table 2: Recognition-based systems (summary).

Scheme	Cognitive Authentication	Use Your Illusion	Story	Passfaces / Face	VIP (type 1)	Déjà Vu	Photographic Authentication	Convex Hull Click	GPI / GPIS
Theoretical space (bits)	○/● 10/73	○ 11	○ 12	○ 13	○ 13	○ 16	◑ 20	◑ 32	◑ 43
User choice resilience	◆	*i.d.*	◇	◇ / ◆	◆	*◇	*◇	*◇	*◇/*◆
Variant response	yes	no	no	no	no	no	yes	yes	no
Server probes	many	1	1	1	1	1	many	many	1
Paper study	–	–	–	–	–	–	–	–	–
Lab study	13×/ 10wk	4×/4wk	–	– –	2×/1wk 3×/4wk	2×/1wk	1×	2×/1wk	2×/1wk
Field study	–	–	≥16wk	≥16wk 10wk	≥16wk	–	–	–	–
Web study	–	–	–	1-5mth 5wk	–	–	–		–
Login time	90-180s	12-26s	–	14-88s	5-†6s	32-36s	†40s	72s	†18s/†19s
Success rate	>95%	89-100%	†85%	72-100%	†11-95%	90-100%	†95-100%	90%	83%/74%
Interference studied	–	–	–	yes	yes	–	–	–	–

Table 3: Cued-recall systems (summary).

Scheme	Jiminy	Suo's scheme	PassPoints	CCP	PCCP	Inkblot Authentication	3D scheme
Theoretical pswd (bits)	○ 9	○ / ◑ 16 / 43	◑ 43	◑ 43	◑ 43	● 94	*i.d.*
User choice resilience	*◇	*◇	◇	◇	*◈	*◈	*i.d.*
Variant response	no	yes	no	no	no	no	*i.d.*
Server probes	1	*i.d.*	1	many	many	1	*i.d.*
Paper study	2×/4 wk	–	–	–	–	–	–
Lab study		–	2×/1wk 3×/6wk 1× 2×/2wk	1×	1× 2×/2wk	3×/1wk	–
Field study	12wk	–	7-9wk	–	–	–	–
Web study	–	–	–	–	–	–	–
Login time	–	–	9-25s	7s	11-89s	–	–
Success rate	†47-73%	–	38-94%	96%	83-94%	†68-80%	–
Interference studied	–	–	yes	–	–	–	–

over time. Issues of accessibility may arise since different user populations, such as the elderly [93], have different requirements. Many of the systems we have discussed implicitly require users with good vision, potentially including good colour vision (for recognizing cues), and good motor skills (for entering sketches or accurate clicks on an image). Design of graphical password systems therefore needs to either address these issues, provide alternatives, or be very aware of the limitations they impose on who will be able to successfully use the software. Because authentication systems by their nature act as gate-keepers to computer systems and services, these issues must be taken very seriously and should be addressed in proposals for new schemes.

8.2 Tasks

Ease of login is the most frequently examined task, but is only one of many. Ideally, usability should be explored along several dimensions. For usability, essential elements to measure and report include: time to create a password, and time to login; memorability (typically through success rates and number of errors made during login over an extended period); and interference, by testing with a normal password load (as opposed to with only one password at a time).

8.2.1 Password Initialization

Authentication systems require initialization. A graphical password can either be assigned or user-selected. Training may be conducted, in part to compensate for the novelty of a scheme relative to well-known approaches like text passwords. Password confirmation is usually involved to ensure that users have not made trivial entry errors, and can accurately remember and enter their password after a short time before testing longer term memorability.

Allowing users to select their own password can aid usability since a password having personal meaning may be easier to remember. However, this design decision has security disadvantages. As discussed later, graphical password systems that suffer from predictability problems due to user choice include the canonical examples of all three main categories: Passfaces, DAS (Pass-Go), and PassPoints. For example, from their study of Face and Story, Davis et al. [30] conclude that user choice leads to predictable patterns that may exploited by attackers.

Allowing user-chosen passwords can also encourage password reuse across accounts. Despite obvious usability advantages (e.g., reduced memory load, and no need to think of new creative passwords for each new account), password reuse implies that an attacker who gains access to an account on a weakly protected system may then have sufficient information to log in to that user's higher value accounts. If permitted, users often reuse passwords verbatim; Florencio et al. [44] found that text passwords are reused on an average of 6 different accounts. Many users also form some common strategy or pattern across accounts [1]. Both situations may be exploited by an attacker who acquires one of the passwords.

Systems which assign randomly selected passwords preclude attacks exploiting predictability, and also eliminate the potential for cross-account password reuse. However, such systems may require time-consuming training to help users remember their passwords (e.g., recall Weinshall [133]). Even with training, such passwords may remain more difficult to remember since opportunities for leveraging are removed. In the Passfaces study of Everitt et al. [41], which assigned passwords to avoid the predictability seen in earlier Passfaces studies, the order of password acquisition and login frequency significantly impacted password memorability. Allowing users to use their own images may improve memorability and encourage positive affective responses [69], but predictability and personalization may weaken security.

It is possible for a system to allow partial user choice in password selection. For example, in PCCP (see Section 6.2),

the middle-ground between allowing user choice and system-assigned passwords led to passwords nearly indistinguishable from random on the measures examined [23]. Further work is needed to evaluate the effect on long-term memorability.

8.2.2 Login

Login should be quick and simple since it is the most common task completed by users of an authentication system. Deviation from this rule may be acceptable under certain circumstances (see section 8.3 below).

Text passwords have an advantage of being ubiquitous, and can be typed in a few seconds or less on standard keyboards. It is thus natural to compare the time to enter a graphical password to that for a text password. Error and success rates on login are the usability measures most often reported in user studies of graphical passwords. Unfortunately, they are often calculated in different ways and measured at different times. For example, some studies consider the trial a success if users can log in within three attempts, while others count only trials that are successful with no errors (i.e., first attempt). To allow comparison, we recommend reporting (at least) success rates for the first attempt and after three attempts, due to the common practice of lockout after three failed attempts.

Memorability issues are important when discussing login performance, as memorability is a main factor determining login success. Measures of memorability address whether passwords can be remembered over short- and long-term and with varying login frequencies. For strategies for testing memorability, see Section 10.1.

Most graphical password studies to date have required users to remember only one password at a time, whereas in real-life users must remember many passwords and may get them confused. In the cognitive psychology literature [5], *memory interference* is "the impaired ability to remember an item when it is similar to other items stored in memory". With authentication, interference occurs when remembering a password for one system impairs the user's memory of a password for another system. This may be of particular concern with graphical passwords since exposure to similar images from multiple concurrent passwords or from password resets may aggravate the problem. Although an important usability concern, published studies [21,24,41,71] evaluating interference from multiple passwords are only now beginning to appear.

8.2.3 Password reset and password change

The tasks of resetting or changing passwords are not typically examined during usability testing of new graphical password schemes, but these are often required in practice when users forget passwords. The process may involve the user interacting only with the system, or may require contact with help desk personnel. Both involve confirming the user's identity through some secondary means, and issuing a new password (which often must be changed immediately on the next login). New text passwords can easily be communicated by phone or through email; graphical passwords cannot be communicated as easily. While this provides protection against some social engineering attacks, it also poses a usability challenge. One solution is to assign temporary non-graphical password during password reset, giving system access to create a new password. Text passwords may also be used as a fall-back if for example some users must, from time to time, log in from legacy systems having text-only interfaces.

System configuration and design of password reset and password change mechanisms can impact memorability, interference, and security of the system. For example, if users are presented with the same, or similar, images as in previous graphical passwords, they may be more likely to confuse the memories of passwords or to reuse passwords. This suggests that reuse of password images should be avoided, and also argues against images being uploaded by users.

Most authentication systems must allow password changes (some systems require this at specified intervals). The usability and security concerns are similar to password reset, except users can complete the task themselves without requiring a temporary password, entering their current graphical password as authentication.

8.2.4 Portable login

Unless restricted to specific environments (e.g., physical presence in a corporate office or at a bank ATM), users of graphical password systems may need to log in from different physical devices or locations. Usability issues to consider include whether the system is suitable for access from devices having different screen sizes or resolutions, and whether local bandwidth constraints impact performance. Moreover, portable login may require a modified login process or completion of additional tasks; these should also be considered and tested.

8.3 Domains

Performance constraints and goals for an authentication system differ depending on the intended environment of use. When presenting a new scheme, the target environment should be clearly declared, to allow comparison of systems intended for similar conditions, and to avoid deploying systems in inappropriate domains.

For high-risk domains such as online banking, security is of utmost importance and it may be acceptable to have a system that is slightly more difficult to use in order to achieve the desired level of security, as long as usability problems do not lead to behaviour triggering other security issues. Conversely, it may be acceptable to have very usable, but lower security schemes for lower risk domains. In fact, this could improve security for high-risk domains if it eliminates the opportunity for password reuse between high- and low-risk systems; it may also help with memorability by reducing chances of password interference. Similarly, infrequently used accounts may be better served by a more memorable scheme that has a relatively long login time if this makes it more likely that the user can log in when needed.

It is unlikely that any single scheme will suit all domains, tasks, and target users, from a combined usability and security viewpoint. Thus, specifying the target environments and applications for newly proposed schemes is important.

9. SECURITY ASPECTS AND ATTACKS

This section discusses standard threats to password-based authentication systems and how they relate to graphical passwords. Attacks are classified as guessing or capture attacks (including malware which captures passwords). We do not discuss attacks which exploit software vulnerabilities to bypass the authentication system entirely, limiting our scope to attacks which directly obtain password credentials.

9.1 Guessing Attacks

Many standard attacks on text passwords convert directly to attacks on some graphical password schemes, once the graphical passwords are encoded in the canonical representation of their password space. For example, since a user response still corresponds to a password, exhaustive search attacks are possible given knowledge of the encoding used (e.g., as binary strings or alphanumeric characters). As such, developers of new graphical schemes must be aware of these, to design against them.

Guessing attacks remain a serious threat [91,103,118], but statistics are scarce (few organizations publicize breaches). An *online guessing attack* requires interaction with the live system; password guesses are entered in turn to see if login succeeds. For graphical as well as text passwords, defenses may be aided by clever use of CAPTCHAs [87]; increasingly delaying (e.g., doubling) system response time on successive incorrect guesses; or limiting, per user account, the number of incorrect login attempts allowed before disabling further attempts. The latter risks locking out legitimate users who forget their password, enables denial-of-service attacks which intentionally provide incorrect passwords, and is less effective against multi-account attacks.

In an *offline guessing attack*, attackers gain access to verifiable text [50] and need not interact with the live system to verify guesses. Schemes vulnerable to offline attack are at higher risk than those requiring online verification, for equivalent password spaces: offline work is not visible, processing trial guesses can be quicker, and pre-computed data structures or special hardware may be used.

Defensive techniques against guessing attacks vary in utility depending on the environment. For most text-based passwords, and some graphical systems (though often unspecified), system-side passwords are stored after processing by a one-way hash function, for added security in case an attacker gains access to this storage. To check if a (userid, password) attempt is correct, the system hashes the password input and compares to the valued stored for that userid. One way to complicate guessing attacks is *iterated hashing* [73], requiring, say, 1000 or many more repeated hashing operations (rather than one); this increases the time to test password candidates online or to pre-compute dictionaries. *Salting* [73] concatenates to a password (before hashing) a user-specific string stored along with the hashed password; this forces hashing for each password guess on a per-salt basis, increasing the cost of pre-computed databases. Most graphical password proposals fail to consider important implementation details such as hashing and salting. Retrofitting such defenses may or may not be possible, depending on design characteristics inherent to individual schemes.

Other defenses, especially important for graphical password schemes subject user choice issues, include *password rules* or policies [73] disallowing weak passwords at creation, encouraging stronger password choices [14, 22], and both reactive and proactive *password checkers* [11, 65]. *System-assigned passwords* are generated randomly to preclude attacks exploiting skewed distributions and use larger portions of the theoretical password space, but have high usability costs: longer training times or increased likelihood that users forget passwords. Mnemonic strategies like Story [30], may improve both usability and security, but often suffer from predictability problems if user choice is allowed.

9.1.1 Exhaustive-search (brute-force) attacks

Exhaustive-search attacks try all elements in a search space, whether representing graphical or text passwords. For user-chosen passwords which are far from equi-probable, dictionary attacks are preferred (see further below) except for small password spaces.

Exhaustive-search optimizations such as Oechslin's *rainbow tables* [80], which trade pre-computation time for storage, are relevant for password cracking. Coarse sequencing optimizations include guessing shorter passwords first (e.g., fewer cell-crossings and strokes in DAS). Fine sequencing optimizations, such as ordering passwords in decreasing expected probability, and favoring subsets expected to hold higher probability passwords [126], are a cross between intelligent brute-force and dictionary attacks.

The advantage to exhaustive offline attacks is that with enough time and computing power, all passwords will be found. However, full search of large password spaces is limited in practice by the time or processing power available; searching only subsets is faster, but doesn't guarantee success. To minimize the threat of exhaustive attacks, the theoretical password space should be too large to search. Note that this is not the case for many recognition-based systems—e.g., the most common configuration of Passfaces has 9-image panels and 4 rounds, yielding only $9^4 = 6561$ passwords. Often such systems require complementary mechanisms such as limiting the number of online guesses per account, or securely combining multiple mechanisms (e.g., TwoStep Authentication [128]). Helping the defender, attacks may require obtaining the image set used, which involves additional effort; the added barrier depends on the size of the image set and the methods required to access it.

9.1.2 Dictionary Attacks and Optimizations

Dictionary attacks on graphical passwords [30, 120] follow a long line of attacks on text passwords (e.g., [43, 65, 73,143,145]). The original idea involved guessing passwords from a relatively short pre-compiled list (*dictionary*) of high-probability candidate password, based on empirical data or assumptions about user behaviour. Massive dictionaries and powerful data structures have created a continuum from small dictionaries to prioritized brute-force attacks, with *smart dictionary attacks* combining time-memory trade-offs of exhaustive attacks with higher success probabilities of prioritized dictionaries, in some cases algorithmically generated [76].

In systems allowing user-choice, dictionary attacks exploit skewed password distributions resulting from certain subsets of passwords being more attractive to non-negligible sets of users. Attacks succeed as users select passwords from predictable, relatively small subsets of the theoretical password space known as *weak password subspaces* [126], which can be enumerated, are small enough to search, and contain a significant fraction of passwords chosen in practice. These are collectively modeled as an *effective password space* including passwords with predicted probabilities higher than some threshold. A theoretical space too large to be exhaustively attacked does not guarantee security; the effective password space must also be too large to search. The challenge here is to understand what composes the effective password space, which remains an open problem even for text passwords. Many graphical password proposals fall to dictionary attacks due to predictable patterns in user choice, as we discuss next.

9.1.3 Guessing Attacks on Specific Graphical Password Schemes

We now discuss some guessing attacks on specific graphical password schemes. Many are adaptations of known strategies discussed above. Replicating all details in a comprehensive review of attacks to date is beyond our scope. Instead, our goal is to both give a brief description of some attack methods used in analysis, and to emphasize that *essentially all new proposals which have been subjected to detailed security analysis have exhibited weaknesses not originally envisioned.* Though disappointing, this follows historical precedent in new security mechanisms; iteratively improving on bold new ideas is often necessary. We focus on results for the exemplars of the three major classes of graphical passwords. Flaws in these raise a red flag for the many others which have seen at best superficial security analysis. Many designers making new graphical password proposals have presented functional details, but no thorough security analysis. We expect this weakness in the literature will be repaired as the research matures. Until then, optimistic but unsupported security assertions by proponents of new schemes are best viewed skeptically.

RECALL-BASED SYSTEMS. In detailed security analysis of DAS and Pass-Go [119, 126], Thorpe used a predictive method built on the reflective symmetry and stroke-count characteristics of passwords, to identify DAS weak password subspaces. Supporting evidence that these subspaces accurately modeled user choice included a small 16-user paper-based DAS study [75], and more convincingly, a 167-user Pass-Go field study [115, 116]. In the Pass-Go study, 40% of users chose passwords falling in a subspace defined primarily by symmetry with respect to a central vertical and horizontal axis (without restrictions on stroke-count), and 72% chose passwords falling in a subspace characterized by 4 or fewer strokes. Similar results were found in the DAS study. The field study also revealed 19% of user-chosen Pass-Go passwords were from a third category (a subspace of about 2^{36} elements), namely, drawings of alphabetic characters or symbols. Populating dictionaries using these subspaces, which range from 2^{31} to 2^{41} elements, allows an exponential decrease in search space vs. the full space of 2^{58} passwords. Thus successful dictionary attacks on Pass-Go (and DAS) require vastly less effort than initially expected or implied by their full password spaces.

RECOGNITION-BASED SYSTEMS. The most prominent security analysis to date for recognition-based systems involved a field study dataset of the Face and Story schemes [30]; passwords in both consisted of four image items. Random subsets containing 80% of the user-chosen passwords were used to construct dictionaries, allowing calculation of how many guesses an attacker following this strategy would take to correctly guess some, or all, of the remaining 20% of dataset passwords. The dictionaries were arranged in nonincreasing order of probability, with the probability distribution over both user choice (as approximated by the dataset itself; see below) and the details of the scheme. Such a probability model may involve simplifying approximations, e.g., assuming that the second and later password elements depend only on the preceding element (a first-order Markov model). The model thus depends on the relative frequencies of pairs of image elements within individual passwords, with the dataset approximating user choice across larger (and different) populations.

For Face, the analysis showed that users tend to select attractive faces of their own race (e.g., Asian, white, black), and selected predictable sets of faces such that an attacker knowing one face could leverage expectations of the face most likely to be selected next in a password. Gender information (male, female) also influenced choice. For the Face dataset, the weakest 25% of user passwords could be guessed in 13 attempts (compared to a full password space of $9^4 = 6561$), and the weakest 10% (corresponding to male participants) in 2 guesses. For Story, the weakest 25% could be guessed in 112 attempts (compared to a full space of $9 \cdot 8 \cdot 7 \cdot 6 = 3024$), and the weakest 10% in 35 guesses. This work highlights the potentially severe (and predictable, thus exploitable) skews in password distribution that may arise from unrestricted user choice in graphical password schemes, and the possibility of coarse personalized attacks, e.g., exploiting knowledge of user race or gender.

CUED-RECALL SYSTEMS. PassPoints and its relatives have attracted the most security analysis among graphical schemes. Efficient dictionary attacks have been enabled by two major weaknesses, both related to user choice. *Hotspots* [36,49] are popular points or areas of an image with higher probability of being chosen by users as click-points. *Patterns* [23] are lines or simple geometric shapes formed by user-chosen click-points in a password. The attacks below target PassPoints itself, as opposed to evolved systems like PCCP.

Success in exploiting hotspots with automated image processing tools has been reported [36, 100]. The most efficient hotspot attacks to date [121,127] harvest from different users a small sample of passwords for target images, using the component click-points to build "human-seeded" attack dictionaries. One such attack uses a first-order Markov model (see above); a second, based on an independent probability model, assumes click-points are independent of their predecessors.

We cite an example result, for a seed dataset harvested from a lab study, and tested on a field study dataset, with full password spaces of 2^{43}. Using dictionaries based on the first-order Markov model, 4% and 10% of field study passwords, respectively, were found within 100 guesses on two representative images. As a second example result, using the first-order Markov-model attack with cross-validation on the field study data found on average 7-10% of user passwords within 3 guesses. While roughly comparable in approach to the random sub-sampling analysis for Face above, the result here is more startling, as the full password space is 2^{43} for PassPoints, compared to just 2^{13} for Face.

The prevalence of simple click-order patterns in PassPoints passwords has also been exploited in customized dictionary attacks, some of which (counter-intuitively) are image independent, as some patterns are evident across a wide range of images. The best such attacks to date are purely-automated attacks [100, 125] not requiring human-seeding. The main patterns explored are variations of "loosely-defined" lines, and sequences of 5 points where consecutive points are constrained only by a fixed distance. To cite two example results, one approach using image-independent patterns found 48-54% of passwords on two representative images from a field study dataset, using dictionaries of about 2^{35} entries; a second approach, combining a simple pattern with an image processing based on a model of visual attention found 7-16% of passwords on the two images, using dictionaries of 2^{26} entries (vs. a full space of 2^{43}).

General comments. Dictionary attacks against recognition and cued-recall graphical password systems may require more effort up-front than against text passwords or (pure) recall-based graphical passwords, since attackers may have to first collect one or more images. Also, images gathered for one system will not help attacks on a second, unless both systems use a common image set.

Offline dictionary attacks of text passwords can be automated using password tools such as Crack [74], John the Ripper [32], and RainbowCrack [107]. Some of these may be modified for online attacks. We expect analogous cracking tools to surface for graphical passwords if the latter come into widespread use. Text password attack tools are often generic; attack tools for some graphical schemes may require system-specific images, but for others, pattern-only attacks [100] are independent of underlying images.

9.2 Capture Attacks

Password capture attacks directly obtain passwords (or parts thereof) by intercepting user-entered data, or tricking users into divulging passwords. For systems with time-invariant login responses, simple replay allows fraudulent access; this motivates consideration of systems with some form of challenge-response variation. However, in the face of attacks which capture login instances (e.g., due to malware, shoulder- surfing attacks, or other interception), even challenge-response knowledge-based authentication schemes appear to face a fundamental limitation, as argued by Coskun and Herley [28], offering at best protection against a limited number of observations. New proposals must consider the following known classes of capture attack; we assume that links over which graphical passwords are sent are encrypted, otherwise simple network sniffing or wire-tapping allows trivial capture.

9.2.1 Shoulder-surfing

Shoulder-surfing [8, 67, 99, 117] is a targeted attack exacerbated by the visual aspect of graphical passwords. As users enter login information, an attacker may gain knowledge about their credentials by direct observation or external recording devices such as video cameras. High-resolution cameras with telephoto lenses and surveillance equipment [67] make shoulder-surfing a real concern if attackers target specific users and have access to their geographic location.

Several existing graphical schemes believed to be resistant or immune to shoulder-surfing have significant usability drawbacks [66, 140], usually in the time and effort required to log in, making them less suitable for everyday authentication. Multi-touch tabletop interfaces support novel approaches offering shoulder-surfing resistant properties [60].

For some graphical passwords, multiple successful logins must be observed to deduce the full password (e.g., when only a subset of user portfolio images are displayed at each login, or if the shared secret is not explicitly revealed at login). Passwords in other schemes can be recovered from one successful login.

9.2.2 Reconstruction

Some attacks involve password reconstruction instead of direct capture [28]. For example one graphical password scheme [133] designed specifically to resist shoulder-surfing, was shown [48] to fall to a SAT (boolean satisfiability problem) solver, which reconstructs user secrets in a few seconds on observing a small number of logins. In general, these and *intersection attacks* [38] involve pooling leaked information gathered from observing or recording several logins for schemes in which the authentication response varies across login instances. Acoustic-based reconstruction attacks on text passwords, such as the password cracker of Berger et al. [12], seem less suited to graphical passwords, though ideas from the reconstruction techniques may apply to graphical schemes involving text input.

9.2.3 Malware

Malicious software includes any unauthorized software installed or downloaded without a user's informed consent, e.g., computer viruses and worms, Trojan horse software including login spoofing, code silently installed upon visiting web sites [89], and mobile code (e.g., JavaScript, Flash components). *Keystroke-loggers* [98] record keyboard input; *mouse-loggers* and *screen scrapers* capture mouse actions and record screen memory, to be sent remotely or made available for retrieval. Many graphical passwords require one or both a mouse-logger and screen scraper for capture, and often a keystroke-logger as well to collect usernames. Keystroke-loggers alone may suffice for schemes like Inkblot Authentication (Section 6), which use keyboard input only. If graphical passwords gain popularity, such malware will likely follow.

9.2.4 Phishing and pharming

Phishing attacks [34] trick users into entering their credentials at a fraudulent website, e.g., by having the user follow a link, in an email or engineered to return as a search engine result. As noted earlier, phishing attacks on recall-based graphical passwords resemble those on text passwords. For phishing attacks on recognition-based or cued-recall systems, specific images must be presented to the user. To do so, a phishing site may conduct earlier server probes to collect the images, or may retrieve and relay information from the legitimate site, in a *man-in-the-middle* (MITM) attack. *Pharming* [57], an advanced form of phishing, subverts the DNS system (by forged DNS responses or DNS cache poisoning) such that domain names are fraudulently resolved to the IP address of an attacker's site. Depending on the password scheme, recording one or more login attempts at a phishing site may provide sufficient information for an attacker to subsequently log in. With a MITM attack, attackers may also log in to the legitimate site at least once by hijacking a single correct authentication response during the attack.

9.2.5 Social engineering

Phishing is a form of social engineering attack [141]; users may be tricked to reveal credentials by any means, e.g., phone calls from a fake help desk or credit company. While such methods may require targeted background work (or knowledge of personal details in *personalized attacks*), this is often easier than otherwise breaking into a system [70].

Text passwords and alphanumeric information are relatively easy to share with colleagues or attackers. For graphical passwords, a frame of reference must first be coordinated to convey the password in sufficient detail for use. This security advantage (complicating social engineering attacks) has usability drawbacks, e.g., complicating password reset by phone, and safe backup storage of passwords. Despite the added difficulty, Dunphy et al. [39] give preliminary evidence

that users can describe PassPoints passwords sufficiently to enable use by others. Other means of sharing a graphical password include taking photos, screen shots, and drawing.

9.3 Security Summary

The first four rows of Tables 1–3 in Section 7 give a summary overview of security. The first two lines address guessing attacks and the next two, capture attacks. More specifically: "Theoretical pswd (bits)" reflects resistance to exhaustive search attacks; "User choice resilience" reflects resistance to dictionary attacks and optimizations, including guessing attacks on specific schemes; "Variant response" relates to replay and reconstruction attacks (including those facilitated by shoulder surfing), as explained in Section 7 and the first paragraph of Section 9.2; and "Server probes" relates to steps necessary for pharming, phishing and other social engineering attacks. As a summary statement on client-side malware, in general it is safest to assume that password systems (text, graphical, or other) are not sufficiently strong to provide security in the presence of client-side malware.

We may also directly compare the security of text passwords to graphical passwords with respect to the attack approaches under Section 9.2. Graphical passwords are in general more vulnerable to shoulder surfing than text passwords; however text passwords are in general at least as, or more vulnerable with respect to reconstruction, pharming/phishing and other social engineering attacks, and malware. For the latter, graphical passwords require slightly more advanced malware.

10. METHODOLOGY FOR EVALUATION

Establishing whether a graphical password system meets its usability and security goals can be challenging. This section summarizes evaluation approaches used, including user studies, with focus on aspects of special concern for examining graphical password systems. Data collected from such user studies is also critical in the security evaluation discussed above. Several general approaches exist for user testing graphical password systems. Each can provide valuable empirical data.

With usability inspection methods (such as *cognitive walkthroughs* [135] and *heuristic evaluations* [79]), evaluators inspect and evaluate usability-related aspects of a system. These are conducted without end users and require a certain level of expertise in usability [79]. They are useful early steps in finding obvious usability problems, but are no substitute for user studies. While user testing is necessary to evaluate usability, it is also critically important in evaluating the practical security of graphical passwords, as well as the interplay between these two dimensions. The challenge lies in designing the tests so that meaningful and representative data is collected. Security tasks are usually not the user's primary task in practice, yet they frequently become a focus when user tests are conducted, which may lead to different behaviour then would happen if the system was deployed in practice. Novelty effects can occur; this can be especially problematic with graphical password selection, since users have yet to develop the coping skills that they may adopt with regular use.

Text passwords, as the most common form of knowledge-based authentication, are often used as a benchmark to assess the usability and security of graphical password schemes. While useful, this comparison is biased by years of user experience with text passwords. They are familiar and comfortable with the login process, can complete it quickly, and have developed a wide range of coping behaviours and strategies to deal with memorability issues. The coping strategies can improve user performance for usability but may also lead to weaker password selection. Complicating matters further, the usable security community lacks definitive and comprehensive results on text passwords so it is difficult to use them as benchmarks.

This raises the issue of user training and familiarization before collecting data for analysis. The type of training, its length, and the instructions given to users can influence their behaviour. Inappropriate training may make users too comfortable and display behaviour not indicative of what would occur in a practical setting, they may become tired of the task and become careless, or they may behave more or less securely based entirely on the instructions (which may not reflect a real life scenario). It is unclear how much training users should receive, if any, before evaluation, but researchers should carefully consider potential effects when interpreting the results of user studies.

The problem of multiple passwords also needs special consideration. Recent publications [24, 41, 71] have tackled this issue but ecological validity remains difficult to achieve. Details such as how passwords are introduced, the number of passwords, similarity between passwords, and login frequency may significantly impact results. Furthermore, interference between different types of graphical passwords has yet to be examined. How to best evaluate multiple password interference remains an open issue.

10.1 Lab studies

Lab studies provide a means to evaluate the success of design decisions in isolation, quantify improvements and performance, discover unexpected usability problems, and identify designs with higher probability of success (or failure) before investing large amounts of time and resources in field studies. While field studies offer superior ecological validity, lab studies have the advantage of being held in a controlled setting and so can be used to establish performance bounds that can indicate whether field tests are worthwhile. The experimenter can ensure that participants are focused on the task at hand, that the study is designed to enable statistical testing of different measures, and that clear comparisons can be made to assess the effectiveness of certain design decisions. For example, a study may have a goal of examining the effectiveness of a new password selection aid. In this case, two versions of the system would be built, differing only in the inclusion or absence of the new selection aid. The system would be instrumented to record the user's passwords and input during password entry, and to include measures such as time to create a new password and number of errors made. With security systems, it is especially important to be relatively confident of a system's design in the lab before deploying it in field studies because of the potential for security and privacy breaches of users' real resources and information if problems occur in a field study.

Besides the predetermined measures, lab studies aim to uncover any unforeseen difficulties encountered by the users across a set of predetermined tasks. These tasks should be carefully chosen to reflect realistic usage scenarios. To maximize ecological validity, the environment should be set up to mimic target environments as closely as possible in tech-

nical details and instructions given. Users should be closely observed as they perform these tasks, as this is how many usability problems are revealed. Researchers must also try to avoid biasing user behaviour, especially when dealing with security, as users may behave more or less securely than usual to "help the researcher". A method called *think-aloud* is often used, where users are encouraged to voice a running commentary as they perform the tasks. Pre/post questionnaires or interviews are useful to gather users' opinions, attitudes, and feedback. These should be a secondary source of information, used in conjunction with observations and system logs, as users' reported views often do not reflect their performance and fail to reveal crucial usability problems.

An often cited guideline, advocating smaller, quicker usability studies—that five users are enough to discover most usability problems [78, 130]—has long been used to justify small usability studies. Recent work revisiting this suggestion highlights that this is often not enough and that in some cases, severe usability problems are only discovered after running a larger group of participants [42, 85, 108]. The likelihood of finding usability problems is not evenly distributed and may vary with the complexity of the system being tested. Some problems only arise under specific circumstances, so a small sample of users may not be sufficient to uncover them. The variability in the problems found by any one user makes it unlikely that five users would discover most usability problems. Faulkner [42] justifies that twenty users "can allow the practitioner to approach increasing levels of certainty that high percentages of existing usability problems have been found in the testing". More participants are also needed for meaningful statistical analysis. When conducting user studies on authentication schemes involving user choice, there is an additional motivation for larger studies: user behaviour patterns which weaken security may only become apparent with a larger sample.

Memorability must be assessed in authentication systems. One approach is to administer distraction tasks within a session, as done in psychological studies on memory. These are intended to clear a user's working memory (short term memory) and simulate the longer passage of time. To be more ecologically valid, many studies have multiple lab sessions, where participants return to log in over the course of several days, weeks, or months. Studies, however, that only require users to remember one password (which often does not protect a meaningful account), raise other ecological validity concerns. Testing multiple passwords raises its own ecological validity issues as noted earlier.

10.2 Field studies

In a field study, the system to be tested is deployed for a group of users who incorporate the system into their regular routine over a period of time (typically a few weeks to a few months), so the advantage is strong ecological validity. Field studies offer the best measure of some important characteristics, such as memorability, in a realistic setting. However, they require a significant investment in resources and time and are preferably undertaken only after success has been reached in a lab environment. A field study allows researchers and designers to observe how the system would operate in real-life and more accurately judge its acceptability, suitability, and usability. With usable authentication research involving passwords, field studies may provide data on what types of passwords users really select when they need to use them regularly, whether passwords are memorable, what unexpected coping strategies arise, and whether the scheme is usable on computer systems with different configurations (e.g., screen sizes). Other issues, such as multiple password inteference or password usage in environments where shoulder-surfing is possible, can also be studied. Real-world usage is of particular concern with security systems because security is often a secondary task [136], enabling (or hindering) access to the user's primary goal. In such cases, user behaviour may vary considerably compared to when users are asked to complete the security tasks in the lab, where it may be their primary focus.

Besides the risk of exposing user resources or information if security vulnerabilities are present and exploited, the data collected from field studies may be affected by factors that are not immediately apparent. It is difficult to know, for example, whether users are employing coping mechanisms such as printing screen captures of passwords. Issues could be explored during interviews or through post-task questionnaires, but researchers must have a suspicion that particular behaviours are occurring to investigate them. Users may not realize that some behaviours are insecure or worthy of mention unless specifically prompted. Another factor to consider is the perceived value of the accounts being protected in the study. While it is inadvisable to risk compromising high-value information in early field studies, the impact of such design choices must be considered when generalizing results.

10.3 Other types of studies

10.3.1 Web-based

Web-based user studies are gaining popularity [6, 41, 44, 71]. While there is no agreed-upon definition for web-based studies, herein we use this term for the case where there is no face-to-face contact with the user study participants at all. Advantages of web studies include: large numbers of participants can be recruited, the participant pool is likely more diverse than in most controlled studies, participants can be prompted to complete tasks at several different times, and participant behaviour may be more natural than in a lab setting. Web-based studies are often cheaper, easier, and faster than traditional controlled studies. Challenges include: great care is needed in getting informed consent from participants (e.g., through a signature or other means of authentication as required by organizational ethics review boards), it is nearly impossible to know if demographics information collected is accurate, it is difficult to enforce adherence to procedures, and the collected data may not reflect real behaviour.

Web-based studies offer one measure of ecological validity, by being held in the participants' natural environment, as opposed to in a controlled lab environment. Additional ecological validity can be gained by integrating realistic tasks and systems, rather than using fabricated tasks. For authentication, studies that focus users on primary tasks other than the actual authentication offer a higher degree of ecological validity than those that simply ask users to log in.

Amazon's Mechanical Turk [3] has been used recently to conduct some usable security studies [59, 105]. This web-based "mTurk" system allows requesters to post "human intelligence tasks" (HITs) which can be accepted and completed by workers for payment. Advantages of using systems like mTurk include: data can be gathered very quickly and

inexpensively from a very large number of users. However, regarding challenges, those mentioned above for other web-based studies remain and may in fact be magnified, such as dealing with skewed demographics (compared to actual target users), and ecological validity issues (is a rapid task completion a primary motivation of mTurk workers?). Additional challenges include designing appropriate short, specific, "micro" tasks that are likely to be completed [64]. For these reasons, the overall suitability of mTurk for authentication studies remains an open question.

10.3.2 Hybrid

In hybrid studies, researchers combine lab studies with tasks completed in participants' regular environments, gaining advantages of both an initial controlled environment and increased ecological validity in subsequent tasks. The tasks are usually invented, but may be designed to approximate realistic tasks. Instructions for follow-up activities may be provided at the end of the initial lab session, or may be sent through email at a later time. For example, in authentication studies, participants may be prompted through email to log in to web-based test systems at various intervals. These passwords may not protect valuable or personal information, but some ecological validity is gained by having users login from within their regular environments. Furthermore, primary tasks can be assigned, such as asking users to comment on a blog or to access subscription-based material, where login with the authentication scheme is simply part of the process.

10.4 Evaluation checklist

To summarize, we list a set of usability and security items to include in presentations of new knowledge-based authentication schemes(whether graphical passwords or otherwise), for evaluating complete proposals.

1. Are target users, domains, and applications clearly identified?
2. Are evaluation parameters, and the theoretical password space, clearly stated?
3. Does the analysis explain the effect of user choice on password distributions, with informed discussion of the effective password space?
4. Does the analysis consider the full range of attacks plausible in the targeted domain and application, and how each attack fails or succeeds? Does a login session leak more information than expected?
5. At minimum, has a lab user study been done (or other types of studies with higher ecological validity), with results compared to appropriate alternatives?
6. Is password interference discussed (e.g., as informed by a user study)?
7. Do the user study and security analysis use the same parameters?

11. FURTHER DISCUSSION AND CONCLUSION

Our tour of graphical password research reveals a rich palette of ideas, but few schemes that deliver on the original promise of addressing the known problems with text passwords. Indeed, review of the first era of graphical password schemes indicates that many of the same problems continue to re-surface. For graphical passwords to advance as a serious authentication alternative, we believe research must be conducted and presented in a manner allowing systematic examination and comparison of each scheme's main characteristics, showing how each meets the usability and security requirements of specific target environments.

Published research in the area of graphical passwords currently lacks consistency, making it difficult to compare or reproduce results. Where reasonable, researchers should choose methods and measures that allow for comparison with other work. Moreover, research proposals and analyses for new systems should include: specific motivation for the work, a description of the system's design including any special instrumentation for prototyping and testing versions, a clear description of the study methodology, and analysis explaining which usability and security aspects are being tested, aside from main results. While early work is often by definition incomplete, a comprehensive evaluation should acknowledge the above points and identify foreseeable issues, even if a full evaluation has not yet been conducted.

Many graphical password systems in the literature to date lack rigorous evaluation in security, usability or both. As Section 9.1.3 notes, significant security flaws have been found in the original versions of all three canonical schemes. A closer look at individual systems has typically revealed less security than promised, matching historical early experience in other areas—usually repaired with maturity. Most systems to date suffer from either small theoretical password spaces (if the system is configured to be usable) or patterns in user choice that reduce the size of the effective password space, highlighting an important insight from the field study and security analysis on Face and Story [30]: many graphical password schemes may require "a different posture towards password selection" than text passwords, where selection by the user is the norm. New designs should thus focus on increasing password entropy without sacrificing usability and memorability.

The main purpose of authentication schemes is to allow system access only by legitimate users. To thoroughly evaluate the security of a graphical password proposal, and to facilitate comparison with alternatives, all standard threats and known attacks should be analyzed, with convincing arguments on how the scheme precludes (or falls to) them. Moreover, such security analysis must be accompanied by concrete experimental studies and usability analysis. For example, a system is of limited interest if it prevents shoulder-surfing but has a password space so small that it falls to a plausible simple brute-force attack. (While shoulder-surfing is a threat in public environments, it is only one of many threats; far less literature has considered more scalable threats such as keystroke loggers or graphical dictionary attacks.) If a system is intended for use in particular environments where some standard threats are not a concern, then the relevant details should be clearly specified. Essential security measures to be reported include: the size of the theoretical password space; the estimated size of the effective password space; details about known or anticipated exploitable patterns in user choice; and an analysis of how the scheme withstands known online and offline attacks. While it may be impossible to prove that graphical passwords can pro-

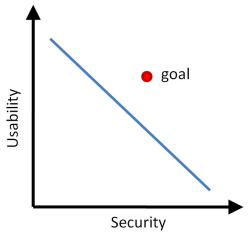

Figure 10: Most graphical password schemes fall along the descending line, where increased security implies decreased usability. The design goal is to increase usability and security simultaneously.

vide greater security than text passwords in the face of the most serious attacks such as resident malware, graphical passwords nonetheless remain of practical interest due to the possibility of offering at least as much security, with the possibility of greater usability and memorability.

In many systems having poor security, user actions compromise security in favour of memorability. The exploitable patterns evident in Passfaces, DAS, and PassPoints passwords result from users trying to select memorable passwords, which in turn increases predictability and facilitates password guessing. A challenge for designers is to identify memory aids for legitimate users, that cannot be leveraged by attackers to guess passwords. Furthermore, systems allowing some degree of user choice should encourage randomization of user-chosen sequences as well as individual items, to avoid divide and conquer guessing attacks. It remains an open question whether systems can be designed such that user choice does not significantly weaken security, or whether a successful combination of system suggestion and user choice can be devised. A complementary method for addressing predictable passwords is the use of so-called "strong" password protocols (e.g., SRP [142], EKE [9]) designed to provide protection against offline guessing attacks by avoiding verifiable text [50]. This can be important for both text and graphical passwords, but their design is notoriously tricky.

For usability, a major concern is multiple password interference. Visual cues provided by graphical passwords along with the potential of human memory processing for images offer reason for optimism, but further research is required to confirm that these can be translated into schemes with increased security and usability, in a realistic setting. As graphical passwords are not widely deployed, it is unknown if we will simply mirror text password problems, where users develop coping strategies, devise and reuse common patterns, and choose minimally secure passwords.

The development of password managers for graphical passwords might address the problem of memory interference. However, such managers may well suffer the same usability and security challenges as their text counterparts noted in Section 1, with additional challenges such as dealing with challenge-response schemes (variant responses) and coping with the variety of password entry requirements. Further consideration of password managers is beyond the scope of this paper. Related to password interference, it would be interesting to investigate user choice if given the opportunity to select both the password schemes and the passwords for multiple accounts, allowing for any number of each type.

We expect tomorrow's ideal graphical password systems may have many of the following desirable characteristics, reflecting lessons learned from proposals to date.

1. Theoretical password space meeting the security policy of the intended domain.

2. Avoidance of exploitable reductions in security due to user choice of passwords, e.g., through persuading password choice towards flatter distributions.

3. At least mild resistance to different types of capture attacks including shoulder surfing and key logging, through variable response (challenge-response) design.

4. Cues aiding memorability, design features minimizing password interference.

5. Usability (e.g., login success rates, login times, password creation times) as close as possible to, or better than, text passwords.

6. Implicit feedback to legitimate users, when passwords are multi-part.

7. Leveraging of pre-existing user-specific knowledge where possible, rather than having users memorize entirely new and/or random information.

In addition to the importance of the evaluation checklist of Section 10.4, and the characteristics immediately above, we emphasize here two additional lessons learned. First, design decisions related to usability should be evaluated jointly with an exploration of their impact on security, since a usable authentication system without adequate security fails to meet its primary purpose. For example, a system where users can choose memorable-but-weak passwords may be usable but can provide a false sense of security. Interface design changes that appear to affect only usability may in fact introduce additional security vulnerabilities.

Second, in assessing usability, apples-to-apples comparison requires comparing schemes of comparable security. Usability comparisons between schemes offering significantly different security propositions must highlight the lack of calibration, to avoid seriously misleading others. For example, the full password space of many recognition-based systems calibrates to that of 4-digit PINs, while recall and cued-recall systems are similar to text passwords of 8-characters-or-more. Longer login times may be acceptable for password-level systems than for PIN-level systems (recalling the Section 7 levels), if the former provide greater security.

Security and usability have historically been viewed as items to be traded off, representing opposite ends of a spectrum: increasing one necessarily decreases the other. Most products and mechanisms to date, including for many graphical password schemes, afford only fixed levers such that, for example, adding extra rounds to Passfaces increases security at the cost of an additional memorability burden since each extra round also exposes users to a new set of decoys. As illustrated in Figure 10, the challenge for the second generation of graphical passwords, and in the *design for usable security* in general, is to find new designs and architectures which afford increases in security and usability together.

Acknowledgments

We thank Kemal Bicakci, and anonymous referees for comments that have helped considerably improve this paper. The first and third authors acknowledge Discovery Grants through the Natural Sciences and Engineering Research Council of Canada (NSERC). The third author is Canada Research Chair in Authentication and Computer Security, and acknowledges NSERC funding thereof. Partial funding from the NSERC Internetworked Systems Security Network (ISSNet) is also acknowledged.

12. REFERENCES

[1] A. Adams, M. A. Sasse, and P. Lunt. Making passwords secure and usable. In *HCI 97: Proceedings of HCI on People and Computers*, pages 1–19, London, UK, 1997. Springer-Verlag.

[2] F. Alsulaiman and A. El Saddik. A novel 3D graphical password schema. In *IEEE Int. Conf. on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, July 2006.

[3] Amazon. Amazon mechanical turk. `http://www.mturk.com/`, 2010.

[4] J. Anderson and G. Bower. Recognition and retrieval processes in free recall. *Psychological Review*, 79(2):97–123, March 1972.

[5] M. Anderson and J. Neely. *Memory. Handbook of Perception and Cognition*, chapter 8, pages 237–313. Academic Press, 2nd edition, 1996.

[6] D. Andrews, B. Nonnecke, and J. Preece. Electronic survey methodology: A case study in reaching hard-to-involve Internet users. *International Journal of Human-Computer Interaction, Lawrence Erlbaum Associates*, 16(2):185–210, 2003.

[7] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. In *USENIX 4th Workshop on Offensive Technologies*, 2010.

[8] M. Backes, M. Durmuth, and D. Unruh. Compromising reflections — or — how to read LCD monitors around the corner. In *IEEE Symposium on Security and Privacy*, 2008.

[9] S. M. Bellovin and M. Merritt. Encrypted key exchange: Password based protocols secure against dictionary attacks. In *IEEE Symposium on Research in Security and Privacy*, 1992.

[10] J. Bentley and C. Mallows. How much assurance does a PIN provide? In Baird and Lopresti, editors, *Human Interactive Proofs (HIP), LNCS 3517, Springer-Verlag*, pages 111–126, 2005.

[11] F. Bergadano, B. Crispo, and G. Ruffo. High dictionary compression for proactive password checking. *ACM Transactions on Information and System Security*, 1(1):3–25, 1998.

[12] Y. Berger, A. Wool, and A. Yeredor. Dictionary attacks using key acoustic emanations. In *13th ACM Conference on Computer and Communications Security (CCS)*, November 2006.

[13] K. Bicakci. Optimal discretization for high-entropy graphical passwords. In *23rd International Symp. on Computer and Information Sciences, IEEE ISCIS 2008*, Istanbul, Turkey, October 2008.

[14] K. Bicakci, N. B. Atalay, M. Yuceel, H. Gurbaslar, and B. Erdeniz. Towards usable solutions to graphical password hotspot problem. In *33rd Annual IEEE International Computer Software and Applications Conference*, 2009.

[15] K. Bicakci, M. Yuceel, B. Erdeniz, H. Gurbaslar, and N. B. Atalay. Graphical passwords as browser extension: Implementation and usability study. In *Third IFIP WG 11.11 International Conference on Trust Management*, Purdue University, USA, June 2009.

[16] J. Birget, D. Hong, and N. Memon. Graphical passwords based on robust discretization. *IEEE Transactions on Information Forensics and Security*, 1(3):395–399, 2006.

[17] G. Blonder. Graphical passwords. U.S. Patent 5,559,961, 1996.

[18] M. Bond. Comments on grIDsure authentication. `http://www.cl.cam.ac.uk/~mkb23/research/GridsureComments.pdf`, March 2008.

[19] S. Brostoff, P. Inglesant, and M. A. Sasse. Evaluating the usability and security of a graphical one-time PIN system. In *BCS Conf. on Human Computer Interaction (British HCI)*, September 2010.

[20] S. Brostoff and M. Sasse. Are Passfaces more usable than passwords? A field trial investigation. In *British Human-Computer Interaction Conference (HCI)*, September 2000.

[21] S. Chiasson, R. Biddle, and P. C. van Oorschot. A second look at the usability of click-based graphical passwords. In *ACM Symposium on Usable Privacy and Security (SOUPS)*, July 2007.

[22] S. Chiasson, A. Forget, R. Biddle, and P. C. van Oorschot. Influencing users towards better passwords: Persuasive Cued Click-Points. In *Human Computer Interaction (HCI), The British Computer Society*, September 2008.

[23] S. Chiasson, A. Forget, R. Biddle, and P. C. van Oorschot. User interface design affects security: Patterns in click-based graphical passwords. *International Journal of Information Security, Springer*, 8(6):387–398, 2009.

[24] S. Chiasson, A. Forget, E. Stobert, P. C. van Oorschot, and R. Biddle. Multiple password interference in text and click-based graphical passwords. In *ACM Computer and Communications Security (CCS)*, November 2009.

[25] S. Chiasson, J. Srinivasan, R. Biddle, and P. C. van Oorschot. Centered discretization with application to graphical passwords. In *USENIX Usability, Psychology, and Security (UPSEC)*, April 2008.

[26] S. Chiasson, P. C. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *15th USENIX Security Symposium*, August 2006.

[27] S. Chiasson, P. C. van Oorschot, and R. Biddle. Graphical password authentication using Cued Click Points. In *European Symposium On Research In Computer Security (ESORICS), LNCS 4734*, pages 359–374, September 2007.

[28] B. Coskun and C. Herley. Can "something you know" be saved? In *Information Security Conference (ISC), LNCS 5222*, pages 421–440. Springer-Verlag, 2008.

[29] F. Craik and J. McDowd. Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3):474–479, July 1987.

[30] D. Davis, F. Monrose, and M. Reiter. On user choice in graphical password schemes. In *13th USENIX Security Symposium*, 2004.

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

[32] S. Designer. John the Ripper password cracker. `http://www.openwall.com/john/`.

[33] R. Dhamija and A. Perrig. Déjà Vu: A user study using images for authentication. In *9th USENIX Security Symposium*, 2000.

[34] R. Dhamija, J. Tygar, and M. Hearst. Why phishing works. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2006.

[35] P. Diggle. *Statistical Analysis of Spatial Point Patterns*. Academic Press: New York, NY, 1983.

[36] A. Dirik, N. Menon, and J. Birget. Modeling user choice in the Passpoints graphical password scheme. In *3rd ACM Symposium on Usable Privacy and Security (SOUPS)*, July 2007.

[37] P. Dunphy, A. Fitch, and P. Olivier. Gaze-contingent passwords at the ATM. In *4th Conference on Communication by Gaze Interaction (COGAIN)*, September 2008.

[38] P. Dunphy, A. P. Heiner, and N. Asokan. A closer look at recognition-based graphical passwords on mobile devices. In *ACM Symposium on Usable Privacy and Security (SOUPS)*, July 2010.

[39] P. Dunphy, J. Nicholson, and P. Olivier. Securing Passfaces for description. In *4th ACM Symposium on Usable Privacy and Security (SOUPS)*, July 2008.

[40] P. Dunphy and J. Yan. Do background images improve "Draw a Secret" graphical passwords? In *14th ACM Conference on Computer and Communications Security (CCS)*, October 2007.

[41] K. Everitt, T. Bragin, J. Fogarty, and T. Kohno. A comprehensive study of frequency, interference, and training of multiple graphical passwords. In *ACM Conference on Human Factors in Computing Systems (CHI)*, April 2009.

[42] L. Faulkner. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, 35(3):379–383, 2003.

[43] D. Feldmeier and P. Karn. UNIX Password Security–Ten Years Later. In *Crypto'89*, August 1989.

[44] D. Florencio and C. Herley. A large-scale study of WWW password habits. In *16th ACM International World Wide Web Conference (WWW)*, May 2007.

[45] D. Florencio and C. Herley. Where do security policies come from? In *Symposium on Usable Privacy and Security*, 2010.

[46] H. Gao, X. Guo, X. Chen, L. Wang, and X. Liu. Yagp: Yet another graphical password strategy. In *Annual Computer Security Applications Conference*, 2008.

[47] J. Goldberg, J. Hagman, and V. Sazawal. Doodling our way to better authentication (student poster). In *ACM Conference on Human Factors in Computing Systems (CHI)*, April 2002.

[48] P. Golle and D. Wagner. Cryptanalysis of a cognitive authentication scheme (extended abstract). In *IEEE Symposium on Security and Privacy*, May 2007.

[49] K. Golofit. Click passwords under investigation. In *12th European Symposium On Research In Computer Security (ESORICS), LNCS 4734*, September 2007.

[50] L. Gong, M. Lomas, R. Needham, and J. Saltzer. Protecting poorly chosen secrets from guessing attacks. *IEEE J. on Selected Areas in Communications*, 11(5):648–656, June 1993.

[51] N. Govindarajulu and S. Madhvanath. Password management using doodles. In *9th International Conference on Multimodal Interfaces (ICMI)*, November 2007.

[52] GrIDsure. GrIDsure corporate website. `http://www.gridsure.com`, Last accessed August 2009.

[53] M. D. Hafiz, A. H. Abdullah, N. Ithnin, and H. K. Mammi. Towards identifying usability and security features of graphical password in knowledge based authentication technique. In *Second Asia International Conf. on Modelling & Simulation*, pages 396–403. IEEE, 2008.

[54] E. Hayashi, N. Christin, R. Dhamija, and A. Perrig. Use Your Illusion: Secure authentication usable anywhere. In *4th ACM Symposium on Usable Privacy and Security (SOUPS)*, Pittsburgh, July 2008.

[55] C. Herley, P. van Oorschot, and A. Patrick. Passwords: If We're So Smart, Why Are We Still Using Them? In *Financial Cryptography and Data Security, LNCS 5628, Springer*, 2009.

[56] A. Hollingworth and J. Henderson. Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1):113–136, 2002.

[57] ICANN Security and Stability Advisory Committee. Domain name hijacking: Incidents, threats, risks, and remedial actions. `http://www.icann.org/en/announcements/hijacking-report-12jul05.pdf`, July 2005.

[58] I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin. The design and analysis of graphical passwords. In *8th USENIX Security Symposium*, August 1999.

[59] P. Kelley, L. Cesca, J. Bresee, and L. Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *CHI'10: Proc. 28th International Conference on Human Factors in Computing Systems*, pages 1573 – 1582, 2010.

[60] D. Kim, P. Dunphy, P. Briggs, J. Hook, J. Nicholson, J. Nicholson, and P. Olivier. Multi-touch authentication on tabletops. In *28th ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1093–1102, Atlanta, USA, April 2010.

[61] W. Kintsch. Models for free recall and recognition. In D. Norman, editor, *Models of Human Memory*. Academic Press: New York, 1970.

[62] B. Kirkpatrick. An experimental study of memory. *Psychological Review*, 1:602–609, 1894.

[63] D. Kirovski, N. Jojie, and P. Roberts. Click passwords. In *Security and Privacy in Dynamic Environments. Proceedings of the IFIP TC-11 21st International Information Security Conference (SEC 2006)*, volume 201, pages 351–363. Boston: Springer, 2006.

[64] A. Kittur, E. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *CHI'08: Proc. 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, 2008.

[65] D. Klein. Foiling the cracker: A survey of, and improvements to, password security. In *2nd USENIX Security Workshop*, 1990.

[66] S. Komanduri and D. Hutchings. Order and entropy in Picture Passwords. In *Graphics Interface Conference (GI)*, May 2008.

[67] B. Laxton, K. Wang, and S. Savage. Reconsidering physical key secrecy: Teleduplication via optical decoding. In *ACM Conference on Computer and Communications Security*, 2008.

[68] S. Madigan. Picture memory. In J. Yuille, editor, *Imagery, Memory, and Cognition: Essays in Honor of Allan Paivio*, chapter 3, pages 65–89. Lawrence Erlbaum Associates, 1983.

[69] M. Mannan, T. Whalen, R. Biddle, and P. van Oorschot. The usable security of passwords based on digital objects: From design and analysis to user study. Technical Report TR-10-02, School of Computer Science, Carleton University, 2010.

[70] K. Mitnick and W. Simon. *The Art of Deception: Controlling the Human Element of Security*. New York: John Wiley & Sons, 2002.

[71] W. Moncur and G. Leplatre. Pictures at the ATM: Exploring the usability of multiple graphical passwords. In *ACM Conference on Human Factors in Computing Systems (CHI)*, April 2007.

[72] F. Monrose and M. Reiter. Graphical passwords. In L. Cranor and S. Garfinkel, editors, *Security and Usability: Designing Secure Systems That People Can Use*, chapter 9, pages 157–174. O'Reilly Media, 2005.

[73] R. Morris and K. Thompson. Password Security: A Case History. *Communications of the ACM*, 22(11):594–597, 1979.

[74] A. Muffett. Crack password cracker. `http://ciac.llnl.gov/ciac/ToolsUnixAuth.html`, 2004.

[75] D. Nali and J. Thorpe. Analyzing user choice in graphical passwords. Technical Report TR-04-01, School of Computer Science, Carleton University, May 2004.

[76] A. Narayanan and V. Shmatikov. Fast dictionary attacks on passwords using time-space tradeoff. In *ACM Conference on Computer and Communications Security (CCS)*, November 2005.

[77] D. Nelson, V. Reed, and J. Walling. Pictorial Superiority Effect. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5):523–528, 1976.

[78] J. Nielsen. *Usability Engineering*. Boston: AP Professional, 1993.

[79] J. Nielsen and R. Mack. *Usability Inspection Methods*. John Wiley & Sons, Inc, 1994.

[80] P. Oechslin. Making a faster cryptanalytic time-memory trade-off. In *Crypto'03*, August 2003.

[81] M. Orozco, B. Malek, M. Eid, and A. El Saddik. Haptic-based sensible graphical password. In *Proceedings of Virtual Concept*, 2006.

[82] A. Paivio. *Mind and Its Evolution: A Dual Coding Theoretical Approach*. Lawrence Erlbaum: Mahwah, N.J., 2006.

[83] A. Paivio, T. Rogers, and P. C. Smythe. Why are pictures easier to recall than words? *Psychonomic Science*, 11(4):137–138, 1968.

[84] Passfaces Corporation. The science behind Passfaces. White paper, `http://www.passfaces.com/enterprise/resources/white_papers.htm`, accessed July 2009.

[85] C. Perfetti and L. Landesman. Eight is not enough. *User Interface Engineering*, 2001.

[86] T. Pering, M. Sundar, J. Light, and R. Want. Photographic authentication through untrusted terminals. *Pervasive Computing*, pages 30–36, January - March 2003.

[87] B. Pinkas and T. Sander. Securing passwords against dictionary attacks. In *9th ACM Conference on Computer and Communications Security (CCS)*, November 2002.

[88] L. Y. Por, X. T. Lim, M. T. Su, and F. Kianoush. The design and implementation of background Pass-Go scheme towards security threats. *WSEAS Transactions on Information Science and Applications*, 5(6):943–952, June 2008.

[89] N. Provos, P. Mavrommatis, M. Abu Rajab, and F. Monrose. All your iFrames point to us. In *17th USENIX Security Symposium*, 2008.

[90] J. G. W. Raaijmakers and R. M. Shiffrin. Models for recall and recognition. *Annual Reviews Psych.*, 43:205–234, January 1992.

[91] D. Ramsbrock, R. Berthier, and M. Cukier. Profiling attacker behavior following SSH compromises. In *37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2007.

[92] K. Renaud. Evaluating authentication mechanisms. In L. Cranor and S. Garfinkel, editors, *Security and Usability: Designing Secure Systems That People Can Use*, chapter 6, pages 103–128. O'Reilly Media, 2005.

[93] K. Renaud. A visuo-biometric authentication mechanism for older users. In *British HCI*, pages 167–182, September 2005.

[94] K. Renaud. Guidelines for designing graphical authentication mechanism interfaces. *International Journal of Information and Computer Security*, 3(1):60–85, June 2009.

[95] K. Renaud. On user involvement in production of images used in visual authentication. *Journal of Visual Languages and Computing*, 20(1):1–15, February 2009.

[96] K. Renaud and A. D. Angeli. My password is here! An investigation into visio-spatial authentication mechanisms. *Interacting with Computers*, 16(4):1017–1041, 2004.

[97] K. Renaud and E. Smith. Jiminy: Helping user to remember their passwords. Technical report, School of Computing, Univ. of South Africa, 2001.

[98] B. Ross, C. Jackson, N. Miyake, D. Boneh, and J. Mitchell. Stronger password authentication using browser extensions. In *14th USENIX Security Symposium*, Baltimore, August 2005.

[99] V. Roth, K. Richter, and R. Freidinger. A PIN-entry method resiliant against shoulder surfing. In *11th ACM Conference on Computer and Communications Security*, 2004.

[100] A. Salehi-Abari, J. Thorpe, and P. van Oorschot. On purely automated attacks and click-based graphical passwords. In *Annual Computer Security Applications Conf. (ACSAC)*, 2008.

[101] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the 'weakest link' – a human/computer interaction approach to usable and effective security. *BT Tech. Journal*, 19(3):122–131, July 2001.

[102] S. Schechter and A. B. Brush. It's No Secret: Measuring the Security and Reliability of Authentication via 'Secret' Questions. In *IEEE Symposium on Security and Privacy*, May 2009.

[103] C. Seifert. Analyzing malicious SSH login attempts. `http://www.securityfocus.com/infocus/1876`, September 2006.

[104] SFR Software. visKey for Pocket PC. `http://www.sfr-software.de/cms/EN/pocketpc/viskey/`.

[105] S. Sheng, M. Holbrook, P. Kumaraguru, L. Cranor, and J. Downs. Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *CHI '10: Proceedings of the 28th International Conference on Human Factors in Computing Systems*, pages 373 – 382, 2010.

[106] R. Shepard. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6:156–163, 1967.

[107] Z. Shuanglei. Project RainbowCrack. `http://www.antsight.com/zsl/rainbowcrack`, 2005.

[108] J. Spool and W. Schroeder. Testing web sites: Five users is nowhere near enough. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2001.

[109] L. Standing, J. Conezio, and R. Haber. Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2), 1970.

[110] E. Stobert, A. Forget, S. Chiasson, P. van Oorschot, and R. Biddle. Exploring usability effects of increasing security in click-based graphical passwords. In *Annual Computer Security Applications Conference (ACSAC)*, 2010.

[111] A. Stubblefield and D. Simon. Inkblot Authentication, MSR-TR-2004-85. Technical report, Microsoft Research, 2004.

[112] X. Suo. A design and analysis of graphical password. Master's thesis, College of Arts and Science, Georgia State University, August 2006.

[113] X. Suo, Y. Zhu, and G. Owen. Graphical passwords: A survey. In *Annual Computer Security Applications Conf. (ACSAC)*, Dec. 2005.

[114] Tafasa. Patternlock. `http://www.tafasa.com/patternlock.html`, Last accessed: September 2010.

[115] H. Tao. Pass-Go, a new graphical password scheme. Master's thesis, School of Information Technology and Engineering, University of Ottawa, June 2006.

[116] H. Tao and C. Adams. Pass-Go: A proposal to improve the usability of graphical passwords. *International Journal of Network Security*, 7(2):273–292, 2008.

[117] F. Tari, A. Ozok, and S. Holden. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *2nd ACM Symposium on Usable Privacy and Security (SOUPS)*, 2006.

[118] J. Thames, R. Abler, and D. Keeling. A distributed active response architecture for preventing SSH dictionary attacks. In *IEEE Southeastcon*, 2008.

[119] J. Thorpe. *On the Predictability and Security of User Choice in Passwords*. PhD thesis, School of Computer Science, Carleton University, January 2008.

[120] J. Thorpe and P. C. van Oorschot. Graphical dictionaries and the memorable space of graphical passwords. In *13th USENIX Security Symposium*, August 2004.

[121] J. Thorpe and P. C. van Oorschot. Human-seeded attacks and exploiting hot-spots in graphical passwords. In *16th USENIX Security Symposium*, August 2007.

[122] E. Tulving and Z. Pearlstone. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5:381–391, 1966.

[123] E. Tulving and M. Watkins. Continuity between recall and recognition. *American Journal of Psych.*, 86(4):739–748, 1973.

[124] T. Valentine. An evaluation of the Passface personal authentication system. Technical report, Goldsmiths College Univ. of London, 1999.

[125] P. C. van Oorschot, A. Salehi-Abari, and J. Thorpe. Purely automated attacks on PassPoints-style graphical passwords. *IEEE Trans. Info. Forensics and Security*, 5(3):393–405, 2010.

[126] P. C. van Oorschot and J. Thorpe. On predictive models and user-drawn graphical passwords. *ACM Transactions on Information and System Security*, 10(4):1–33, 2008.

[127] P. C. van Oorschot and J. Thorpe. Exploiting predictability in click-based graphical passwords. *Journal of Computer Security*, 2011.

[128] P. C. van Oorschot and T. Wan. TwoStep: An authentication method combining text and graphical passwords. In *4th International MCETECH Conference on eTechnologies*, 2009.

[129] C. Varenhorst. Passdoodles: A lightweight authentication method. MIT Research Science Institute, July 2004.

[130] R. Virzi. Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors*, 34:457–468, 1992.

[131] K.-P. L. Vu, R. Proctor, A. Bhargav-Spantzel, B.-L. Tai, J. Cook, and E. Schultz. Improving password security and memorability to protect personal and organizational information. *International Journal of*

Human-Computer Studies, 65:744–757, 2007.

[132] R. Weber. The Statistical Security of GrIDsure. Technical report, University of Cambridge, 2006.

[133] D. Weinshall. Cognitive authentication schemes safe against spyware (short paper). In *IEEE Symposium on Security and Privacy*, May 2006.

[134] R. Weiss and A. De Luca. PassShapes – utilizing stroke based authentication to increase password memorability. In *NordiCHI*, pages 383–392. ACM, October 2008.

[135] C. Wharton, J. Bradford, R. Jeffries, and M. Franzke. Applying cognitive walkthroughs to more complex user interfaces: Experiences, issues, and recommendations. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 1992.

[136] A. Whitten and J. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *8th USENIX Security Symposium*, 1999.

[137] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. Authentication using graphical passwords: Basic results. In *11th International Conference on Human-Computer Interaction (HCI International)*, July 2005.

[138] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. Authentication using graphical passwords: Effects of tolerance and image choice. In *1st Symposium on Usable Privacy and Security (SOUPS)*, July 2005.

[139] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, 63(1-2):102–127, 2005.

[140] S. Wiedenbeck, J. Waters, L. Sobrado, and J. Birget. Design and evaluation of a shoulder-surfing resistant graphical password scheme. In *International Working Conference on Advanced Visual Interfaces (AVI)*, May 2006.

[141] M. Workman. Gaining access with social engineering: An empirical study of the threat. *Information Systems Security, Taylor & Francis Group*, 16(6):315–331, 2007.

[142] T. Wu. The secure remote password protocol. In *Network and Distributed System Security Symposium (NDSS)*, 1998.

[143] T. Wu. A Real-World Analysis of Kerberos Password Security. In *Proceedings of the 1999 Network and Distributed System Security Symposium (NDSS)*, February 1999.

[144] J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password memorability and security: Empirical results. *IEEE Security & Privacy Magazine*, 2(5):25–31, 2004.

[145] J. Yan, A. Blackwell, R. Anderson, and A. Grant. The memorability and security of passwords. In L. Cranor and S. Garfinkel, editors, *Security and Usability: Designing Secure Systems That People Can Use*, chapter 7, pages 129–142. O'Reilly Media, 2005.