Graphical Passwords: Learning from the First Generation

Robert Biddle, Sonia Chiasson, P.C. van Oorschot

***Abstract*—Numerous graphical password schemes have recently been proposed as alternatives to traditional text 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 should address, review methodological issues related to empirical evaluation, and identify areas for further research and improved methodology.**

***Index Terms*—Computer security, access control, user interface, human factors.**

I. Introduction

Beginning around 1999, numerous 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, parts of images, or sketches.

Despite the large number of options for authentication, text passwords remain the most common choice for several reasons [2], [3]. For example, 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 [4]–[8].

When text password users adopt unsafe coping strategies [9]–[11], such as reusing passwords across accounts to help with memorability, the resulting decrease in security cannot be successfully addressed by simply strengthening, in isolation, the underlying technical security aspects of a system. Usability issues often significantly impact the real-world security of the system. User interface design decisions may unintentionally sway user behaviour, often 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 (e.g., longer or more complex) passwords, precluding users from resorting to unsafe coping practices.

Surveys of graphical passwords circa 2005 are available from Suo et al. [10] and Monrose and Reiter [9]. More recently, Hafiz et al. [12] briefly summarize and categorize 12 schemes. Renaud [13] reviews numerous graphical password systems and offers usability guidelines for their design.

In this paper we provide a comprehensive review of the first ten years of published research on graphical passwords, and reflect on it. Reflection clearly shows that the graphical nature of schemes does not by itself avoid the problems typical of text password systems. However, while this first generation of graphical password schemes presents some familiar problems, we see an emerging second generation beginning to leverage the graphical elements in new ways to avoid the old problems.

We begin with an overview classifying schemes into three main categories — based on recall, recognition, and cued-recall — and centered on a primary exemplar of each category. We selectively discuss further schemes and extensions offering interesting additional characteristics and improvements, or where significant usability studies or security analysis has allowed a better understanding. We systematically review usability requirements and features for comparative analysis, and highlight specialized analysis techniques. We consider threat models, catalogue known attack strategies, and discuss the suitability of different schemes for various environments. 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.

Looking to the future, we consider methodological issues for evaluation of proposals, discuss challenges related to empirical evaluation, and extract lessons that can be learned from the research to date. We believe that if graphical passwords are to advance as a serious authentication alternative, research must be conducted and presented more 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.

Authors shown in alphabetical order; corresponding author is Chiasson, and paper is based on a chapter from her thesis [1]. Email: chiasson@scs.carleton.ca.

Version: October 2, 2009. Technical Report TR-09-09, School of Computer Science, Carleton University, Ottawa, Canada.

II. 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 [14]–[17]. The most widely accepted theory explaining this difference is the *dual-coding theory* [18], 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* [19]. We base our classification on these categories.

Recall requires that a person remember information without cueing. With recognition, a person is provided with the information and has to decide whether 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 whether they are similar and differ only in their retrieval difficulty [20]. It is generally accepted, however, that recognition is an easier memory task than recall [21], [22]. In cued-recall, an external cue is provided to help remember information. Tulving and Pearlstone [23] explain that items in human memory may be available but not accessible for retrieval. Their results show that previously inaccessible information in a pure recall situation can be retrieved with the aid of a retrieval cue.

III. 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 VIII.

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 smaller dictionary of likely passwords) so as to obtain an acceptable success rate within a manageable number of guesses. Guessing attacks may be conducted online through the intended login interface or offline if some verifiable text [24] (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 three 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 attack where users are tricked into entering their credentials at a fraudulent website that records users' input. Malware attacks use unauthorized software installed on client computers or servers to capture keyboard, mouse, or screen output, which is then parsed to find login credentials.

As will be seen in the following sections, early graphical password systems tended to focus on one particular strength, for example being resistent 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 reentry. 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, for storage to provide additional security in case the password file is compromised.

IV. RECALL-BASED SYSTEMS

Recall-based graphical password systems are occasionally referred to as *drawmetric systems* [25] 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 (which may arguably act as a mild memory cue). Recall is a difficult memory task [26] 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 memory. With text passwords, there is evidence that users often include the name of the system as part of their passwords [27], [28]. 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 related to recall-based systems here; see Section VIII for background and additonal 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 to 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. It is therefore possible that 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, such experimental results have been reported for password recovery mechanisms such as personal verification questions [29].

The following subsections provide an overview of recall-based graphical password schemes in the literature to date, centered on Draw-A-Secret [11]. Others are variations of it.

A. Canonical Example: Draw-A-Secret

Draw-A-Secret (DAS) [11] 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 [30] 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. Limitations of this preliminary study 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.

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 has cardinality 2^{58} [11]. This is often stated as 58 bits for brevity, but should not be mis-interpreted 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 VIII). 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.

Fig. 1. Sample Draw-A-Secret password [11]

In summary, the DAS design 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 [31] 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.

B. Other recall-based schemes

BDAS, proposed by Dunphy and Yan [32], 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. [33] proposed 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 [34], [35] is similar to DAS, allowing users to

create a freehand drawing as a password, but 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 [36] 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 access to both original and entered doodles to test if they are sufficiently similar.

Weiss and De Luca [37] proposed a similar system, PassShapes. 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 are acceptable according to the authors, but no security analysis has been reported.

The Pass-Go scheme (see Figure 2) designed by Tao [38] was motivated by an expected DAS usability issue: the difficulty of accurately duplicating sketches whose lines cross near ("too close" [11]) 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 system to date for which testing in a field study has been reported. Results of the large 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 VIII) may be less effective but attacks which exploit patterns [31], [39], for example, remain a concern.

A similar scheme was proposed by Orozco et al. [40], using a haptic input device that measures pen pressure while users draw their password. While this is intended to help protect against shoulder-surfing (an observer would have difficulty distinguishing variances in pen pressure), their user study showed

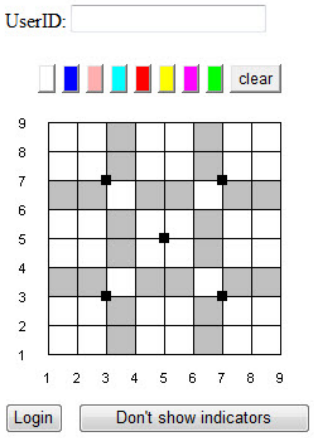

Fig. 2. Login screen for Pass-Go [38]

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. [41] 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 [42], a commercial product, displays a 5×5 grid of digits. For their password, users select and memorize a pattern consisting of an ordered subset of the 25 grid squares, and enter the corresponding digits therein using the 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. In a summary of a usability study [43] posted online, the reported login success rate exceeds 92% after 36 days. An initial security analysis by Weber [44] reported that grIDsure passwords were much more secure than traditional PINs, especially against shoulder-surfing. Independent analysis by Bond [45] notes several weaknesses in the scheme.

A grid-based system resembling a mini Pass-Go has also been deployed commercially for screen-unlock on Google Android cell phones. Rather than entering a 4-digit PIN, users touch-draw their password on a 3×3 grid.

These later recall schemes offer design and understanding that goes beyond that in 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.

V. Recognition-based Systems

Recognition-based systems, also known as *cognometric systems* [25] or *searchmetric systems* [13], generally require that users memorize a portfolio of images during password creation, and then to log in, must recognize their images from among decoys. Humans have exceptional ability to recognize images previously seen, even those viewed very

briefly [46], [47]. From a security perspective, such systems are not suitable replacements for text password schemes, 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). Recognition-based systems have been proposed using various types of images, most notably: faces, random art, everyday objects, and icons. Renaud [13] discusses specific security and usability considerations, and offers usability design guidelines focusing on recognition-based systems.

Phishing attacks are somewhat more difficult with recognition-based systems because the system must present the correct set of images to the user before password entry. This can be accomplished with a man-in-the-middle (MITM) attack, as noted in Section VIII, 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 information to the legitimate site, retrieve the panel of images from that site 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 seems to be 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 the images selected by users are larger discrete units that may be more easily identifiable and there are relatively few images (indeed, the theoretical password space is small). 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. The variation in 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 — e.g., 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 thus may be available to anyone gaining access to the stored information. An attacker with access to system-side files may gain the advantage of access to 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.

A. Canonical Example: PassFaces (and Faces)

The recognition-based system studied most extensively to date is PassFaces [48]. 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. Each round must be executed correctly for a successful login. The original test systems involved $n = 4$ rounds of $M = 9$ images per panel, with one image per panel from the user portfolio. The user portfolio contained exactly 4 faces, so all portfolio images were used during each login. The cardinality of the theoretical password space for PassFaces is M^n, with $M = 9$, $n = 4$ yielding $6561 \approx 2^{13}$ passwords.

In a study with 77 users, Valentine [49] 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 field study of Brostoff and Sasse [50] 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. [51] conducted a large field study where students used one of two graphical password schemes to access class material: Faces (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 VIII. To avoid this problem, a commercial PassFaces product [48] 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 [48] reports that password creation takes 3-5 minutes for a panel of 9 faces and 5 rounds.

Dunphy et al. [52] 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 portfolio 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. [53] 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 necessary. For further resistance against shoulder-surfing, Dunphy et al. [54] proposed and tested a version of PassFaces using eye-gaze as input at a simulated ATM

Fig. 3. PassFaces system. Left: sample panel from the original system [51]. Right: panel with decoys similar to the image from the user's portfolio [52].

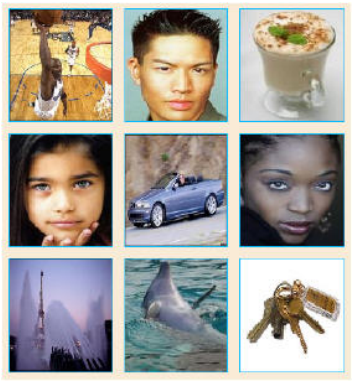

Fig. 4. Sample panel for the Story system [51].

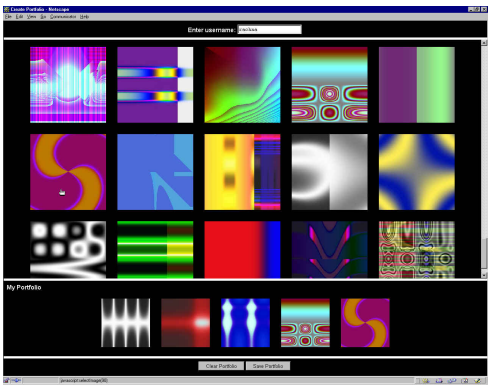

Fig. 5. Screenshot of the Déjà Vu graphical password system [56]

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. [55] 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.

B. Other recognition-based schemes

Story (see Figure 4) was proposed by Davis et al. [51] as a comparison system for PassFaces. 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. Images in their user study contained everyday objects, places, or people. 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 images in their set. In the test system, a panel had 9 images and a password involved selecting a sequence of 4 images from this panel.

Story was user-tested along with Faces in a field study. Davis et al. [51] 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 [56] (see Figure 5), users select and memorize a subset of "random art" images from a larger sample to create 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 is displayed, 5 of which belong 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 the images from their portfolio. 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 [56] 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 rigourously tested. Déjà Vu is somewhat more resistant to shoulder-surfing than previously described schemes, as only a portion of the user's portfolio is revealed during each login. Several logins would need to be observed to identify all images in a user's portfolio. Participants in the user study found it difficult to describe their portfolio images and users who had 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 password. 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 [57] proposed a graphical password 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 belonging to their previously

Fig. 6. Sample panel of Weinshall's cognitive authentication scheme [57]

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 right or bottom edge of the panel, 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 within a certain number of rounds, the login fails.

Users receive a system-assigned portfolio of 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 on over a period of 10 weeks.

The main claim [57] of resisting shoulder-surfing was proven false [58] (see Section VIII). Nonetheless, Weinshall's 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. Here N is the number of images in a panel, and M is the number of portfolio images displayed, e.g., N=80, M=30, yields $\binom{80}{30} = 2^{73}$ 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 effective size of 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 here only briefly. In the VIP system of De Angeli et al. [25], [59], 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 of Pering et al. [60], 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 pool of images collected from other users. Use Your Illusion, by Hayashi et al. [61], also requires that users select portfolio images from among panels of decoys; the selected images are distorted after original selection, the idea being that the legitimate user can still recognize the images despite the distortion, while the distortions create difficulties for others. The distortion is intended to protect against social engineering and shoulder-surfing attacks. In the Convex Hull Click Scheme of Wiedenbeck et al. [62], users once again 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.

Renaud [63] completed a field study comparing different types of user involvement in selecting the portfolio images for recognition-based schemes. Users could select images from a photo archive, could take their own photos, or could 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 from users to identify the correct response. A key issue with these early recognition-based schemes is the compromise between the size of the theoretical password space and usability in terms of memorability and time to log in. 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 [55] 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 the decoys also become familiar.

VI. Cued-recall Systems

Cued-recall systems typically require that users remember and target specific locations within a presented image. This

feature, intended to reduce the memory load on users, is an easier memory task than pure recall. Such systems may also be called *locimetric* [25] due to their reliance on identifying specific locations. This is a different memory task than simply recognizing an image as a whole. Hollingworth and Henderson [64] 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 1996 patent [65]. The PassPoints successor of that scheme launched research in the cued-recall subclass that we call *click-based graphical passwords*.

These 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 would be necessary to identify the image location. The screen scraper may 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.

A. Canonical Example: PassPoints

The literature on cued-recall graphical password systems is dominated by PassPoints [66]–[68] and its variations. During creation of a PassPoints password (see Figure 7), users are presented with an image. A password is a sequence of any $n = 5$ user-selected click-points (pixels) on this image. 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 prompt of the location of the originally chosen click-points. Note that this is not an optimal cued-recall scenario: users are presented with 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* [69], *centered discretization* [70], and *optimal discretization* [71] are possible alternatives. Kirovski et al. [72] 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. [66]–[68] 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. [73] conducted a lab study and a large 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. Later security analyses found it to be vulnerable to hotspots and simple patterns within images [39], [74]–[78], as elaborated in Section VIII. Bicakci et al. [79] conducted a lab study where a PassPoints password 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.

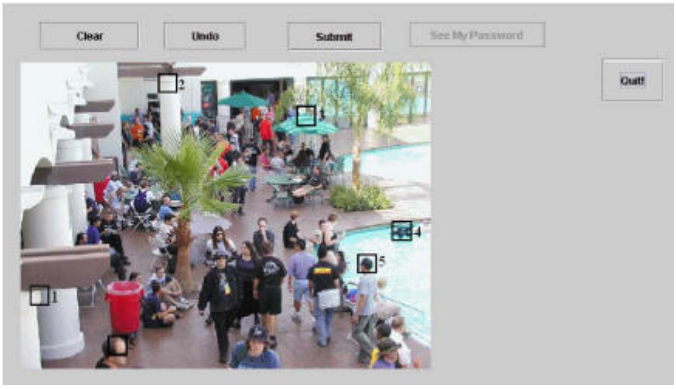

Fig. 7. PassPoints password example [67]. The 5 numbered boxes (not ordinarily visible to users) illustrate the tolerance area around click-points.

A commercial version of PassPoints for the PocketPC is available from visKey [80] 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.

B. Other cued-recall variants

PassPoints has received attention from others, who have proposed modifications. To address shoulder-surfing, Suo [81] proposes a shoulder-surfing resistant version as follows. During login, the image is blurred except for a small focus area. Rather than using a mouse to select click-points, 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) [82] is a click-based scheme where users select one click-point on each of 5 images

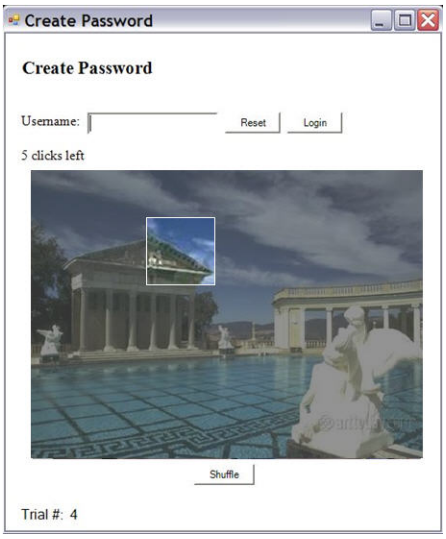

Fig. 8. Persuasive Cued Click-Points. During password creation, users select a click-point from the highlighted viewport or press the shuffle button to relocate the viewport.

presented in sequence, one at a time; this provides *one-to-one cueing*. Each image after the first is chosen using a deterministic function of the current image, the co-ordinates of the user-entered click-point, and a user identifier. Users receive immediate 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* [82] is not helpful to an attacker who does not know the expected sequence of images.

In a lab-based user study [82], 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 in click-points revealed that users tended to select click-points falling within known hotspots [83]. Further analysis [39] showed that simple patterns of click-points were eliminated (cf. PassPoints above).

Persuasive Cued Click-Points (PCCP) [83] 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 randomized viewport positions is to flatten the distribution of click-points across multiple users, to reduce the effects of hotspots.

A lab-based user study [83] found that click-points selected by PCCP participants did not fall into known hotspots, new hotspots were not formed, and PCCP passwords did not form simple patterns of click-points [39]. Login success rates were similar to the original CCP system. On average, participants took 50 seconds to create a password (an increase caused primarily by participants who shuffled repeatedly, though most shuffled relatively infrequently), and 8 seconds to log in.

Proposed implementations of PassPoints, CCP, and PCCP use a grid-based discretization algorithm, as mentioned earlier, for determining whether login click-points are within tolerance. In system-side storage for verification, these passwords can be hashed; additional information, however, is stored in a manner accessible to the system, such as a grid identifier (for each click-point) 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 information about the grid identifiers to their advantage.

Inkblot Authentication [84] (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 order, as cues during login and users enter each of their 2-character responses. It was suggested [84] 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 [84] that inkblots should be abstract enough that an attacker seeing the inkblots would not have an advantage in guessing a user's password.

Similarly, Jiminy [85], [86] 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 must 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 an open question.

Alsulaiman and El Saddik [87] 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 [87] 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.

Fig. 9. Inkblots used in the Inkblot Authentication user study [84]

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 [88] from spatial statistics has been used [83] 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 [39].

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 the existence of hotspots and geometric 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.

VII. 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 to date towards an accepted standard for evaluating the usability of graphical password schemes, nearly every system evaluated (if at all), has been tested 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 known, or widely practised, in the graphical password literature to date.

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 may result in a false sense of security. Interface design changes that appear to affect only usability may in fact introduce additional security vulnerabilities.

The remainder of this discussion is organized under the headings: target users, tasks, domains.

A. Target Users

As with other systems, 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 over time. There are also issues of accessibility that arise with graphical passwords since different user populations, such as the elderly [89], may 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.

B. 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).

1) Password Initialization: Authentication systems require initialization. A graphical password can either be assigned, or selected by the user. Training may be conducted, perhaps at least to compensate for the novelty of the scheme in comparison with more well-known approaches such as

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 chosen password having personal meaning may be easier to remember. However, this design decision has security disadvantages. As discussed later under security, graphical password systems that suffer from predictability problems related to user choice include the canonical examples of all three main categories: PassFaces, DAS (Pass-Go), and PassPoints. For example, from their study of Faces and Story, Davis et al. [51] conclude that allowing user choice leads to predictable patterns that can be 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. [90] found that text passwords are reused on an average of 6 different accounts. Many users also form some common strategy or pattern across accounts [91]. Both situations may be exploited by an attacker who acquires one of the passwords.

Systems which assign randomly selected passwords remove predictability, and also eliminate the potential for cross-account password reuse. However, such systems may require a time-consuming training process to help users remember their passwords (e.g., recall Weinshall's scheme [57]). 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. [55], which assigned passwords to avoid the predictability seen in earlier PassFaces studies, the order of password acquisition and login frequency significantly impacted password memorability.

It is possible for a system to allow partial user choice in password selection. For example, in PCCP (see Section VI-B), the middle-ground between allowing user choice and system-assigned passwords led to passwords nearly indistinguishable from random on the measures examined [39]. Further work is needed to evaluate the effect on long-term memorability.

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 VII-C below).

Text passwords have an advantage of being ubiquitous, and can be typed in a few seconds. It is thus natural to compare the time to enter a graphical password to that for a text password. Recognition-based schemes typically have the slowest password entry times, as users iterate through several rounds of images. On the other hand, some cued-recall schemes have been shown to have login times nearly as fast as text passwords [28].

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 points in time. 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. For the purpose of comparison, we recommend that, at minimum, success rates be reported 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 since memorability is a main factor determining whether login will be successful. Measures of memorability address whether passwords can be remembered over short- and long-term and with varying login frequencies. Strategies for testing memorability are discussed in Section IX-A.

Interference issues are also important. 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 [92], *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 [28], [55], [59] evaluating interference from multiple passwords are only now beginning to appear.

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 the password reset and password change mechanism 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 graphical password images should be avoided, and also argues against images being uploaded by users.

Similar to reset, most authentication systems must accommodate password change (some systems require this at specified intervals). The usability and security concerns are similar, except users can complete the task themselves without

requiring a temporary password, entering their current graphical password as authentication.

4) Portable login: Unless restricted to a very specific environment (e.g., while physically present in a corporate office, or at a bank ATM), it can be expected that 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.

C. Domains

Performance constraints and goals for an authentication system differ depending on the intended environment of use. While the highest level of security may be unnecessary for a newspaper subscription, due to the low risks and minor consequences of a security breach, an online banking site requires much stronger security. 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.

Ideally, an authentication system would be both highly secure and highly usable. Indeed, the ultimate goal of usable security is to increase both usability and security. However, in practice to date, the designs of many systems offer only the usual trade-off: the cost of increased security is decreased usability. To illustrate, many recognition-based graphical password schemes, when configured as proposed and tested in the literature, have theoretical password spaces approximately the same size as 4-digit PINs. Increasing the number of images per round or the number of rounds results in a larger password space but decreases usability and memorability. For PassFaces to achieve a theoretical password space comparable in size to 8-character passwords of alphanumeric characters (no symbols), it would require 62 images per panel and 8 rounds (versus the commonly reported 9 images per panel, and 4 rounds). This would result in longer login times and significantly complicate the recognition process. However, for environments where PIN-equivalent security suffices, schemes like PassFaces remain suitable.

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 result in improved 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 opportunities for 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 seems quite unlikely that any single scheme will be ideal for all domains, tasks, and target users, from a combined usability and security viewpoint. As such, specifying the particular environments and target applications for newly proposed schemes is important.

VIII. Security Aspects and Attacks

The main purpose of authentication mechanisms is to allow system access to only legitimate users. To thoroughly evaluate the security of a proposed graphical password system, and to facilitate comparison with alternatives, all standard threats and known attacks should be addressed. 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 simple brute-force attack that is a legitimate threat. 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 made and 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 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 in order to bypass the authentication system entirely, limiting our scope to attacks which directly obtain password credentials.

A. Guessing Attacks

Guessing attacks remain a serious threat [93]–[95], although statistics are scarce (few organizations publicize breaches). An *online guessing attack* requires interaction with the live system; usually each password guess is entered in turn to see if login is successful. Defenses against such attacks can be ameliorated by clever use of CAPTCHA's [96], [97]; increasingly delaying (e.g., doubling) the system response time on each successive incorrect guess; or limiting, per user account, the number of incorrect login attempts allowed before locking the account from further login attempts. The latter has usability costs: legitimate users who forget their password may be locked out. Moreover, locking out means that effective denial-of-service (DoS) attacks can be launched by intentionally entering incorrect passwords to prevent legitimate users from accessing their accounts [96]. Also, locking out is less effective against multi-account attacks [96], where instead of targeting a specific account, attackers try some number of guesses on each of many arbitrary accounts, seeking success on at least some accounts. This attack strategy works for both text and graphical password systems.

In an *offline guessing attack*, attackers gain access to verifiable text [24] and need not interact with the live system to test guesses for correctness. Schemes vulnerable to offline attacks are at a higher risk than those requiring online verification as offline work is not visible, and trial guesses can be processed much more quickly. Attacks may exploit pre-computed data structures and special hardware.

Defensive techniques against guessing attacks are numerous, and vary in utility depending on the environment. System-side passwords are typically 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 login attempt is correct, the system hashes the login input and tests for a match to the stored value. One technique to slow down guessing attacks is *iterated hashing* [4], requiring, say, 1000 or many more password hashing operations (rather than one); this increases the time to test individual password candidates online, and to pre-compute dictionaries. Another is *salting* [4], which 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-user (per-salt) basis, adding to the cost of pre-computed data structures. Designing authentication protocols to resist offline guessing attacks by avoiding verifiable text — such as *encrypted key exchange* (EKE), SRP and the like [24], [98], [99] — can be important for both text and graphical passwords, but is notoriously tricky.

Other long-standing defensive techniques include *password rules* or policies [4] to disallow weak passwords at their creation time, and both reactive and proactive *password checkers* (e.g., [5], [100]). *System-assigned passwords* — generated randomly, to preclude attacks exploiting password distributions — force use of the entire theoretical password space, but with high usability costs: longer training times and increased likelihood that users forget passwords. Mnemonic strategies like passphrases [7], [101] and Story [51] may potentially improve both usability and security, but may also suffer from predictability problems if user choice is allowed.

1) Exhaustive-search (brute-force) attacks: The defining characteristic of an exhaustive-search attack is that it exhaustively guesses all passwords within a password space. Such attacks are most often discussed in the context of cryptographic *key search* [4], where typical keys are system-assigned and equi-probable. For user-chosen passwords (which are far from equi-probable), except for small password spaces, dictionary attacks are preferred, as we discuss further below.

Exhaustive-search optimizations such as Oechslin's *rainbow tables* [102], which trade pre-computation time for storage, have been used for both cryptographic key search and alphanumeric password cracking. Coarse sequencing optimizations include first guessing shorter passwords and (for text passwords) lower-case only. Fine sequencing optimizations, which prioritize in decreasing order of expected probability, and favor specific subsets expected to hold higher probability passwords, are considered dictionary attacks.

The advantage to exhaustive attacks is that with enough time and computing power, all passwords will be found (unless an online attack is detected and stopped before completed). However, full search of large password spaces is infeasible. "Guaranteed" success may thus require more time or processing power than is available; searching only subsets is faster, but doesn't guarantee success.

To minimize the threat of exhaustive attacks, the set of all passwords allowed within a system (the *theoretical password space*) should too large to search. However, this is not the case for many recognition-based systems — e.g., the standard configuration of PassFaces has 9-image panels and 4 rounds, yielding only $9^4 = 6561$ passwords. In practice, such systems require complementary mechanisms such as limiting the number of online guesses per account, or multi-factor authentication. Helping the defense, 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.

2) Dictionary Attacks and Optimizations: Dictionary attacks on graphical passwords [51], [103] follow a long line of attacks on text passwords (e.g., [4], [5], [104]–[106]).

The original idea involved guessing passwords from a relatively short pre-compiled list (*dictionary*) of character sequences considered high-probability candidates, based on empirical data or assumptions about user behaviour. For online attacks, candidates of higher expected probability were ordered earlier. For systems involving (iterated) hashing or salting, these operations often allowed pre-computation, yielding "encrypted" dictionaries. Then for offline attacks, upon obtaining a list of hashed passwords (e.g., the Unix `/etc/passwd` password file), dictionary matches are found by table lookup. These ideas have been extended as technology has provided low-latency web access to login servers, greater computing power and storage, and improved search techniques. 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 [107]. The original defining aspect of a dictionary attack [4], a database, can also be replaced by algorithmic enumeration [107].

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 — *weak password subspaces* [31] 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; to prevent successful attacks, the effective password space must also be too large to search. The knowledge gap here is to understand what composes the effective password space, a problem still at best only vaguely understood for text passwords. Many graphical password proposals are susceptible to dictionary attacks due to predictable patterns in user choice, as we discuss next.

3) Specific Attacks on Graphical Password Schemes: We highlight that significant security issues are now known in the exemplars of each of the three major classes of graphical passwords. Moreover, few other schemes have received serious independent (if any) security scrutiny. Thus, it should be recognized that security claims made by proponents of various graphical password schemes are often optimistic.

RECALL-BASED SYSTEMS. DAS and Pass-Go have been studied with respect to dictionary and predictive attacks [31],

[38], [103], [108]. DAS passwords have been categorized into classes based on characteristics such as symmetry and a small number of strokes. Using this classification it was shown [31] that a large number of passwords from a paper-based study [30] and a subsequent Pass-Go field study [109], fall within such predictable categories. The field study also found [109] that a high percentage of Pass-Go users chose passwords from a third category, namely, drawings of alphabetic characters or symbols. It appears prudent to conclude that such predictable categories of relatively small cardinality will allow attackers to identify candidate passwords of higher probability, leading to efficient dictionary attacks. The security of Pass-Go thus warrants further study, as does that of DAS.

RECOGNITION-BASED SYSTEMS. For PassFaces, the analysis of user choice by Davis et al. [51] showed that users tend to select attractive faces of their own race; and that users selected predictable sets of faces such that an attacker knowing one face could determine the face most likely to be selected as the next password part. Because users tend to select predictable images, successful dictionary attacks may be expected, as well as personalized attacks, e.g., if attackers know a user's race or gender. Davis et al. [51] guessed 10% of passwords created by male participants in 2 guesses. A major conclusion was that many graphical password schemes, including Faces, may require "a different posture towards password selection" than text passwords, where selection by the user is the norm. As noted in Section V (which also mentions user choice issues in the Story scheme [51]), a phishing attack on PassFaces requires a MITM attack.

CUED-RECALL SYSTEMS. PassPoints' users tend to select passwords containing popular points (hotspots) or following simple patterns [39], [74]–[78]. *Hotspots* are areas of the image with higher probability of being chosen by users as individual click-points. *Simple patterns* are simple geometric shapes formed by the 5 click-points in a password. Both can be leveraged to launch efficient dictionary attacks. While partial success in locating hotspots has been reported using automated image processing techniques, a more efficient attack collects a small sample of passwords, on the image in question, from different users. Hotspots are extracted from this to build an attack dictionary — a small one wherein all password components are hotspots, or a larger one wherein some clickpoints are hotspots with others unconstrained. Many PassPoints passwords also follow geometric patterns [39], [77], such as a straight line; these can be exploited to prioritize guesses, and as the patterns are evident across a wide range of images, they may be used even without prior knowledge or analysis of the images.

Dictionary attacks against recognition and cued-recall graphical password systems require more effort up-front than against text passwords or recall-based graphical passwords, since attackers must first collect one or more of a set of images. Images gathered for one system will not help attacks on a second system, unless both systems use the same image set.

Text password crack tools (Crack [110], John the Ripper [111], RainbowCrack [112], and many others) are available to automate offline dictionary attacks. Some of these may be modified for online attacks. Similar cracking tools would likely surface for graphical passwords if the latter gain widespread usage. Text password attack tools are often generic, while some graphical attack tools may require system-specific images (though others, like the pattern-only attacks [77], have image-independent dictionaries).

B. Capture Attacks

Password capture attacks involve directly obtaining passwords, or part thereof, by capturing login credentials entered by the user, or tricking the user into divulging passwords. We assume that links over which graphical passwords are sent are encrypted, otherwise simple network sniffing or wire-tapping allows trivial capture. New graphical password proposals must consider the following known classes of capture attack.

1) Shoulder-surfing: Shoulder-surfing [53], [113]–[115] is a targeted attack exacerbated by the visual aspect of graphical passwords. As the user enters login information, an attacker may gain knowledge about their credentials through direct observation or external recording devices such as video cameras. As examples of the range of related attacks, text information can be gleaned from telephoto images of computer screens reflected on nearby items [114], while physical keyboard entry has been identified from telephoto lens images as far as 195 feet away [115]. High-resolution cameras with telephoto lenses and surveillance equipment make shoulder-surfing a real concern if attackers target specific users and have access to their geographic location. While problematic in public environments, shoulder-surfing may not be as serious a threat in private environments; far less academic attention has been devoted to more relevant threats such as keystroke loggers or graphical dictionary attacks.

For some recognition-based graphical passwords, multiple successful logins must be observed to deduce the full secret — 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 graphical systems can be gathered from observing or recording one successful login.

Existing graphical schemes believed to be resistant or immune to shoulder-surfing [62], [116] have significant usability drawbacks, usually in the time and effort required to log in, making them less suitable for everyday authentication.

2) Reconstruction: Some attacks involve password reconstruction rather than direct capture. For example, Weinshall's scheme [57], designed specifically to resist shoulder-surfing, was shown by Golle and Wagner [58] to fall to a SAT (boolean satisfiability problem) solver, which reconstructs user secrets in a few seconds on observing a small number of logins. Acoustic-based reconstruction attacks on text passwords, such as the password cracker of Berger et al. [117], seem less suited to graphical passwords, though ideas from the reconstruction techniques may be of use.

3) Malware: Malicious software includes any unauthorized software installed for malicious purposes and without a user's informed consent, including computer viruses and worms, Trojan horse software including login spoofing, code silently installed as a result of visiting web sites [118], and mobile code in the form of JavaScript, ActiveX, or Flash components.

Such malware may gather password information. *Keystroke-loggers* [119] record keyboard input; *mouse-loggers* and *screen scrapers* capture mouse actions and record screen memory, to be sent remotely or made available for retrieval. Text passwords can be captured using only a keystroke-logger. Most graphical password systems require one or both of a mouse-logger and screen scraper to capture passwords, and often a keystroke-logger as well to collect usernames. Keystroke-loggers alone may suffice for schemes like Inkblot Authentication (Section VI), which use keyboard only. If graphical passwords gain popularity, such malware will likely do so also.

4) Phishing and pharming: Phishing attacks [120] 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 recognition-based and cued-recall systems, specific images must be presented to the user. To do so, a phishing site may retrieve and relay information from the legitimate site, in a *man-in-the-middle* (MITM) attack. *Pharming* [121], 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 web site. Depending on design of 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 correct authentication response provided during the attack.

5) Social engineering: Phishing is one form of social engineering attack [122], [123] — for malicious purposes, tricking users 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 (and knowledge of personal details in *personalized attacks*), this is often easier than otherwise breaking into a system [122].

Text passwords and alphanumeric information are relatively easy to share, with colleagues or attackers. Sharing is more difficult for graphical passwords; a frame of reference must be coordinated, before conveying the password in sufficient detail to be used. This security advantage (complicating social engineering attacks) has usability drawbacks, e.g., preventing password reset by phone, and complicating safe backup storage of passwords. Despite the additional difficulty, Dunphy et al. [52] have preliminary evidence that users can sufficiently describe their PassPoints password to enable someone else to enter it. Other means of sharing a graphical password include taking photos, screen shots, and drawing.

IX. Methodology for Evaluation

As with other types of authentication mechanisms, 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.

With usability inspection methods (such as *cognitive walkthroughs* [124] and *heuristic evaluations* [125]), 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 [125]. 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 almost inevitably become a focus when user tests are conducted, which may lead to behaviour that is not representative of what would happen if the system were 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 if using the system regularly.

Since text passwords are the most common knowledge-based authentication mechanisms, they are often used as a benchmark to assess the usability and security of graphical password schemes. While useful, this comparison is biased because users have years of 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 provided to users can influence their behaviour. Users may be more comfortable and display behaviour 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 given 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 take into account potential effects when interpreting the results of user studies.

The problem of testing for multiple passwords also needs special consideration. Recent publications [28], [55], [59] 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 the frequency of login may have significant impact on the study 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.

There are three general approaches to user testing graphical password systems: lab studies, field studies, and hybrid studies. Each can provide valuable empirical data.

A. 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 choice of 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 unforseen 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 technical 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 potentially 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 [126], [127] — has long been used to justify small usability studies. Recent work revisits this assumption, highlighting that this is often not enough and that in some cases, severe usability problems are only discovered after running a larger group of participants [128]–[130]. 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 number of problems found by studying any one user also makes it unlikely that a sample of five users would discover most usability problems. Faulkner [128] 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". 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 graphical password studies have multiple lab sessions, where participants return at fixed intervals to re-enter their passwords over the course of several days, weeks, or months. Such studies, however, that only require that users remember a single 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.

B. 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, whether the scheme is usable on computer systems with different configurations (e.g., screen sizes), and whether circumstances such as interference from multiple passwords or password use in environments where shoulder-surfing is possible causes problems not apparent in the lab. Real-world usage is of particular concern with security systems because security is often a secondary task [131], 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 already have a suspicion that particular behaviours are occurring in order to investigate them. Users may not necessarily realize that some behaviours are insecure or worthy of mention unless specifically prompted.

C. Other types of studies

1) Web-based: Other types of user studies are gaining popularity, for example, unsupervised web-based studies [55], [59], [90], [132]. The advantages are that 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 to consider 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.

2) Hybrid: In hybrid studies, researchers combine lab studies with tasks completed in participants' regular environment, gaining advantages of both an initial controlled environment and increasing ecological validity in the subsequent tasks. The tasks are usually fictitious, 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 enter their passwords 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.

X. Further Discussion and Conclusion

Our tour of graphical password research to date has revealed a rich palette of ideas, but few schemes that truly deliver on the original promise of addressing the problems seen in text passwords. Indeed, careful examination of the first generation of graphical password schemes indicates that many of the same problems continue to re-surface.

In assessing usability, an apples-to-apples comparison requires comparing schemes of equivalent security (Figure 10).

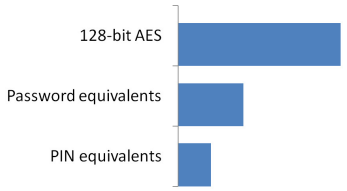

Fig. 10. Abstraction (log-scale) for keeping relative security in context. Usability should be compared with schemes of comparable security levels.

It is less meaningful to compare the usability of two schemes offering vastly different security propositions, and if done, this should be explicitly acknowledged. For example, most recognition-based systems have theoretical password spaces comparable to 4-digit PINs, while recall and cued-recall systems have theoretical password spaces more similar in size to 8-character-or-more text passwords. Somewhat longer login times may be acceptable for password-equivalent systems than for PIN-equivalent systems, if they provide greater security where it is needed.

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, analysis that explains 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 proposed graphical password systems lack rigorous evaluation in security or usability (or both). A closer look at individual systems has typically revealed less security than promised, matching historical experience in other areas. Significant security flaws have been found in all three canonical examples (original versions) discussed in this paper. These and other proposed systems 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. New designs should focus on increasing entropy without sacrificing usability and memorability.

In many systems having poor security, users appear to have compromised 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

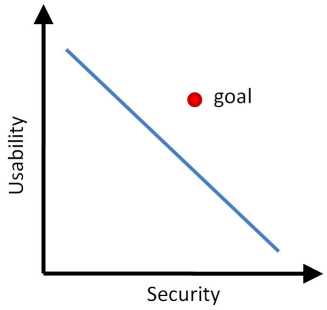

Fig. 11. Most graphical password schemes to date fall along the descending line, where increased security implies decreased usability. The goal of design for usable security is to increase both usability and security simultaneously.

devised. A complementary method for addressing predictable passwords is the use of so-called "strong" password protocols (e.g., SRP [99], EKE [98]) designed to provide protection against offline dictionary attacks.

For usability, a major concern is multiple password interference. Since graphical passwords are not widely deployed, it is unknown whether we will simply mirror the problems with text passwords where users develop coping strategies, devise and reuse common patterns, and choose minimally secure passwords. The 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.

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 but at the cost of an additional memorability burden since each additional round also exposes users to a new set of decoys. As illustrated in Figure 11, the challenge for the second generation of graphical password schemes, and in the *design for usable security* in general, is instead to find new designs and architectures which afford increases in security and usability together.

ACKNOWLEDGMENTS

The first author acknowledges funding of a Discovery Grant through the Natural Sciences and Engineering Research Council of Canada (NSERC). The third author is Canada Research Chair in Internet Authentication and Computer Security, and acknowledges NSERC funding of this chair, a Discovery Grant, and a Discovery Accelerator Supplement. Partial funding from the NSERC Internetworked Systems Security Network (ISSNet) is also acknowledged.

REFERENCES

[1] S. Chiasson, "Usable authentication and click-based graphical passwords," Ph.D. dissertation, School of Computer Science, Carleton University, December 2008.

[2] K. Renaud, "Evaluating authentication mechanisms," in *Security and Usability: Designing Secure Systems That People Can Use*, L. Cranor and S. Garfinkel, Eds. O'Reilly Media, 2005, ch. 6, pp. 103–128.

[3] 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.

[4] R. Morris and K. Thompson, "Password Security: A Case History," *Communicaitons of the ACM*, vol. 22, no. 11, pp. 594–597, 1979.

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

[6] 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*, vol. 19, no. 3, pp. 122–131, July 2001.

[7] J. Yan, A. Blackwell, R. Anderson, and A. Grant, "Password memorability and security: Empirical results," *IEEE Security & Privacy Magazine*, vol. 2, no. 5, pp. 25–31, 2004.

[8] J. Bentley and C. Mallows, "How much assurance does a PIN provide?" in *Human Interactive Proofs (HIP), LNCS 3517, Springer-Verlag*, H. Baird and D. Lopresti, Eds., 2005, pp. 111–126.

[9] F. Monrose and M. Reiter, "Graphical passwords," in *Security and Usability: Designing Secure Systems That People Can Use*, L. Cranor and S. Garfinkel, Eds. O'Reilly Media, 2005, ch. 9, pp. 157–174.

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

[11] 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.

[12] 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*. IEEE, 2008, pp. 396–403.

[13] K. Renaud, "Guidelines for designing graphical authentication mechanism interfaces," *International Journal of Information and Computer Security*, vol. 3, no. 1, pp. 60–85, June 2009.

[14] B. Kirkpatrick, "An experimental study of memory," *Psychological Review*, vol. 1, pp. 602–609, 1894.

[15] S. Madigan, "Picture memory," in *Imagery, Memory, and Cognition: Essays in Honor of Allan Paivio*, J. Yuille, Ed. Lawrence Erlbaum Associates, 1983, ch. 3, pp. 65–89.

[16] A. Paivio, T. Rogers, and P. C. Smythe, "Why are pictures easier to recall than words?" *Psychonomic Science*, vol. 11, no. 4, pp. 137–138, 1968.

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

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

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

[20] J. Anderson and G. Bower, "Recognition and retrieval processes in free recall," *Psychological Review*, vol. 79, no. 2, pp. 97–123, March 1972.

[21] E. Tulving and M. Watkins, "Continuity between recall and recognition," *American Journal of Psych.*, vol. 86, no. 4, pp. 739–748, 1973.

[22] W. Kintsch, "Models for free recall and recognition," in *Models of Human Memory*, D. Norman, Ed. Academic Press: New York, 1970.

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

[24] L. Gong, M. Lomas, R. Needham, and J. Saltzer, "Protecting poorly chosen secrets from guessing attacks," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 5, pp. 648–656, June 1993.

[25] 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*, vol. 63, no. 1-2, pp. 128–152, 2005.

[26] F. Craik and J. McDowd, "Age differences in recall and recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 13, no. 3, pp. 474–479, July 1987.

[27] 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*, vol. 65, pp. 744–757, 2007.

[28] 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.

[29] 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.

[30] D. Nali and J. Thorpe, "Analyzing user choice in graphical passwords," School of Computer Science, Carleton University, Tech. Rep. TR-04-01, May 2004.

[31] P. C. van Oorschot and J. Thorpe, "On predictive models and user-drawn graphical passwords," *ACM Transactions on Information and System Security*, vol. 10, no. 4, pp. 1–33, 2008.

[32] 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.

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

[34] 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.

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

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

[37] R. Weiss and A. De Luca, "PassShapes – utilizing stroke based authentication to increase password memorability," in *NordiCHI*. ACM, October 2008, pp. 383–392.

[38] H. Tao and C. Adams, "Pass-Go: A proposal to improve the usability of graphical passwords," *International Journal of Network Security*, vol. 7, no. 2, pp. 273–292, 2008.

[39] 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*, vol. 8, no. 5, 2009.

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

[41] 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*, vol. 5, no. 6, pp. 943–952, June 2008.

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

[43] M. A. Sasse, "GrIDsure usability trials," http://www.gridsure.com/uploads/UCL%20Report%20Summary%20.pdf, accessed August 2009.

[44] R. Weber, "The Statistical Security of GrIDsure," http://www.gridsure.com/uploads/Stats%20report%20-%20Richard%20Weber.pdf, University of Cambridge, Tech. Rep., June 2006.

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

[46] L. Standing, J. Conezio, and R. Haber, "Perception and memory for pictures: Single-trial learning of 2500 visual stimuli," *Psychonomic Science*, vol. 19, no. 2, p. 7374, 1970.

[47] D. Nelson, V. Reed, and J. Walling, "Pictorial Superiority Effect," *Journal of Experimental Psychology: Human Learning and Memory*, vol. 2, no. 5, pp. 523–528, 1976.

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

[49] T. Valentine, "An evaluation of the Passface personal authentication system," Goldsmiths College Univ. of London, Tech. Rep., 1999.

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

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

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

[53] 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.

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

[55] 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.

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

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

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

[59] 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.

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

[61] 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.

[62] 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.

[63] K. Renaud, "On user involvement in production of images used in visual authentication," *Journal of Visual Languages and Computing*, vol. 20, no. 1, pp. 1–15, February 2009.

[64] A. Hollingworth and J. Henderson, "Accurate visual memory for previously attended objects in natural scenes," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 28, no. 1, pp. 113–136, 2002.

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

[66] 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.

[67] ——, "PassPoints: Design and longitudinal evaluation of a graphical password system," *International Journal of Human-Computer Studies*, vol. 63, no. 1-2, pp. 102–127, 2005.

[68] ——, "Authentication using graphical passwords: Basic results," in *11th International Conference on Human-Computer Interaction (HCI International)*, July 2005.

[69] J. Birget, D. Hong, and N. Memon, "Graphical passwords based on robust discretization," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 3, pp. 395–399, 2006.

[70] 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.

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

[72] 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)*, S. Fischer-Hbner, K. Rannenberg, L. Yngstrm, and S. Lindskog, Eds., vol. 201. Boston: Springer, 2006, pp. 351–363.

[73] S. Chiasson, R. Biddle, and P. C. van Oorschot, "A second look at the usability of click-based graphical passwords," in *3rd ACM Symposium on Usable Privacy and Security (SOUPS)*, July 2007.

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

[75] 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.

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

[77] A. Salehi-Abari, J. Thorpe, and P. C. van Oorschot, "On purely automated attacks and click-based graphical passwords," in *24th Annual Computer Security Applications Conference (ACSAC)*, 2008.

[78] P. C. van Oorschot and J. Thorpe, "On predicting and exploiting hotspots in click-based graphical passwords," School of Computer Science, Carleton University, Tech. Rep. TR-08-21, November 2008.

[79] 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.

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

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

[82] 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*, September 2007, pp. 359–374.

[83] 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.

[84] A. Stubblefield and D. Simon, "Inkblot Authentication, MSR-TR-2004-85," Microsoft Research, Tech. Rep., 2004.

[85] K. Renaud and E. Smith, "Jiminy: Helping user to remember their passwords," School of Computing, Univ. of South Africa, Tech. Rep., 2001.

[86] K. Renaud and A. D. Angeli, "My password is here! An investigation into visio-spatial authentication mechanisms," *Interacting with Computers*, vol. 16, no. 4, pp. 1017–1041, 2004.

[87] F. Alsulaiman and A. El Saddik, "A novel 3D graphical password schema," in *IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, July 2006.

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

[89] K. Renaud, "A visuo-biometric authentication mechanism for older users," in *British HCI*, September 2005, pp. 167–182.

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

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

[92] M. Anderson and J. Neely, *Memory. Handbook of Perception and Cognition*, 2nd ed. Academic Press, 1996, ch. 8, pp. 237–313.

[93] 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.

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

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

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

[97] P. C. van Oorschot and S. Stubblebine, "On countering online dictionary attacks with login histories and humans-in-the-loop," *ACM Trans. on Info. and System Security*, vol. 9, no. 3, pp. 235–258, 2006.

[98] 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.

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

[100] F. Bergadano, B. Crispo, and G. Ruffo, "High dictionary compression for proactive password checking," *ACM Transactions on Information and System Security*, vol. 1, no. 1, pp. 3–25, 1998.

[101] C. Kuo, S. Romanosky, and L. Cranor, "Human selection of Mnemonic Phrase-based Passwords," in *2nd ACM Symposium on Usable Privacy and Security (SOUPS)*, July 2006.

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

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

[104] D. C. Feldmeier and P. R. Karn, "UNIX Password Security – Ten Years Later," in *Crypto'89*, August 1989.

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

[106] J. Yan, A. Blackwell, R. Anderson, and A. Grant, "The memorability and security of passwords," in *Security and Usability: Designing Secure Systems That People Can Use*, L. Cranor and S. Garfinkel, Eds. O'Reilly Media, 2005, ch. 7, pp. 129–142.

[107] A. Narayanan and V. Shmatikov, "Fast dictionary attacks on passwords using time-space tradeoff," in *12th ACM Conference on Computer and Communications Security (CCS)*, November 2005.

[108] J. Thorpe and P. C. van Oorschot, "Towards secure design choices for implementing graphical passwords," in *20th Annual Computer Security Applications Conference (ACSAC)*, December 2004.

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

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

[111] S. Designer, "John the Ripper password cracker," http://www.openwall.com/john/.

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

[113] 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.

[114] 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.

[115] B. Laxton, K. Wang, and S. Savage, "Reconsidering physical key secrecy: Teleduplication via optical decoding," in *15th ACM Conference on Computer and Communications Security (CCS)*, 2008.

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

[117] 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.

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

[119] 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.

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

[121] 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.

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

[123] M. Workman, "Gaining access with social engineering: An empirical study of the threat," *Information Systems Security, Taylor & Francis Group*, vol. 16, no. 6, pp. 315–331, November 2007.

[124] 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.

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

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

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

[128] L. Faulkner, "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing," *Behavior Research Methods, Instruments, & Computers*, vol. 35, no. 3, pp. 379–383, 2003.

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

[130] 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.

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

[132] 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*, vol. 16, no. 2, pp. 185–210, 2003.