Exploring the Design Space of Graphical Passwords on Smartphones

Florian Schaub, Marcel Walch, Bastian Könings, and Michael Weber
Institute of Media Informatics
Ulm University
89069 Ulm, Germany
{ florian.schaub | marcel.walch | bastian.koenings | michael.weber }@uni-ulm.de

ABSTRACT

Smartphones have emerged as a likely application area for graphical passwords, because they are easier to input on touchscreens than text passwords. Extensive research on graphical passwords and the capabilities of modern smartphones result in a complex design space for graphical password schemes on smartphones. We analyze and describe this design space in detail. In the process, we identify and highlight interrelations between usability and security characteristics, available design features, and smartphone capabilities. We further show the expressiveness and utility of the design space in the development of graphical passwords schemes by implementing five different existing graphical password schemes on one smartphone platform. We performed usability and shoulder surfing experiments with the implemented schemes to validate identified relations in the design space. From our results, we derive a number of helpful insights and guidelines for the design of graphical passwords.

Categories and Subject Descriptors

H.1.2 [**Models and principles**]: User/machine systems—*Human factors*; K.6.5 [**Management of Computing and Information Systems**]: Security and protection—*Authentication*; H.5.2 [**Information interfaces and presentation**]: User interfaces—*Interaction styles*

General Terms

Experimentation, Security, Human Factors.

Keywords

Authentication; graphical passwords; mobile interaction; multitouch; shoulder surfing; smartphone; usability.

1. INTRODUCTION

Graphical authentication mechanisms have the potential to overcome certain issues with text-based passwords, such as password memorability and the lack of recall cues, because visual representations are more memorable and easier to recall [24]. Graphical passwords and smartphones with touchscreens seem a natural fit, as they often require direct selection or manipulation of visual elements. In contrast, text passwords have to be entered on virtual keyboards on which typing effort varies between characters, with special characters requiring up to four touch events [44]. So far, not many graphical password schemes are in actual use on smartphones. A notable exception is Android's Pattern Lock, which requires users to draw a symbol on a 3x3 grid to unlock the phone's screen. While simple to use, passwords based on drawing patterns are not resistant against shoulder surfing [16]. Many shoulder surfing enhancements have been proposed, as will be discussed later on. While graphical passwords are generally distinguished by the method of password memorization as recall, recognition, and cued-recall schemes [4], additional aspects, such as the interaction method and presentation of cues, also affect security, usability and shoulder surfing resistance [15, 51]. The resulting design space is quite large as evidenced by the number of proposed graphical password schemes. Smartphone-specific features, like single- and multi-touch input or sensors, further extend this design space, but also introduce constraints, e.g., smaller screen size and reduced pointer accuracy compared to desktop use [12].

As pointed out by Biddle et al. [4], diversity in design aspects, variations in evaluation setup, and different target platforms make it difficult to compare graphical password schemes for mobile devices in terms of usability and security. We address this issue by analyzing the design space for graphical passwords on smartphones, in order to support researchers and developers who design, implement or use graphical password schemes on such devices. The aspects of our design space are grounded in existing work and related studies. In order to show the applicability of the design space, we re-implemented five existing schemes on the same smartphone platform for sake of comparison and analyze how smartphone capabilities impact certain design aspects based on our implementation experience with those schemes. We further performed a comparative user study with these schemes to assess how different instantiations of design aspects affect usability and shoulder surfing. Based on the study results, we discuss relations between different aspects in the design space and derive guidelines for the design of graphical password schemes on smartphones.

We discuss related work in Section 2 before presenting the design space in Section 3. In Section 4, we show how

Symposium on Usable Privacy and Security (SOUPS) 2013, July 24–26, 2013, Newcastle, UK.

graphical password schemes can be mapped onto the design space, and how constraints of the same platform impact the implementation of each scheme. We discuss our results of usability and shoulder surfing experiments performed with those schemes in Sections 5 and 6. The discussion in Section 7 relates the study results to the proposed design space and identifies respective implications, which result in a set of insights and guidelines for graphical password schemes on smartphones. Section 8 concludes the paper.

2. RELATED WORK

Graphical passwords are knowledge-based authentication mechanisms. Effects of different aspects on knowledge-based authentication have been studied extensively, with a major focus on text passwords, resulting in many general recommendations and design guidelines. Instead of repeating those here, we primarily focus on related work pertaining directly to graphical passwords.

Bonneau et al. [6] provide an extensive analysis of different proposals for replacing text passwords in web authentication. They identify a set of requirements for password replacement schemes focusing on usability, deployability, and security. In regard to graphical passwords, they analyzed Persuasive Cued Click Points [10] as an exemplary scheme. They find that graphical passwords are not effortless in terms of memorability but offer advantages over text-passwords as images can be used as cues for different passwords. They further point out that graphical passwords are easy to learn, but typically require longer entry times than text passwords—at least in the web and desktop context. They also note that graphical passwords have low accessibility, because they rely on the recognition of and interaction with visual elements. Graphical passwords are considered not resilient to physical observation as evidenced by many experiments on shoulder surfing and are potentially vulnerable to guessing attacks. However, Bonneau et al. do not discuss the effect of specific design aspects on the characteristics of graphical passwords. Herley & van Oorschot [27] note that it is unlikely to achieve all security, usability, and economic requirements in one authentication scheme. Instead, requirements need to be considered and weighted in light of the specific application domain—authentication on smartphones in our case—a point also supported by others [4, 15]. They also caution against viewing usability and security as a one-dimensional tradeoff. For example, certain usability improvements may increase the potential of observation attacks, but the associated risk may be lower than for brute force guessing attacks. Our proposed design space provides a more nuanced understanding of this relationship in the context of smartphones.

Biddle et al.'s survey on graphical passwords [4] provides a comprehensive overview of existing graphical password schemes with focus on memorability, security, and usability aspects. They further analyze relations between security and usability aspects. In terms of security, they consider a scheme's *theoretical password space*, *user-choice resilience* (i.e., if the password distribution is skewed by user chosen passwords), *variant response* (i.e., does password entry vary between attempts to thwart replay attacks), and the need for *server probes* (i.e., how many requests to the server are required in order to mount a phishing attack). In terms of usability, they compare required *entry time* and *success rate* for logins. In terms of memorability, they assess what kind of studies have been performed (lab, field, or web), and if *password interference* has been studied. They find that results from different user studies are often hard to compare due to variations in study setup and assessed parameters. Nevertheless, they derive a list of design recommendations for graphical password schemes. Similarly, De Angeli et al. [15] and Renaud [40] propose specific design guidelines for recognition-based graphical passwords based on user studies. Those guidelines, where applicable to smartphones, will be discussed inline with the design space in Section 3.

Beyond those general guidelines, few domain- or platform-specific design guidelines have been proposed. Kim et al. [31] study design considerations for authentication in collaborative multi-touch interaction on tabletops. Based on the domain's design opportunities and constraints, they identify potential strategies for improving shoulder surfing resistance by reducing visibility and dissipating attention of observers, subdividing authentication actions, and knowledge transformation. While their proposed mechanisms are optimized for tabletops and require more screen estate than available on smartphones, their general strategies can also inform graphical password design for smartphones. Dunphy et al. [20] specifically consider the design of recognition-based graphical password schemes on mobile devices. They analyze the effects of the number of images presented to the user for recognition on shoulder surfing resistance and propose methods for utilizing the user's personal images.

3. DESIGN SPACE

The discussion in Section 2 shows that most existing studies and guidelines relating to design aspects of graphical passwords do not explicitly consider the target platform, although the need for doing so is commonly recognized [4, 15, 27]. This juxtaposition is not caused by oversight but rather a desire to formulate guidelines pertaining to password characteristics that are applicable and adaptable to arbitrary platforms. We argue, however, that explicit consideration of the target platform is necessary to yield a description of the design space that accurately reflects intricate relationships between design aspects and capabilities facilitated by or inherent to the target platform, as evidenced by the few existing platform-specific studies [20, 31]. In this section, we map out the design space for graphical passwords on smartphones by relating existing research and generic guidelines to smartphone characteristics. The result is a comprehensive description of this design space and its complexity.

We distinguish three major groups of aspects that are part of the design space for graphical passwords on smartphones, which are highlighted in different colors in Figure 1. *Design features* correspond to available design choices in the process of developing a graphical password scheme, e.g., how the user interacts with the scheme or the spatiotemporal arrangement of visual elements. The platform-specific capabilities (*smartphone capabilities*), such as screen size, extend or restrict the set of design features, their available parameters, and how they can be realized on specific devices. Both, design features and smartphone capabilities impact the *password characteristics* of the resulting graphical password scheme. From another perspective, specific requirements for password characteristics can also determine meaningful design features. By assessing aspects of each group in detail, our analysis of the design space revealed nuanced influences and complex relations between these groups

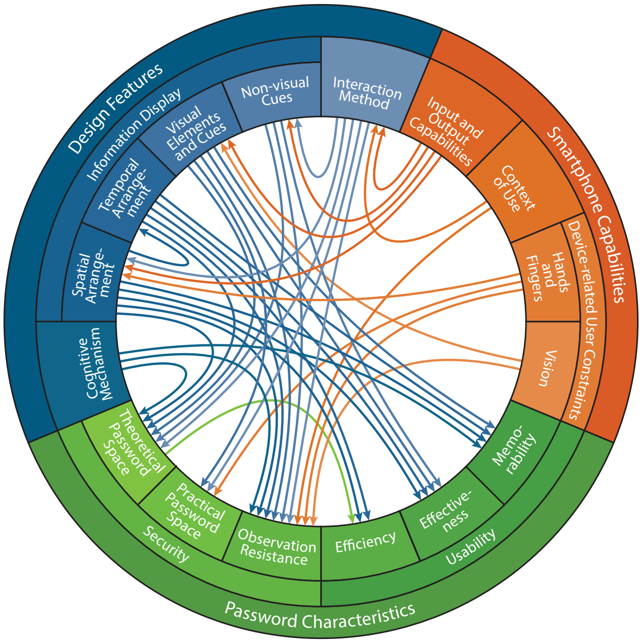


Figure 1: The components of the smartphone-specific design space for graphical passwords and their interrelations.

(shown as arrows in Fig. 1), which we will discuss in detail. We begin with the discussion of password characteristics, because they are mainly influenced by aspects from the other two groups, and these relations are easier to understand once those characteristics have been defined.

3.1 Password Characteristics

Security and usability are two major characteristics of knowledge-based authentication mechanisms, such as graphical passwords schemes. However, security and usability should not be seen as a one-dimensional trade-off [27]. Graphical password schemes typically strive to enhance usability without compromising security [4].

3.1.1 Security

Multiple security aspects have to be considered in the design of an authentication scheme [6], ranging from encoding of stored passwords to resistance against offline and over-the-network attacks. Here, we focus only on interaction-related or smartphone-specific security aspects.

Theoretical password space.
The theoretical password space of a graphical password scheme is a security strength indicator determined by the *total number of possible passwords*. The theoretical password space assumes an equiprobable distribution of passwords. Its size corresponds to the theoretical computational effort of an adversary guessing the password with exhaustive search over all potential passwords (brute force). As for text passwords, the size of the theoretical password space depends on the number of *available characters* and the *password length*. For graphical passwords, the "character set" is defined by the employed visual elements and how a password is specified. The password length corresponds to the length of a recognition or recall sequence. The theoretical password space should correspond to the security required by the intended domain or application [4].

Practical password space.
Statistical password distributions are often not equiprobable due to scheme-dependent predictability of user choices [5]. Davis et al. [14] identified gender-specific selection biases in recognition-based graphical passwords using face images. Recall schemes based on drawing shapes have been shown to suffer from predictive patterns [35, 52]. In cued-recall schemes that require users to select points on an image, users have been found to prefer salient image features (*hot spots*) [25, 53]. Thus, the practical password space is defined by the *probability of an adversary guessing the password* based on the statistical distribution of passwords for a given scheme [5]. Skewed password distributions, predictive models, and graphical dictionaries [51] can be leveraged in such attacks, resulting in a practical password space smaller than the theoretical one, which reduces the practical security of a scheme. For many schemes, optimizations have been proposed as design features to influence user choice towards flatter and less predictable password distributions [4].

Observation resistance.
Observation attacks [6], also called capture attacks [4], aim to obtain a user's password through observation of the login process. In the context of graphical passwords, *shoulder surfing* is considered a major threat [49], i.e., a bystander obtaining the password by observing the user entering it. However, *internal observation* (i.e., key logging) should also be considered [4, 6]. *Video-based shoulder surfing* records the login process on video for later analysis. Many design features have been proposed to increase resistance of graphical passwords against observation attacks.

3.1.2 Usability

Text passwords and PINs are commonly used as benchmarks for alternative authentication methods. Therefore, graphical passwords should strive to match, or better exceed, the usability of text passwords [4]. Usability can be assessed qualitatively or quantitatively. Qualitative methods provide insights on user satisfaction. Quantitative metrics are particularly helpful in assessing the effect of design features on usability. Common quantitative metrics of password usability are efficiency, effectiveness, and memorability, as they are affected by many design features.

Efficiency.
Efficiency is commonly defined by the *entry time* required by a user to complete a login task [4]. Entry time should be low to facilitate efficient authentication, but must be balanced against security requirements. In comparison to text passwords, entry times of many graphical password schemes are considered too long for practical use [20, 6].

Effectiveness.
Effectiveness reflects how well users can perform a specific task with a mechanism. In case of password schemes, legitimate users should be able to authenticate without error. The *success rate* for entering a password correctly without errors is a common metric for effectiveness [15].

Memorability.

Dual coding theory [38] suggests that graphical representations are easier to remember than text, because they induce a visual and verbal code in the user's memory. Studies have indeed shown that graphical passwords are memorable over long intervals [7], but also uncovered the issue of *password interference* when multiple, similar graphical passwords are used [23]. Memorability can be supported by leveraging pre-existing user-specific knowledge rather than requiring users to memorize new or random information [4].

3.2 Design Features

Design of a graphical password scheme is dominated by the used cognitive mechanism, how information is presented, and user interaction. We discuss each aspect in detail.

3.2.1 Cognitive mechanism

Knowledge-based authentication can leverage different cognitive mechanisms. Graphical passwords are commonly categorized into recall, recognition, and cued-recall schemes [4, 15]. *Recall schemes* require users to reproduce a secret from their recollection. Typically, users are either required to draw a shape from memory or repeat a sequence of actions [47]. Examples of recall schemes are Draw-A-Secret (DAS) [29] and Pass-Go [48], as well as Android's Pattern Unlock. In *recognition schemes* users have to recognize a sequence of images or shapes, usually embedded in a grid of decoy images to detract attention of observers [15, 20]. Examples are VIP [15] and Use Your Illusion (UYI) [26]. *Cued-recall schemes* typically require users to select target points in an image or a sequence of images [4, 15]. The image serves as a cue to support memory recall. Ideally, cues are only helpful to the legitimate user but not to observers [4].

The cognitive mechanism impacts a scheme's theoretical password space, observation resistance, and memorability. Recognition and cued-recall passwords are typically easier to remember than recall ones, assuming the same password strength. However, recognition schemes have a small theoretical password space and are prone to shoulder surfing [4].

3.2.2 Information display

Compared to text-based logins, the display of information on a graphical login screen provides more design dimensions. The spatial and temporal arrangement of visual elements, as well as their design allow for diverse variations.

Spatial arrangement.

Almost all graphical password schemes rely on a *grid structure* in order to discretize user input. Recognition schemes usually display a number of images on a grid; recall and cued-recall schemes match the user's input to a grid, which must provide sufficient tolerance to accommodate click variations. A visible grid provides an additional cue for recall. Grid size is influenced by screen size and resolution, as well as by the employed input method (e.g., touch input requires larger tolerance margins than mouse input). Grid size in turn influences the theoretical password space and all usability characteristics. Thorpe & van Oorschot [50] analyzed the effect of different grid sizes on DAS [29]. They show that larger grid sizes increase the password space, but also incur additional recall effort and potentially reduce memorability.

Additionally, visual elements can be *randomized* or displayed at *fixed* positions. De Angeli et al. [15] find that fixed positions for each challenge in a recognition scheme decrease observation resistance but increase effectiveness.

Temporal arrangement.

Graphical password schemes either provide a *single challenge* (draw one shape, select one image), *multiple challenges on a fixed background* (select multiple click points on one image), or *multiple challenge rounds with changing cues* (on each screen select a click point on an image). Thus, similar to a text password, the resulting password is typically a sequence of interactions. The length of the sequence influences the scheme's theoretical password space.

Multi-round graphical password schemes can utilize the next challenge to provide *implicit feedback* about the correctness of the user's previous action [4]. Assuming a fixed sequence of images, an unexpected cue image in the new round indicates that previous input was wrong [11, 10, 41]. Due to its implicit nature, this feedback should only be recognizable and useful to the legitimate user [10].

Variable response design can further enhance observation resistance. Instead of users selecting the same sequence of images at each login attempt in a recognition scheme, only a subset from a *key image portfolio* is used [15]. However, key image portfolios are vulnerable to *intersection attacks* in which an adversary observes multiple login attempts and gains exploitable information about differences in appearance frequencies or preferences of key images and decoy images [19]. Intersection attacks can be hampered by also drawing decoy images from a *decoy image portfolio* [20].

Visual Elements and Cues.

Visual elements and cues are essential components of a graphical password scheme. Biddle et al. [4] recommend that cues should support memorability and that design features should aim to minimize password interference. In terms of designing visual elements one has to consider the entropy, similarity, and familiarity of visuals, as well as potential effects of user choice on the practical password space. Similar to the grid size, the *size of cues* or images and their *level of detail* influence security and usability of recognition and cued-recall schemes. More images per screen provide more options and increase the theoretical password space. More visual elements also enhance shoulder surfing resistance [20], as an adversary's attention is dissipated by the increased amount of superfluous information [31]. For recall schemes, Zakaria et al. [55] propose shoulder surfing defenses (decoy and disappearing strokes) that dissipate an observer's attention while a user is drawing a secret.

Chosen image sets also affect security and usability. High *similarity* of key and decoy images in recognition schemes and cues in cued-recall schemes can adversely affect effectiveness and memorability, as users may have difficulties determining the correct images or click points [15]. Yet, when key images are too different from decoys, predictive models can be used to predict them [53, 51]. Thus, key and decoy image sets must balance memorability and observation resistance. Nameable and distinctive images are generally easier to remember [15]. Using *familiar cues*, e.g., a user's personal photos, has the additional advantage that users are not required to memorize new and more or less random information [4]. When personal images are used, they must be filtered to remove too similar images [20] (e.g., two photos taken only moments apart) and decoy images should be automatically chosen to balance similarity with the user's personal images [21]. With this strategy, key images (personal photos) are well recognizable by the user, yet sufficiently similar to deter an observer. Other approaches to enhance observation resistance include blurring images [26] and displaying only segments of an image [36].

An issue that affects the practical password space is the predictability of *user choice* when creating a password, as users exhibit certain preference biases for image selection in recognition schemes [14] and drawing shapes in recall schemes [52], as well as preferring salient image features (*hot spots*) in click point selection [53, 10]. Letting the system select a password for the user (*system choice*) would eliminate these biases but adversely affect memorability [15]. Instead, *persuasive mechanisms* have been proposed to help users choose less predictable passwords. Dunphy et al. [22] experimented with background images for draw-a-secret schemes to increase shape diversity. For click point-based cued-recall schemes, Chiasson et al. [10] propose viewports highlighting parts of the image to guide click point selection and Bulling et al. [8] mask salient image regions to encourage users to pick less salient points. In a related approach for text passwords, Schechter et al. [45] construct a password popularity oracle to warn users when selecting a too popular password.

Non-visual cues.

Non-visual cues can increase observation resistance by complementing visual elements with additional information not displayed on the screen. One example is users receiving *audio instructions* via earphones whether to lie or answer the current challenge truthfully [43]. *Tactile and audio cues* can also be used to assign meaning to a plain interface [2]. Basically, the challenge is subdivided into a visual and a non-visual part and users have to transform their secret according to the non-visual cue. Non-visual cues enhance observation resistance, but they are restricted by the input and output capabilities provided by a device. The required knowledge transformation can also impair effectiveness [31].

Some schemes also require users to transform their secret according to visual cues. Examples are Convex Hull Click [54], where a user selects items around the recognized key image, and WYSWYE [30], where the user has to derive a pattern from a grid of key and decoy images by eliminating rows and columns without key images. Due to required screen estate, knowledge transformation based on non-visual cues is typically preferred for mobile devices.

3.2.3 Interaction method

The majority of graphical password schemes require users to either make a selection or draw something. Influenced by available input capabilities, most existing graphical password schemes are based on *single pointer* interaction. However, current smartphones support *multitouch* interaction that can also be leveraged in graphical password entry to increase the theoretical password space, e.g., by using multiple fingers for click point selection [41] or drawing [37]. It should also be considered if multitouch interaction requires *one or two hands.* Touch interaction also provides novel attack vectors. Oily residue on the screen facilitates *smudge attacks*, which are a particular threat to draw a secret schemes [1].

Reduced visibility of the interaction can increase observation resistance. In multi-finger interaction, the user's hand may occlude larger portions of the input area [41]. It has also been suggested to use magnetic gestures [42] or the back of the device [18] for authentication, whereby the device occludes the hand. Gaze-based input has been shown to increase observation resistance as it eliminates finger-based interaction completely in favor of using the user's gaze direction [32] or eye gestures [16] for selection tasks.

Depending on a smartphone's available sensors, *biometrics* information can be used to increase resistance. For example, recall-based drawing schemes can be extended to not only authenticate a user based on the drawn shape but also on *how* it has been drawn, based on finger pressure [34, 17] and the effective finger size [17] on a capacitive touchscreen. Other sensors, such as accelerometer, gyroscope, location, and camera, could also potentially enhance authentication.

3.3 Smartphone Capabilities

The available parameters for design features are typically constrained by the capabilities of the specific target platform or smartphone model and the context of use.

3.3.1 Input and output capabilities

A device's input and output capabilities directly relate to the interaction methods it supports. While multitouch screens are common in current smartphones, there may be restrictions in the number of *concurrent touch points* a device can process, e.g., some older smartphones only support 2–4 fingers. *Screen size and resolution* of the device affect spatial arrangement and display of visual elements in terms of how many elements can be displayed and at what detail, which also impacts the theoretical password space. Graphical password schemes must be able to adapt accordingly [12].

Some *advanced interaction methods* that have been proposed to enhance observation resistance, such as gaze detection, require *additional sensors* that may not be available on all devices. However, current smartphones offer a range of sensors that can be effectively combined and utilized to facilitate ideas such as back of device authentication and biometric detection. While speech recognition is also becoming a common place feature, its utility for password entry is not straightforward as voice interaction is prone to observation attacks, e.g., eavesdropping.

In terms of supported output capabilities, current smartphones are all quite similar. They feature a relatively large screen, speakers, an earphone jack, vibration motors, and support a number of communication protocols, such as Bluetooth and Wifi, which could be used to relay output to other devices. For non-visual cues, such as audio instructions, one has to keep in mind that they must only be accessible to the user, i.e., the user must be wearing earphones, which may not be the case at all times.

3.3.2 Context of use

Besides capabilities and constraints of the device, the *context of use* should also be considered in the design of graphical password schemes. Smartphones are often used in public spaces, such as malls, universities, or busses, where the user's interaction with the phone can be easily observed by others [46]. High resolution or "retina" displays not only make screen content more easily readable by the user but also by observers. Thus, the shoulder surfing risk depends on the context of use. Context of use may also constrain the interaction methods users feel comfortable with and are willing to use (e.g., speech input on a crowded bus).

3.3.3 Device-related user constraints

Size and form factor of smartphones further necessitate to consider user constraints in relation to smartphone interaction. While hand and vision constraints are most prevalent at the moment, future interaction methods may require consideration of other user constraints.

Hands and fingers.

Direct touch manipulation results in *lower input accuracy* compared to pointer manipulation with a mouse, because finger tips produce larger pointer blobs and the finger occludes the actual point of interaction. On the other hand, *occlusion* by the user's hand and fingers may reduce visibility for shoulder surfers and enhance observation resistance.

While multitouch interaction can increase the theoretical password space, the effect is constrained to anatomically possible *finger and hand postures*, potentially resulting in a reduced practical password space. Similarly, in most situations not all fingers can be used for input, as the user must also hold the device. Implications by the user's preference for one- or two-handed operation should also be considered.

Vision.

Vision impairments and *color blindness* should be considered in the design of visual elements and cues. Reflections of the screen in the user's eyes or glasses could also potentially be exploited for shoulder surfing attacks, as has been shown for text passwords on mobile devices [39].

4. THE DESIGN SPACE IN PRACTICE

The discussion of the design space and its graphical representation (see. Figure 1) highlight the complexity of interrelations between different aspects of graphical password design. Almost all aspects influence or are influenced by more than one other aspect. Nevertheless, certain general properties emerge. Smartphone capabilities mainly impact design features, which in turn impact the password characteristics of a given scheme. In practice, the smartphone capabilities are typically determined by a specific target platform or device. What remains variable are the design features and how they are utilized by a specific graphical password scheme. In order to better understand how design features are impacted by smartphone capabilities, we implemented five different graphical password schemes that cover various design features. Hereby, we opted for already existing schemes, which we re-implemented on the same smartphone platform in order to gain comparable insights about the effects of smartphone constraints and capabilities on the instantiation of design features. Subsequently, we performed a comparative user study in order to assess how the combination of fixed smartphone capabilities and varying design features impact security and usability characteristics of the different schemes. The results are a number of insights and guidelines to aid navigation of the design space.

4.1 Target Platform

The target platform for our implementations was Android with API level 8 (Android 2.2.x). We used a Samsung Galaxy Nexus (GT-I9250) for assessing device-specific capabilities. The device has a 4.65 in Super AMOLED screen with a resolution of 1.280 × 720 pixels. The capacitive touchscreen supports multitouch with up to ten fingers. The smartphone has a range of sensors, such as accelerometer, gyroscope, barometer, digital compass, proximity sensor, dual microphones, as well as back and front cameras.

4.2 Implemented Password Schemes

In order to compare the effects of smartphone capabilities on different design features, we selected schemes covering all three categories of cognitive mechanisms and a wide range of design features. As a recall scheme, we chose the basic version of *Pass-Go* [48]. *Pass-Go* is a draw a secret scheme, which is similar to, but slightly more complex than Android's pattern lock. We chose *Use Your Illusion (UYI)* [26] as a representative recognition scheme, because it explicitly aims to enhance observation resistance by addressing shoulder surfing threats. *TAPI*[1] [13] combines elements from recognition (recognize image on grid) and cued-recall schemes (select partitions of images), which makes it an interesting hybrid approach to consider. *Cued Click Points (CCP)* [11, 10] is a canonical example of a cued-recall scheme in which users select points on a sequence of images. We further considered *MIBA*[2] [41], a cued-recall scheme leveraging multitouch input. Thus, the chosen schemes cover a large variety of information display methods and different types of touch interaction. We omitted non-visual cues and more advanced interaction methods in order to retain a managable number of schemes by focusing on commonly available interaction and information display methods. A shared focus on touch interaction and visual information display of all schemes ensures comparability of implementation aspects, as well as experimental results.

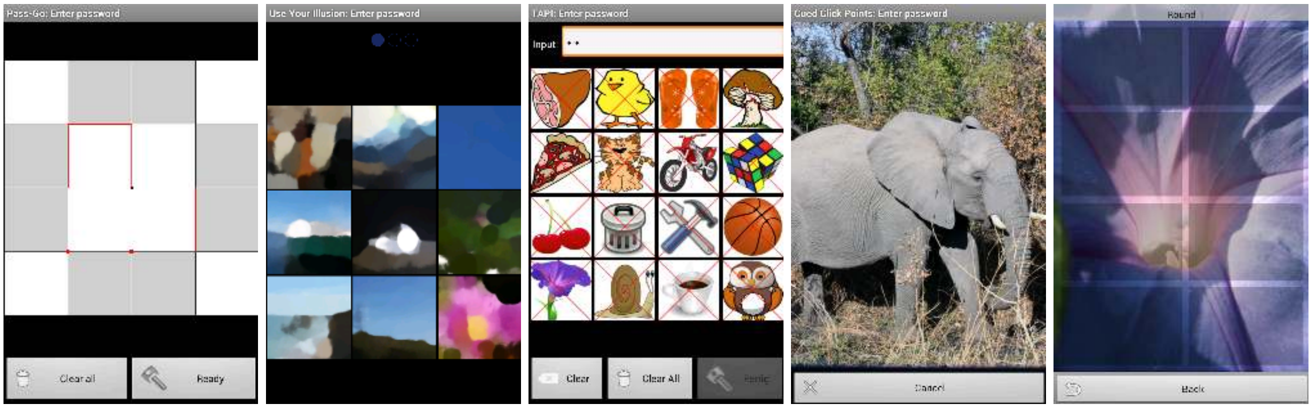
For each scheme, we implemented Android activities for password enrollment and login according to the descriptions in the respective original papers. We defined a common password enrollment process implemented for all five schemes. First, the system prompts the user to create a password of a specific strength, before the user can start to enter the new password. The enrollment process ends with a "remember password" dialog, which shows the created password again to support memorization. While a text password is shown in plaintext, we used scheme-specific graphical metaphors to visualize created graphical passwords, e.g., highlighting selected areas on an image. The login activity challenges the user to input the password. Some of the implemented schemes recognize the end of the password on their own, others provide a finish button, according to each scheme's proposal. After password entry has been completed, a dialog shows if authentication was successful or not.

In the following, we discuss and categorize each scheme in relation to the design space. We particularly focus on design features (cognitive mechanisms, information display, interaction method), which are impacted by smartphone constraints. Aspects that are not directly influenced by the smartphone, such as choice of the image set, are of lesser interest here. See Appendix A for a table summarizing the mappings between schemes and design space.

4.2.1 Pass-Go

Tao & Adams [48] propose Pass-Go as a recall scheme focused on intersection points of a grid. The user can either

[1] TAPI = Touchscreen Auth. using Partitioned Images
[2] MIBA = Multitouch Image Based Authentication



(a) Pass-Go (b) UYI (c) TAPI (d) CCP (e) MIBA

Figure 2: Login activities of the implemented graphical password schemes.

draw dots on intersection points of the grid or connect intersection points with strokes, as shown in Figure 2a. Points and lines have to be drawn in the right order to authenticate successfully. *Pass-Go* also employs simple visual cues by augmenting the grid with a fixed background pattern reminiscent of a Go board to assist users in locating the correct intersection points. Tao & Adams further suggest Color Pass-Go to increase the theoretical password space by allowing users to choose different colors for their dots and strokes. An additional button provides the option to hide drawing indicators on the screen. Sensitive areas around intersection points are used to make user input snap to intersection points. The original Pass-Go implementation uses a 9×9 grid and sensitive areas with radius $\frac{d}{4}$, where d is the width of a grid cell.

Pass-Go has been optimized for mouse input in the web context. In order to adequately support touch input on the smaller smartphone display, we reduced the grid resolution to 5×5 and increased the radius of sensitive areas to $\frac{d}{3}$. In order to gain more screen estate for the grid, we removed the “hide drawing indicators” button and the color selectors, as only 3% of participants in Tao & Adams’s deployment study made use of this option. While 30% voluntarily used multiple colors in their study, the potential increase of the password space must be traded off against usability on the smaller screen. As a result, our implementation is also closer to Android’s PatternLock, which is likely based on Pass-Go.

4.2.2 UYI

UYI [26] is a recognition-based scheme. The login screen displays 9 images randomly positioned in a 3×3 grid. The user must recognize and select a key image amongst decoy images. In the original proposal, a UYI password consists of three challenges [26], i.e., 3 key images must be recognized on 3 screens. An indicator on top shows the current position in the sequence. The login process automatically ends after the last challenge. As there is no option to correct input, the user has to always complete the whole login process, even after noticing a mistake. Key and decoy images are photos. However, during login the user sees only distorted versions of them, as shown in Figure 2b. The idea is that a legitimate user, who has seen the original key images at time of enrollment, can recognize their blurred and abstracted versions, while memorizing the distorted versions is more difficult for an observer. Thus, level of detail is reduced to improve shoulder surfing resistance.

While Hayashi et al.’s implementation was optimized for a mobile device without touchscreen, the adaptation to single touch input was straightforward requiring no changes to the scheme. In the login process, UYI does not provide visual feedback when a picture is selected except for changing to the next challenge. Our implementation also gives haptic feedback on image selection with a short vibration. Hayashi et al. do not describe the layout of their enrollment view. Our prototype allows the user to horizontally browse through a gallery of 27 photographs. Selected images and their distorted versions are shown below. Images can be deselected again if they do not appeal to the user.

4.2.3 TAPI

TAPI [13] is a hybrid scheme that could be considered recognition-based and cued-recall. The login screen displays a 4×4 grid of icons. Each icon is divided into four parts to enhance shoulder surfing resistance, as shown in Figure 2c. The user has to recognize the key icon and select the correct part (cued-recall). A password consists of a sequence of challenges, whereby icons do not change their position.

Citty & Hutchings [13] modeled the layout of TAPI after a typical PIN pad. Hence, the input bar on the top, which indicates how many image parts have been selected already. Buttons allow to correct the last entry, clear all, and submit the complete password. No further adaptations to the scheme were required, because the original TAPI implementation was also based on Android and touch input.

4.2.4 CCP

CCP [11] is a canonical example of a cued-recall scheme. The CCP login screen, shown in Figure 2d, displays an image on which the user can freely select a point (click point). A CCP password consists of a sequence of click points on different images (five in the original proposal), with one click point per image. The position of a click point determines the subsequent image to provide implicit feedback about wrongly selected points (an unexpected image is shown). The login process ends automatically when the number of click points matches the password length or can be aborted by the user. CCP theoretically requires a large picture portfolio to provide a unique picture for every possible click-point [11]. Our implementation uses a portfolio containing only 34 images. In the enrollment phase, if the image corresponding to a click point position has already been used, linear probing is used to find an unused one. This ensures that passwords consist of a sequence of unique images but does not prevent potentially showing a used image after a "wrong" click point. User choice has a large impact on the practical password space of CCP due to a preference for image hot spots. Chiasson et al. suggest Persuasive CCP (PCCP) [10] as an improvement. For enrollment, users can only set a click point within an randomly positioned viewport, which results in less predictable passwords.

In CCP, mouse clicks on the image are mapped with centered discretization [9] onto an invisible 24×17 grid of tolerance squares (each 19×19 px), whereby each value has a two dimensional offset determined at enrollment. In order to adapt this for touch input, we had to enlarge the size of tolerance squares to 98×98 px, resulting in a 8×10 grid and a reduction of the theoretical password space.

4.2.5 MIBA

MIBA [41] is a cued-recall scheme similar to CCP. It has been developed for smartphones and uses multitouch input to enable simultaneous selection of up to 4 click points. The login process is based on CCP, i.e., the image sequence is determined by the position of click points and provides implicit feedback. Errors can be corrected with a back button. As shown in Figure 2e, images are overlaid with a 2×4 grid of semitransparent rectangles to better support users in selecting click points with multiple fingers. Input is accepted when fingers are lifted off the screen, which allows to reposition fingers without causing input events. Ritter et al. determined grid layout and number of fingers as a trade-off between a larger theoretical password space (more grid cells, more fingers) and anatomically possible hand postures. MIBA uses a shift-function [28] to increase the theoretical password space. A long touch activates a shift round, which should be imperceptible to an observer but indicated to the user with a short vibration. The login process ends when the correct password is entered. An infinite sequence of images is shown for wrong inputs.

4.3 Effects on Password Characteristics

The implementation of these schemes on one smartphone platform showed that the smaller screen size and the move from mouse to touch input requires a reduction in grid size, as larger targets are required. By reducing the number of additional buttons and interaction elements, we were able to increase screen estate for the actual password input. The minimal interface and relatively large grid size of CCP stand out in this regard. TAPI's grid is also quite large with 8×8 image partitions. With the exception of Pass-Go, all schemes rely on multiple challenges. Three of them employ changing visual cues, while TAPI shows a fixed set of icons. MIBA leverages multitouch to increase the theoretical password space but the interaction with multiple fingers is restricted by feasible hand postures.

4.3.1 Theoretical password space

In order to compare the theoretical password spaces of the implemented schemes, we considered the password length required by each scheme to achieve 14 bit and 42 bit strength, under the assumption of equiprobable password distribution.

Table 1: Password lengths for same password strength

Scheme	**14 bit password**		**42 bit password**	
	rounds	# clicks	rounds	# clicks
Pass-Go	1	length 2	1	length 6
UYI	3	3	10	10
TAPI	3	3	7	7
CCP	3	3	9	9
MIBA	1	1-2	3	3-6
PIN	1	4	1	13

14 bit strength corresponds to the security of a four-digit PIN, which is commonly used to unlock phones. 42 bit are equivalent to the strength of a seven-character text password consisting of numbers, lowercase, and uppercase characters. Table 1 provides an overview of the total number of clicks and rounds required in each scheme to achieve the respective strength. For Pass-Go, we report the required stroke length rather than clicks. For comparison, we also list the required effort for a PIN of the same strength. Note that we focus on theoretical password strength to simplify comparison. The practical password space of schemes may be smaller due to effects of user choice on password distribution, in which case longer passwords than reported in Table 1 are required to achieve 14 bit and 42 bit strength.

Pass-Go and MIBA require the shortest password length to reach 14 bit and 42 bit. Due to the potential combination of four fingers and the shift function, MIBA requires only 3 challenges for 42 bit, compared to 7 and 9 rounds for TAPI and CCP. All implemented graphical password schemes require less clicks than entering a comparable PIN.

4.3.2 User study

We performed a between-subjects lab study to compare observation resistance and usability aspects of the implemented schemes, as well as PIN-entry as a baseline. In our study, we focused on the effects of design features and smartphone capabilities on shoulder surfing resistance, efficiency, and effectiveness. Independent variables in our study were *password scheme* and *password strength* (14 and 42 bit).

Recruitment and Demographics.

We recruited 60 participants (14 female, 46 male, aged 18–32 years) from our campus population. The majority of participants had a computer science background. We assigned each participant to one password scheme, evenly distributed across six groups. Each group consisted of 2 women and 8 men with the exception of the Pass-Go group (4 female, 6 male). Slightly more than half of the participants owned a smartphone with touchscreen; 4–6 smartphone owners were in each group. Of the 32 smartphone owners, 19% used PIN and 31% used the Android Unlock Pattern to protect their personal phone. There was no PIN user in the PIN group. Unlock Pattern users where distributed evenly across graphical schemes (1–2 per group) and 4 of them in the PIN group. Participants in the PIN group were explicitly instructed not to use any of their real-world PINs (e.g., ATM).

Procedure.

The study was performed at our university lab. During sessions, only one participant and the experimenter were present. At the beginning of each session, participants were assigned to a group, signed a consent form, and provided demographic information. Afterwards, they were introduced to the assigned scheme, including what constitutes a password, description of the enroll and login process, as well as explanation of any scheme-specific features, e.g., in case of MIBA, multi-finger use, correction key, shift function, and that input determines the next image. Participants were also briefed that they had to solve a mental rotation task between the actual tasks. Participants were instructed to wear their glasses, if necessary. The usability and shoulder surfing experiments are described in detail in Sections 5 and 6, together with their results. All participants used the same smartphone (see Sec. 4.1) to complete the experiments. For the study, we extended the implemented password schemes in such a way that they stored entered passwords and user inputs enriched with timestamps, as well as the result of a login challenge with the required duration and number of clicks or strokes. Each session lasted 20–35 minutes. Participants received chocolate at the end of the session.

5. USABILITY EXPERIMENT

Our usability experiment pertains to initial use, as all participants were novices in the use of their assigned password scheme, with the obvious exception of the PIN group.

The experiment started with a tutorial phase, in which participants had to enter a provided 14 bit password followed by a provided 42 bit password on the smartphone. The application continued to challenge the participants until both passwords had been entered successfully. Next, participants had to create their own 14 bit password. Once enrolled, the participant had to enter the created password five times. Similar to Chiasson et al. [11], we shortly distracted participants with mental rotation tasks (MRTs) between login attempts to clear their working memory. Then, participants had to repeat the same process with a 42 bit password (enrollment, five attempts with MRTs). The usability experiment ended with Lewis' Post-Study System Usability Questionnaire (PSSUQ) [33].

Thus, we assessed usability with a combination of quantitative and qualitative metrics. A scheme's efficiency is measured by the *entry time* required for a login, effectiveness is assessed with the login *success rate*. The usability questionnaire provides qualitative data on *user satisfaction* based on participant ratings on 7-point Likert scales. Due to the variations in design features, we expected significant differences between schemes for these metrics.

5.1 Entry time

Figure 3 shows the entry time distributions of the different schemes over all successful login attempts for 14 bit and 42 bit passwords. As we balanced schemes in their password strength, differences in entry time indicate which schemes require less effort for authentication. Non-parametric Kruskal-Wallis tests show significant differences between groups for entry time of, both, 14 bit passwords ($H(5)=22.77$, $p<.01$) and 42 bit passwords ($H(5)=32.03$, $p<.01$). We performed post-hoc analysis with Games-Howell tests.

UYI users needed the longest for password entry. UYI median entry time for 14 bit passwords was 8.7s, which is significantly slower than Pass-Go ($p<.01$, $r=.85$), PIN ($p<.01$, $r=.85$), and MIBA ($p=.01$, $r=.71$) with strong effects. For the 42 bit password, UYI (Mdn=21.7s) was also significantly slower than Pass-Go ($p<.01$, $r=.85$) and MIBA ($p<.01$, $r=.85$). Apparently, UYI's distortion and randomized placement of key images hampered the participants' ability to quickly recognize and locate them. Pass-Go and MIBA perform significantly better with their fixed grid and button positions.

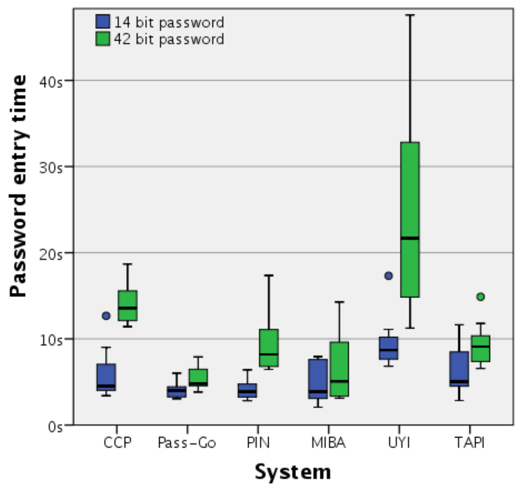


Figure 3: Password entry time for 14 and 42 bit passwords.

Table 2: Successful login rates

Scheme	14 bit password		42 bit password	
	mean	std. dev.	mean	std. dev.
PIN	1.00	.000	.98	.063
Pass-Go	1.00	.000	.82	.346
TAPI	.92	.140	.94	.097
UYI	.90	.105	.71	.389
MIBA	.82	.199	.68	.434
CCP	.66	.378	.42	.416

For the 42 bit password, entry time of CCP was significantly slower (Mdn=13.6s) than Pass-Go ($p<.01$, $r=.82$), MIBA ($p=.02$, $r=.62$), and TAPI ($p=.05$, $r=.68$) with large effects. Pass-Go was also significantly faster than TAPI for 42 bit password entry ($p<.01$, $r=.71$), with results exhibiting a strong effect as well. These results suggest that visible grids provide an advantage when entering stronger passwords, and that grids with fewer elements are more usable.

In general, Pass-Go was most efficient for both password strengths (Mdn_{14}=4.0s, Mdn_{42}=4.8s). MIBA performed also well (Mdn_{14}=3.9s, Mdn_{42}=5.1s), but PIN entry showed comparable performance for 14 bit passwords (Mdn_{14}=3.9s).

5.2 Success rate

The success rate is the average of successful logins over all attempts of one participant. Table 2 shows the mean values per group for both password strengths. Kruskal-Wallis tests indicate significant differences in success rate for short passwords ($H(5)=20.31$, $p<.01$) and long ones ($H(5)=16.92$, $p<.01$). However, post-hoc analysis for the 14 bit password revealed no significant pairwise differences. For the 42 bit password, Games-Howell post-hoc analysis shows that the success rate of CCP is significantly below the one of PIN ($p=.02$, $r=.78$) and TAPI ($p=.03$, $r=.69$).

It is not surprising that the PIN group has the highest success rate considering the familiarity of entering PINs. Of the graphical password schemes, Pass-Go, TAPI, and UYI

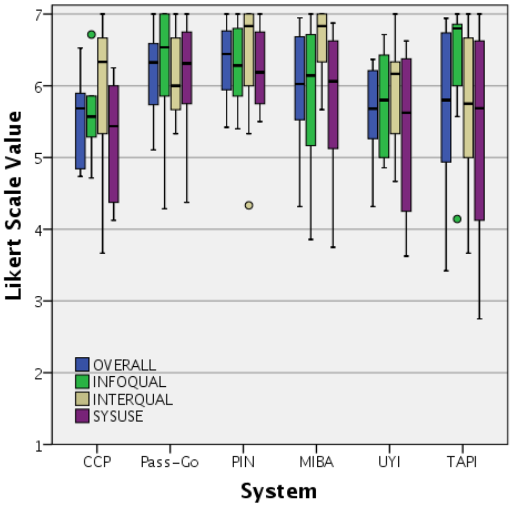


Figure 4: User satisfaction results (PSSUQ).

showed over 90% successful logins for short passwords; for long passwords only TAPI. Overall, TAPI was the most effective, albeit not the most efficient scheme in our study and the only scheme that achieved higher success rates for longer passwords than short ones. Pass-Go was the most sufficient and is also highly effective, but our results suggest that effectiveness decreases for longer drawings. The results for UYI suggest that three distorted images can be recognized well, which matches the UYI configuration proposed by [26], but that UYI is less suitable for stronger passwords. The success rates of MIBA and CCP may suggest that implicit feedback through the sequence of images is less effective than expected in helping users recognize and correct mistakes.

5.3 User Satisfaction

The PSSUQ consists of 19 items, which result in four scales measuring user satisfaction [33]: perceived usefulness of the scheme in completing the given tasks (*SysUse*), perceived quality of displayed information (*InfoQual*) and interface elements (*InterQual*), as well as overall satisfaction with the scheme (*Overall*). Figure 4 shows the results for each scheme. An ANOVA found no significant differences between schemes for any of the scores, as all median scores were in the range of 5.44–6.8 on a 7-point Likert scale (7=*best*). Thus, our participants perceived all tested schemes as usable, despite the significant differences in terms of entry time and success rate. Note, however, that the results in Figure 4 reflect non-comparative perceptions, as each participant interacted only with one scheme. In a within-subjects study, results would likely show stronger differences, but could also suffer from training and preference biases.

6. SHOULDER SURFING EXPERIMENTS

In the shoulder surfing experiments, we also assessed passwords with 14 and 42 bit strength. Participants acted as shoulder surfers and the experimenter played the victim. This setup has the advantage that the experimenter could train password entry beforehand, to ensure consistent entry speed and body posture for all participants. Thus, we evaluated a casual observer's ability to recognize a password entered by a trained user, compared to evaluating an expert shoulder surfer with a novice user in the reverse setup.

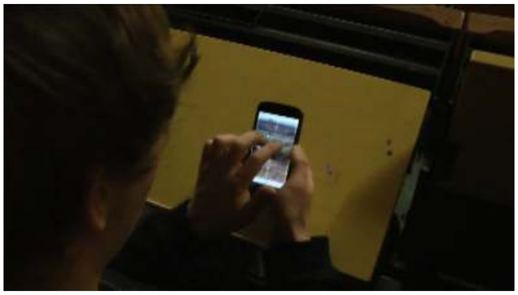


Figure 5: Video of a login attempt with the MIBA scheme.

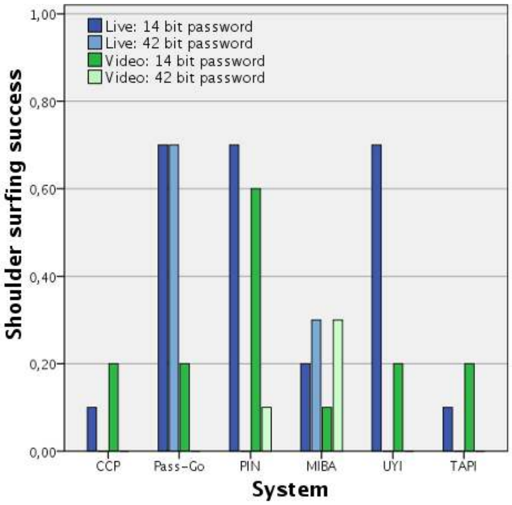


Figure 6: Shoulder surfing success rate (live and video).

The shoulder surfing experiments were performed after the usability experiment. The right-handed experimenter sat at a table holding the smartphone as if entering a password and the participant could position herself left, right, or behind the experimenter. The participant also received paper and pencil to take notes, similar to Tari et al.'s study [49]. When the participant stated to be ready, the experimenter entered the trained password once. Subsequently, the smartphone was passed to the participant, who had three attempts to enter the password. This procedure was first carried out with a 14 bit password and then with a 42 bit one.

Afterwards, a similar experiment with the same procedure followed in which the participant had to recognize 14 and 42 bit password input recorded on video. The idea was to study if video and live shoulder surfing produce comparable results. The person in the video entered the password with the left hand and the camera was positioned over the person's right shoulder as shown in Figure 5—a scene that favors the shoulder surfer and could realistically occur, e.g., in a lecture theater.

6.1 Shoulder Surfing Success Rate

Due to the differences in how passwords are entered in the different schemes, definition of a fair and comparable distance metric between entered and correct password for all schemes is quite difficult and somewhat arbitrary. Thus, we opted for a robust binary metric to measure shoulder surfing success, which is 1 if the participant entered the correct password within three attempts and 0 otherwise. Figure 6 summarizes the normalized shoulder surfing success results for 14 and 42 bit passwords and live and video observation.

We look at the live shoulder surfing results first. Kruskal-Wallis tests indicated significant differences for 14 bit passwords ($H(5)$=19.76, p<.01) and 42 bit ones ($H(5)$=29.26, p<.01). We performed post-hoc analysis with Games-Howell tests. For the 14 bit password, participants had the least shoulder surfing success for CCP and TAPI. Both schemes are significantly more resistant to shoulder surfing than Pass-Go, PIN and UYI (all: p=.05, r=.60), which were most susceptible to shoulder surfing for short passwords. For the 42 bit password, Pass-Go was significantly more susceptible to shoulder surfing than CCP, PIN, UYI and TAPI (all: p=.01, r=.72), which were not guessed correctly at all. The live shoulder surfing results show that Pass-Go passwords, even 42 bit ones, are easy to shoulder surf, which suggests that drawn patterns are easier to observe than click selections. CCP and TAPI were shoulder surfing resistant even for 14 bit passwords. The reason might be, that the exact click point is always covered by the clicking finger.

The results for video shoulder surfing exhibit no significant differences. However, the success rate of video observations are lower for almost all schemes than the respective live results, with a few exceptions deviating by only 1–2 observations. A reason might be that fine details of Pass-Go patterns and exact click points are harder to perceive on video. With 42 bit passwords, only MIBA and PIN had any successful guesses for the video observation. Likely due to their larger buttons compared to the other schemes, as well as the long touch required to activate MIBA's shift mode.

7. DISCUSSION

The results of our experiments show that almost all tested graphical password schemes are comparable to PIN entry in terms of usability, but are more resistant against shoulder surfing, at least for 42 bit passwords. The question we are addressing now is how the obvious differences in design features of these schemes impact password characteristics and what general guidelines can be derived from the gained insights in order to make the rather complex design space of Figure 1 more accessible and manageable.

7.1 Gained Insights and Guidelines

Cognitive mechanism: The results from our lab study show that Pass-Go, the only recall scheme, is highly effective and efficient, but also significantly easier to shoulder surf than the other schemes. A potential explanation is that drawing patterns on the smaller, smartphone-optimized grid are easy to remember for legitimate users as well as observers. Analyzed cued-recall and recognition schemes also showed good usability, but are also more resistant against observation attacks. These results suggest that **(1) cued-recall and recognition schemes strike a better balance between security and usability and are preferable over drawing-based recall schemes on smartphones.** Thus, the quite popular drawing-based Android Pattern Unlock is suboptimal in terms of security.

Spatial arrangement: The usability results show that smaller grids with fewer and larger items cause fewer login errors. In general, touch input requires larger grid items than mouse input. However, the cued-recall schemes CCP and TAPI have relatively small grid items and exhibit reasonable password entry times, due to more elements per screen (larger theoretical password space). Furthermore, CCP and TAPI turned out to be most resistant to shoulder surfing by casual observers in our study, likely because the user's finger occludes the pressed item on the touchscreen. Therefore, **(2) grid items or touch targets should be as small as still usable to enhance security** by increasing the theoretical password space and observation resistance. The results of UYI show that **(3) randomized positioning of items is not preferable in terms of usability or security**, because users need more time to locate their key image, which also gives observers time to do the same.

Temporal arrangement: The entry times observed in our study are conservative, as entry time would likely decrease with frequent use of a scheme. Nevertheless, UYI exhibits the worst entry time for 42 bit passwords due to the high number of challenges (10) required to reach this password strength. Adjustment of UYI's grid size could likely improve entry time. Thus, **(4) spatial and temporal arrangement must be balanced to maintain usability for stronger passwords**. Our study results provide **(5) no evidence that changing cues are preferable over fixed backgrounds in terms of security or usability**. Indeed, TAPI with its fixed background is highly effective and also hard to shoulder surf by casual observers for longer passwords. MIBA and CCP both support implicit feedback, but our results indicate no particularly positive effect on the login success rate. Further studies are required to determine respective reasons. None of the analyzed schemes employed variable response design (only subset of key image portfolio). While variable response design would likely improve observation resistance, it might also reduce entry time, similar to the randomized image placement of UYI. However, our results do show that **(6) longer passwords increase observation resistance**, because they overwhelm the working memory of casual observers. Further experiments are required to determine if that also holds true for trained observers. Furthermore, video shoulder surfing seems less practical than expected, due to the required shooting angle and lighting conditions. While these issues can be partially addressed with image processing and automated video analysis, the required capabilities are likely not associated with the adversary model of a casual observer.

Visual Elements and cues: Our results indicate that **(7) visible grids improve usability (entry time) for stronger passwords**, as Pass-Go, MIBA and TAPI were significantly faster than CCP for 42 bit passwords. UYI exhibited the slowest password entry times, which could either be caused by the randomization or distortion of images. Yet, our results do not show any differences in performance that could be directly attributed to the type of visual cue (images, icons, or distorted images). As mentioned above, smaller items seem to improve shoulder surfing resistance.

Non-visual cues: As none of the analyzed schemes employed advanced non-visual cues, we cannot comment on their effectiveness. However, activation of MIBA's shift function was relatively easy to recognize by shoulder surfers, due to the longer time (1s) the user's fingers remained on the display. Shortening the threshold for shift activation could improve shoulder surfing resistance for MIBA.

Interaction method: Quantitative and qualitative results show that drawing and single touch are very usable.

However, draw interaction is more susceptible to shoulder surfing. Multitouch interaction increases the theoretical password space and therefore reduces the number of challenges required for stronger passwords. Yet, some participants reported initial difficulties with MIBA's multitouch interaction. Combination of simple interaction methods with additional authentication factors, e.g., biometrics, could be a viable alternative to complex multi finger interaction.

7.2 Limitations

Our analysis of the design space and representative graphical password schemes was mainly focused on design and interaction aspects and their implications for usability and security. However, implementations of graphical password schemes must also consider additional aspects, such as secure password storage and offline attacks, reliable data models for graphical passwords, or online verification requirements. Bonneau et al.'s extensive list of requirements for authentication schemes [6] is a helpful resource in that regard. Our selection of password schemes for practical evaluation covered most but not all design features of the design space. Non-visual cues and advanced interaction methods, such as biometrics, were not considered in favor of a manageable number of schemes, which were well comparable due to a shared focus on visual information display and touch interaction. It would be worthwhile to further explore the design space in future studies that look more specifically at non-visual cues and advanced interaction methods in order to extend the set of guidelines.

Concerning our usability and shoulder surfing study, it should be noted that our results are not representative of the general population, due to the fact that the majority of participants were males in their early twenties with a computer science background. Thus, our usability results are likely better than for the general population, due to above average technology affinity. Yet, the obtained results were still suitable for our goal of differentiating effects of design features on password characteristics by comparing different schemes.

In our setup, participants acted as shoulder surfers, thus, our results and insights pertain to casual observers. Further experiments are required to compare our results with the performance of trained observers, who may achieve higher success rates. For interpretation of our study results, we assumed equiprobable password distributions and refrained from analyzing the effects of user chosen passwords on observation resistance, because our sample size (10 users per scheme) was too small to obtain a meaningful estimate of the practical password space. While practical password space analysis has been performed for some of the schemes (e.g., hot spot analysis of CCP [10]) comparative experiments with larger sample sizes could provide further insights regarding the influence of different design features and smartphone constraints on user choice and practical password space.

Our efficiency and effectiveness results are conservative as they reflect initial use. Further long-term and deployment studies could provide additional insights on training effects. Of further interest would be the analysis of effects of design features and smartphone capabilities on password memorability in different schemes, as well as on password interference with multiple user-chosen passwords. Respective experiments comparing multiple schemes are needed to further refine the insights and guidelines discussed above.

8. CONCLUSIONS

We provided multiple contributions in this paper. We analyzed and described the design space for graphical passwords on smartphones in detail. Our detailed discussion of password characteristics, design features, smartphone capabilities, and their interrelations provides a comprehensive picture of the available means and constraints when designing graphical passwords for smartphones. We implemented five existing graphical password schemes on the same smartphone platform in order to demonstrate that the proposed design space is expressive enough to capture all aspects of a graphical password scheme and meaningful enough to guide and support design and development of such schemes. The discussion of these schemes and their implementations shows how smartphone capabilities actually impact different design features, as well as the diversity of graphical password schemes. In our user study, we further assessed these schemes in usability and shoulder surfing experiments to validate identified relations in the design space and gain further insights on which design parameters are suitable to effectively balance security and usability characteristics of graphical passwords. Our discussion of results leads to a number of helpful guidelines for the design of graphical passwords. Further comparative studies could provide additional insights to refine and complement those guidelines, e.g., by studying effects on memorability, analyzing practical password space, and the effects of more advanced interaction methods. The outlined design space supports design choices by highlighting anticipatable effects of different design decisions on usability and shoulder surfing resistance.

While graphical passwords are unlikely to fully replace text passwords on smartphones in the near future [6], promising applications are convenient and secure device access, as well as protection of password managers for text passwords, which has also been suggested in the browser context [3]. Thus, in practical use, graphical passwords may obviate the need to type text passwords on smartphones even without fully replacing them.

An interesting observation in our study was that 50% of smartphone owners did not use any authentication method on their personal device. One explanation could be a lack of awareness about the risks associated with sensitive information on smartphones. Another reason could be that existing authentication methods are too burdensome for regular use, highlighting the importance of research on authentication methods that are secure, easy to use, and less burdensome than PIN or password entry. Our analysis of the design space for graphical passwords shows that there are vast opportunities for improvement and innovation. Especially the utilization of multitouch, non-verbal cues, and authentication with multiple factors is still under explored. However, the outlined design space also highlights the importance of balancing multiple design aspects rather than focusing on singular aspects, such as shoulder surfing resistance.

9. ACKNOWLEDGMENTS

The authors would like to thank all study participants, our shepherd Robert Biddle, as well as Daniel Ritter for providing the original MIBA source code. Background images used for MIBA and CCP are courtesy of Sito Wijngaarden. Images used for UYI are courtesy of Christoph Straub.

10. REFERENCES

[1] A. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. In *Workshop on Offensive Technologies (WOOT '10)*. USENIX Assoc., 2010.

[2] A. Bianchi, I. Oakley, V. Kostakos, and D. S. Kwon. The phone lock: audio and haptic shoulder-surfing resistant PIN entry methods for mobile devices. In *Proc. Conf. on Tangible, embedded, and embodied interaction (TEI '11)*. ACM, 2011.

[3] K. Bicakci, M. Yuceel, B. Erdeniz, H. Gurbaslar, and N. Atalay. Graphical passwords as browser extension: Implementation and usability study. In *Third IFIP WG 11.11 International Conference on Trust Management (IFIPTM '09)*. Springer, 2009.

[4] R. Biddle, S. Chiasson, and P. Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys*, 44(4):1–41, 2012.

[5] J. Bonneau. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *2012 IEEE Symposium on Security and Privacy*, number Section VII, pages 538–552. IEEE, May 2012.

[6] J. Bonneau, C. Herley, P. C. V. Oorschot, and F. Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *IEEE Symp. on Security and Privacy*. IEEE, 2012.

[7] S. Brostoff and M. A. Sasse. Are Passfaces more usable than passwords? a field trial investigation. In *Proc. BCS-HCI '00*. Springer, 2000.

[8] A. Bulling, F. Alt, and A. Schmidt. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In *Proc. CHI '12*. ACM, 2012.

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

[10] S. Chiasson, E. Stobert, A. Forget, R. Biddle, and P. C. van Oorschot. Persuasive Cued Click-Points: Design, implementation, and Evaluation of a Knowledge-Based Authentication Mechanism. *IEEE Trans. Depend. and Secure Comp.*, 9(2):222–235, 2011.

[11] S. Chiasson, P. van Oorschot, and R. Biddle. Graphical Password Authentication Using Cued Click Points. In *Proc. ESORICS '07*. Springer, 2007.

[12] U. Cil and K. Bicakci. gridwordx: Design, implementation, and usability evaluation of an authentication scheme supporting both desktops and mobile devices. In *Workshop on Mobile Security Technologies (MoST '13)*, 2013.

[13] J. Citty and D. R. Hutchings. TAPI: touch-screen authentication using partitioned images. Tech. Report 2010-1, Elon University, 2010.

[14] D. Davis, F. Monrose, and M. K. Reiter. On User Choice in Graphical Password Schemes. In *Proc. USENIX Security Symposium*. USENIX Assoc., 2004.

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

[16] A. De Luca, M. Denzel, and H. Hussmann. Look into my eyes!: can you guess my password? In *Proc. SOUPS '09*. ACM, 2009.

[17] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch me once and I know it's you! In *Proc. CHI '12*. ACM, 2012.

[18] A. De Luca, E. von Zezschwitz, N. D. H. Nguyen, M. Maurer, E. Rubegni, M. P. Scipioni, and M. Langheinrich. Back-of-Device Authentication on Smartphones. In *Proc. CHI '13*. ACM, 2013.

[19] R. Dhamija and A. Perrig. Déjà Vu: a user study using images for authentication. In *Proc. USENIX Security Symposium*. USENIX Association, 2000.

[20] P. Dunphy, A. P. Heiner, and N. Asokan. A closer look at recognition-based graphical passwords on mobile devices. In *Proc. SOUPS '10*. ACM, 2010.

[21] P. Dunphy and P. Olivier. On automated image choice for secure and usable graphical passwords. In *Proc. Annual Comp. Security Applications Conf. (ACSAC '12)*. ACM, 2012.

[22] P. Dunphy and J. Yan. Do background images improve "draw a secret" graphical passwords? In *Proc. CCS '07*. ACM, 2007.

[23] K. M. Everitt, T. Bragin, J. Fogarty, and T. Kohno. A comprehensive study of frequency, interference, and training of multiple graphical passwords. In *Proc. CHI '09*. ACM, 2009.

[24] A. G. Goldstein and J. E. Chance. Visual recognition memory for complex configurations. *Perception Psychophysics*, 9(2):237–241, 1970.

[25] K. Golofit. Click Passwords Under Investigation. In *Proc. ESORICS '07*. Springer, 2007.

[26] E. Hayashi, R. Dhamija, N. Christin, and A. Perrig. Use Your Illusion: secure authentication usable anywhere. In *Proc. SOUPS '08*. ACM, 2008.

[27] C. Herley and P. van Oorschot. A Research Agenda Acknowledging the Persistence of Passwords. *IEEE Security & Privacy Magazine*, 10(1):28–36, 2012.

[28] W. Jansen, S. Gavrila, V. Korolev, R. Ayers, and R. Swanstrom. Picture password: A visual login technique for mobile devices. Tech. report NISTIR 7030, NIST, 2003.

[29] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin. The design and analysis of graphical passwords. In *Proc. USENIX Security Symposium*. USENIX Assoc., 1999.

[30] R. A. Khot, P. Kumaraguru, and K. Srinathan. WYSWYE: Shoulder Surfing Defense for Recognition based Graphical Passwords. In *Proc. OzCHI '12*, 2012.

[31] D. Kim, P. Dunphy, P. Briggs, J. Hook, J. Nicholson, J. Nicholson, and P. Olivier. Multi-touch authentication on tabletops. In *Proc. CHI '10*. ACM, 2010.

[32] M. Kumar, T. Garfinkel, D. Boneh, and T. Winograd. Reducing Shoulder-surfing by Using Gaze-based Password Entry. In *Proc. SOUPS '07*, 2007.

[33] J. R. Lewis. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7(1):57–78, 1995.

[34] B. Malek, M. Orozco, and A. El Saddik. Novel

shoulder-surfing resistant haptic-based graphical password. In *Proc. EuroHaptics '06*, 2006.

[35] D. Nali and J. Thorpe. Analyzing user choice in graphical passwords. Tech. report TR-04-01, Carleton University, 2004.

[36] J. Nicholson, P. Dunphy, and L. Coventry. A security assessment of tiles: a new portfolio-based graphical authentication system. In *Proc. CHI '12 Extended Abstracts*. ACM, 2012.

[37] I. Oakley and A. Bianchi. Multi-touch passwords for mobile device access. In *Proc. UbiComp '12*. ACM, 2012.

[38] A. Paivio. Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45(3), 1991.

[39] R. Raguram, A. M. White, D. Goswami, F. Monrose, and J.-m. Frahm. iSpy: automatic reconstruction of typed input from compromising reflections. In *Proc. CCS '11*. ACM, 2011.

[40] K. Renaud. Guidelines for designing graphical authentication mechanism interfaces. *Int. Journal of Info. and Comp. Sec.*, 3(1):60, 2009.

[41] D. Ritter, F. Schaub, M. Walch, and M. Weber. MIBA: Multitouch image-based authentication on smartphones. In *Proc. CHI '13 Extended Abstracts*. ACM, 2013.

[42] A. Sahami Shirazi, P. Moghadam, H. Ketabdar, and A. Schmidt. Assessing the vulnerability of magnetic gestural authentication to video-based shoulder surfing attacks. In *Proc. CHI '12*. ACM, 2012.

[43] H. Sasamoto, N. Christin, and E. Hayashi. Undercover: authentication usable in front of prying eyes. In *Proc. CHI '08*. ACM, 2008.

[44] F. Schaub, R. Deyhle, and M. Weber. Password entry usability and shoulder surfing susceptibility on different smartphone platforms. In *Proc. Mobile and Ubiquitous Multimedia (MUM '12)*. ACM, 2012.

[45] S. Schechter, C. Herley, and M. Mitzenmacher. Popularity is everything: A new approach to protecting passwords from statistical-guessing attacks. In *Proc. 5th USENIX Workshop on Hot Topics in Security (HotSec '10)*. USENIX Assoc., 2010.

[46] J. Seifert, A. De Luca, B. Conradi, and H. Hussmann. TreasurePhone: Context-sensitive user data protection on mobile phones. In *Proc. Pervasive '10*, 2010.

[47] X. Suo, Y. Zhu, and G. S. Owen. Graphical Passwords: A Survey. In *Proc. Annual Comp. Security Applications Conf. (ACSAC '05)*. IEEE, 2005.

[48] H. Tao and C. Adams. Pass-Go: A Proposal to Improve the Usability of Graphical Passwords. *Int. J. Network Security*, 7(2):273–292, 2008.

[49] F. Tari, A. A. Ozok, and S. H. Holden. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proc. SOUPS'06*. ACM, 2006.

[50] J. Thorpe and P. van Oorschot. Towards Secure Design Choices for Implementing Graphical Passwords. In *Proc. Annual Comp. Security Applications Conf. (ACSAC '04)*. IEEE, 2004.

[51] J. Thorpe and P. C. van Oorschot. Graphical dictionaries and the memorable space of graphical passwords. In *Proc. USENIX Security Symposium*. USENIX Assoc., 2004.

[52] P. C. van Oorschot and J. Thorpe. On predictive models and user-drawn graphical passwords. *ACM TISSEC*, 10(4):1–33, 2008.

[53] P. C. van Oorschot and J. Thorpe. Exploiting Predictability in Click-based Graphical Passwords. *Journal of Computer Security*, 19(4):669–702, 2011.

[54] S. Wiedenbeck, J. Waters, L. Sobrado, and J.-c. Birget. Design and Evaluation of a Shoulder-Surfing Resistant Graphical Password Scheme. In *Proc. Conf. Advanced visual interfaces (AVI '06)*, 2006.

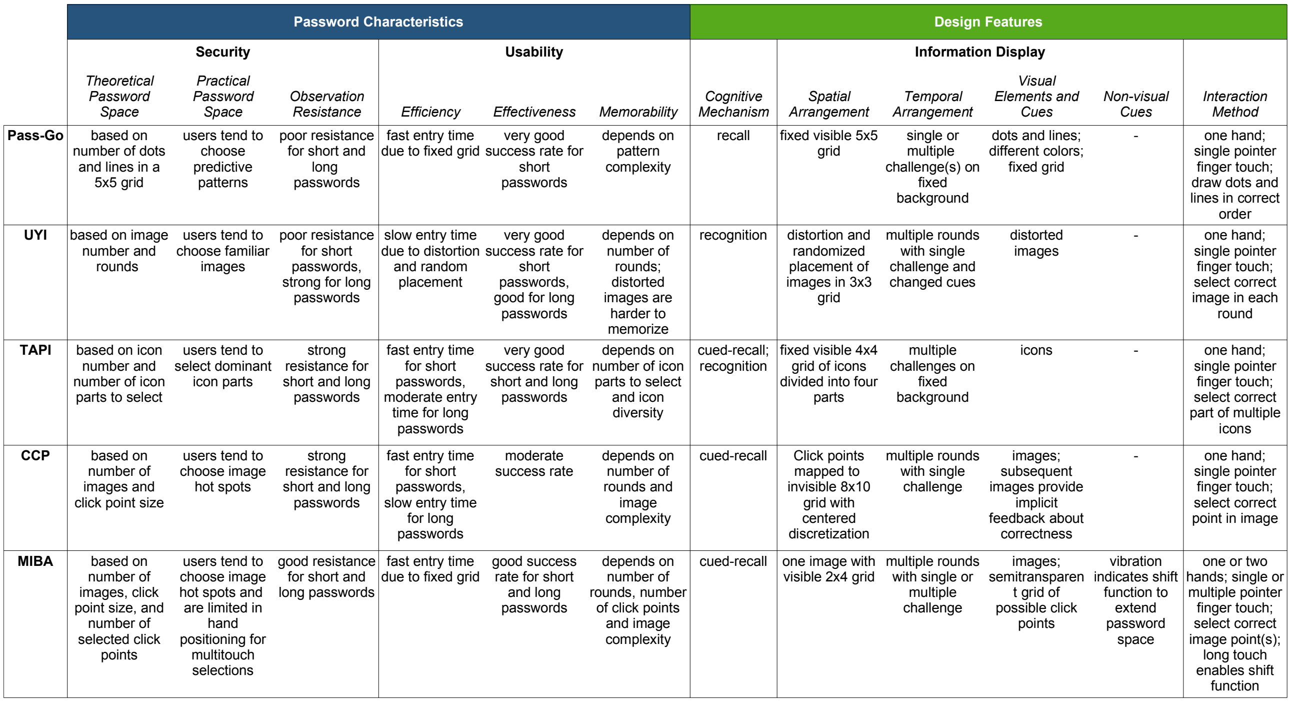
[55] N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan. Shoulder surfing defence for recall-based graphical passwords. In *Proc. SOUPS '11*. ACM, 2011.

APPENDIX

A. DESIGN SPACE MAPPINGS

Table 3 summarizes how each of the evaluated graphical password schemes maps onto the design space and how specific design space features are instantiated. Due to the fact that the smartphone capabilities were the same for all password schemes (see Section 4.1), the table provides only the mapping for password characteristics and design features.

Table 3: Device space mapping of graphical password schemes.



	Password Characteristics						Design Features					
	Security			Usability				Information Display				
	Theoretical Password Space	*Practical Password Space*	*Observation Resistance*	*Efficiency*	*Effectiveness*	*Memorability*	*Cognitive Mechanism*	*Spatial Arrangement*	*Temporal Arrangement*	*Visual Elements and Cues*	*Non-visual Cues*	*Interaction Method*
Pass-Go	based on number of dots and lines in a 5x5 grid	users tend to choose predictive patterns	poor resistance for short and long passwords	fast entry time due to fixed grid	very good success rate for short passwords	depends on pattern complexity	recall	fixed visible 5x5 grid	single or multiple challenge(s) on fixed background	dots and lines; different colors; fixed grid	-	one hand; single pointer finger touch; draw dots and lines in correct order
UYI	based on image number and rounds	users tend to choose familiar images	poor resistance for short passwords, strong for long passwords	slow entry time due to distortion and random placement	very good success rate for short passwords, good for long passwords	depends on number of rounds; distorted images are harder to memorize	recognition	distortion and randomized placement of images in 3x3 grid	multiple rounds with single challenge and changed cues	distorted images	-	one hand; single pointer finger touch; select correct image in each round
TAPI	based on icon number and number of icon parts to select	users tend to select dominant icon parts	strong resistance for short and long passwords	fast entry time for short passwords, moderate entry time for long passwords	very good success rate for short and long passwords	depends on number of icon parts to select and icon diversity	cued-recall; recognition	fixed visible 4x4 grid of icons divided into four parts	multiple challenges on fixed background	icons	-	one hand; single pointer finger touch; select correct part of multiple icons
CCP	based on number of images and click point size	users tend to choose image hot spots	strong resistance for short and long passwords	fast entry time for short passwords, slow entry time for long passwords	moderate success rate	depends on number of rounds and image complexity	cued-recall	Click points mapped to invisible 8x10 grid with centered discretization	multiple rounds with single challenge	images; subsequent images provide implicit feedback about correctness	-	one hand; single pointer finger touch; select correct point in image
MIBA	based on number of images, click point size, and number of selected click points	users tend to choose image hot spots and are limited in hand positioning for multitouch selections	good resistance for short and long passwords	fast entry time due to fixed grid	good success rate for short and long passwords	depends on number of rounds, number of click points and image complexity	cued-recall	one image with visible 2x4 grid	multiple rounds with single or multiple challenge	images; semitransparen t grid of possible click points	vibration indicates shift function to extend password space	one or two hands; single or multiple pointer finger touch; select correct image point(s); long touch enables shift function