Memorability of Cued-Recall Graphical Passwords with Saliency Masks

Florian Alt[1], **Mateusz Mikusz**[2], **Stefan Schneegass**[3], **Andreas Bulling**[4]

[1]LMU Munich – Group for Media Informatics (Amalienstrasse 17, 80333 München, Germany)
[2]Lancaster University (School of Computing & Communications, Lancaster University, Lancaster, UK)
[3]University of Stuttgart – VIS (Pfaffenwaldring 5a, 70569 Stuttgart, Germany)
[4]Max Planck Institute for Informatics (Campus E1 4, 66123 Saarbrücken, Germany)

ABSTRACT

Cued-recall graphical passwords have a lot of potential for secure user authentication, particularly if combined with saliency masks to prevent users from selecting weak passwords. Saliency masks were shown to significantly improve password security by excluding those areas of the image that are most likely to lead to hotspots. In this paper we investigate the impact of such saliency masks on the memorability of cued-recall graphical passwords. We first conduct two pre-studies (N=52) to obtain a set of images with three different image complexities as well as real passwords. A month-long user study (N=26) revealed that there is a strong learning effect for graphical passwords, in particular if defined on images with a saliency mask. While for complex images, the learning curve is steeper than for less complex ones, they best supported memorability in the long term, most likely because they provided users more alternatives to select memorable password points. These results complement prior work on the security of such passwords and underline the potential of saliency masks as both a secure and usable improvement to cued-recall gaze-based graphical passwords.

Author Keywords

User authentication; Cued-recall graphical passwords; Memorability; User study; Saliency masks

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*Input devices and strategies*; K.6.5 Computing Milieux: Security and Protection—*Authentication*

INTRODUCTION

Graphical passwords have long been investigated as a means for user authentication (see [7, 27] for extensive reviews). Graphical password systems for authentication either rely on a recognition task (the user has to recognize one or more graphical entities that together constitute their password), a memory recall task (the user has to freely remember their password), or a cued-recall task (the user is presented cues, for example, an image, that help them to recall their password).

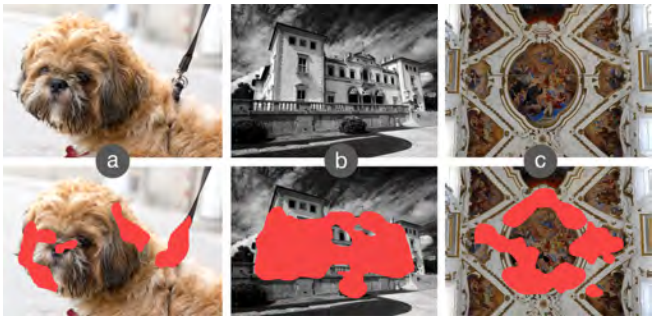

Figure 1: In this work we compare the memorability of graphical passwords defined on images with low (a), medium (b) and high (c) complexity. The bottom row shows the saliency mask versions of the images.

Cued-recall graphical passwords in particular have considerable advantages over traditional approaches, such as alphanumeric passwords or PINs, as they leverage the vast capacity and capabilities of the visual system [3, 16]. A key advantage is their increased resistance to guessing attacks, due to the potentially larger theoretical password space [10]. In addition, memorability of multiple graphical passwords is substantially more effective than multiple PIN numbers and can be further improved by using mnemonics to aid their recall [23].

Previous research suggests that the security of cued-recall graphical passwords consisting of several password points defined on a single image can be significantly increased by masking out the most salient areas of the image during password selection [10]. These saliency masks can be automatically calculated using a computational model of visual attention and can effectively prevent users from choosing weak password points that fall into attentional hotspots.

While saliency masks improve password security it is still an open question what impact they have on the ability of users to remember their passwords over long periods of time. In this work we investigate the impact of saliency masks on short-term and long-term password memorability in a study with 26 participants. Furthermore, our investigation provides insights into how saliency masks impact on users' password selection strategies.

MUM '16, Deceomber 12 – 15, 2016, Rovaniemi, Finland

ACM 978-1-4503-4860-7/16/12...$15.00
DOI: http://dx.doi.org/10.1145/3012709.3012727

CONTRIBUTION STATEMENT

This work contributes (1) a user study and analysis on the impact of saliency masks on the memorability of cued-recall graphical passwords, both in the short term and in the long term, and (2) a quantitative and qualitative comparison of graphical passwords on images with different complexities with and without saliency masks as well as 4-digit PINs.

RELATED WORK

This work draws upon prior research on graphical passwords and techniques to improve password memorability.

Memorability of Graphical Passwords

Password memorability has long been identified as one of the most critical usability issues [9] with 4-5 regularly used passwords per user considered to be a maximum [1]. Graphical passwords were shown to have considerable memorability advantages over alphanumeric passwords or PINs as they leverage the capabilities of the visual system [3, 16].

PassFaces is a commercial graphical password authentication system in which users select four previously self-chosen face images (the password) out of a larger set of images. Brostoff and Sasse evaluated PassFaces and found that participants made fewer login errors compared to users of alphanumeric passwords, despite accessing the system less frequently [8]. Similar findings were obtained by Davis et al. who performed a brief memorability comparison between PassFaces and passwords consisting of sequences of face images [13].

Dhamija and Perrig proposed a system were users authenticated by recognizing previously seen images [15]. They compared their system with user-generated alphanumeric passwords and 4-digit PINs and found that, after one week, participants only made 5 % errors with their system compared to more than 30 % with passwords and PINs.

Widenbeck et al. presented a click-based graphical password scheme, called PassPoints, that allowed users to define passwords by selecting an arbitrary sequence of points on a single image [31]. They compared their scheme to alphanumeric passwords in a longitudinal study over six weeks and found that both password types were equally memorable, but graphical passwords required more time to input. In follow-up work they investigated the influence of the error margin in clicking on the password points as well as the type of image on memorability [30]. They found that while the type of image did not have any influence, small error margins significantly reduced password memorability – most likely because users did not memorize password points accurately enough.

Schneegass et al. presented SmudgeSafe [25], a graphical authentication scheme that applies graphical transformations to the image on which password points are selected. In this way, the resistance against so called smudge-attacks is enhanced. In the context of a usability study participants rated the memorability of the approach to be very high (5 on a 5-Point Likert scale). Data were collected using experience sampling from users who had downloaded and installed the app from GooglePlay. Hence, no information is available as to how long they had previously used the system.

Chiasson et al. conducted a series of studies on the security and usability of click-based graphical passwords that involved users to select one point per image but for a sequence of several images. In an initial study they found that performance was very good in terms of speed, accuracy, and number of errors and that users preferred their approach over PassPoints [12]. In an additional laboratory and large-scale field study they were able to validate their earlier usability claims with respect to password memorability but they also found that the choice of image significantly influenced login success rates [11].

Techniques to Improve Password Memorability

Several researchers investigated techniques to improve memorability of both alphanumeric and graphical passwords. Recently, Zhang et al. argued that interference between different passwords in memory is one of the major challenges to multiple-password recall [33]. They compared two methods to alleviate interferences and found that providing the user with the first letter of their alphanumeric password led to significant improvement of memorability. In a similar study, Vu et al. evaluated the accuracy for recalling alphanumeric passwords, derived from the first letters of the words of a sentence, and found these to be more memorable than random passwords [28]. Yan et al. performed a large-scale study on the memorability and security of phrase-based passwords, i.e. memorable phrases condensed into passwords [32]. They found that phrase-based passwords were as secure as random passwords but significantly easier to remember.

Weiss et al. presented an authentication method, called PassShapes, that involved users in drawing geometric shapes constructed of combinations of eight strokes [29]. They showed that PassShapes increased memorability when users could practice their shapes several times – an effect that even increased over time. Lin et al. presented a variation of the draw-a-secret scheme originally proposed by Jermyn et al. [20] that used a qualitative mapping between user strokes and password to improve security without decreasing memorability [22].

Moncur et al. investigated whether users find multiple graphical passwords more memorable than multiple PINs and also compared two memory augmentation strategies for increasing the memorability of graphical passwords [23]. They found that multiple graphical passwords could be more easily remembered than multiple PINs and that memorability could be further improved by using mnemonics to aid their recall.

Summary

All of these studies investigated memorability as well as techniques to improve the memorability of alphanumeric and cued-recall graphical passwords. However, only few of them evaluated long-term password memorability over the duration of several weeks, like for example Wiedenbeck et al. [30, 31] and De Luca et al. [14]. In addition, there is no prior work that investigated the impact of user-rated image complexity on memorability.

The goal and novel contribution of this work is to evaluate memorability for saliency masks in the long-term that were so far only shown to increase the security of cued-recall graphical passwords [10].

SALIENCY MASKS

To generate the saliency masks, we used a Graph-Based Visual Saliency (GBVS) model (see [18] for details on GBVS and [17] for the MATLAB toolbox we used). GBVS was shown to predict human fixations on natural images with superior performance to the original visual saliency algorithm presented by Itti et al. [19]. The saliency masks were calculated using the default parameters of the toolbox. The greyscale heat maps returned by the GBVS algorithm were first normalised and thresholded at the 0.5 level so as to separate salient and non-salient areas of the images. The salient areas were then used as saliency masks that were overlaid in red onto the original images (see bottom row of Figure 1).

While previous work focused on the security of saliency masks [10], in this work we quantitatively evaluate the impact of saliency masks on long-term memorability of graphical passwords. As a baseline, we use 4-digit PINs that are common when it comes to user authentication, for example, at ATMs. Specifically, we investigate how saliency masks impact on the number of false login attempts, the accuracy of password point selection and whether there is a learning effect.

USER STUDIES

Memorability of graphical passwords defined on a single image may be influenced by the complexity of that image as well as by the specific selection of password points. We therefore conducted two pre-studies to obtain a random set of images with three user-rated image complexities (pre-study 1), and real passwords, i.e. passwords defined by users, to prevent from potential biases that manual selection of passwords may have introduced (pre-study 2). These two pre-studies were followed by the main study to investigate long-term graphical password memorability with and without saliency masks.

Pre-Study 1: Image Selection

A large number of different metrics for quantifying image complexity were proposed in the past. For example, while some researchers used gray-level dependent metrics, such as contrast and feature distribution, others proposed edge-dependent metrics (for example, target edge strength, average contour length, or edge characteristics), or shape/size-dependent metrics (for example, number of pixels of target, the aspect ratio of the target, or largest target size) [5, 6, 24].

These metrics were developed to meet the requirements of a specific application or use case, such as automated target recognition in images, and do not account for the user's subjective perception and cognitive abilities. We hence opted to follow a user-centered approach for image selection: An independent group of people was first asked to rate the complexity of a large image set from which we then manually selected three images with user-rated low, medium, and high complexity.

We automatically retrieved a random set of 20 different images from flickr[1]that were under public domain license. The images mainly depict sceneries and persons similar to the actual preferences of users when selecting images for graphical passwords [2]. The selection of images was not restricted in any way but we made sure to obtain images with a resolution of at least 1600×1200 pixels. All oversized images were manually scaled down to that resolution if need be. To obtain complexity ratings for these images, we created a website that displayed two random but always different images next to each other. Participants were then asked to click on the image they considered to be more complex (see Figure 2a). It is important to note that we didn't provide any definition of complexity as we wanted participants to take their decision as quickly as possible and based solely on their intuitive understanding and subjective notion of "image complexity".

[1] http://www.flickr.com/

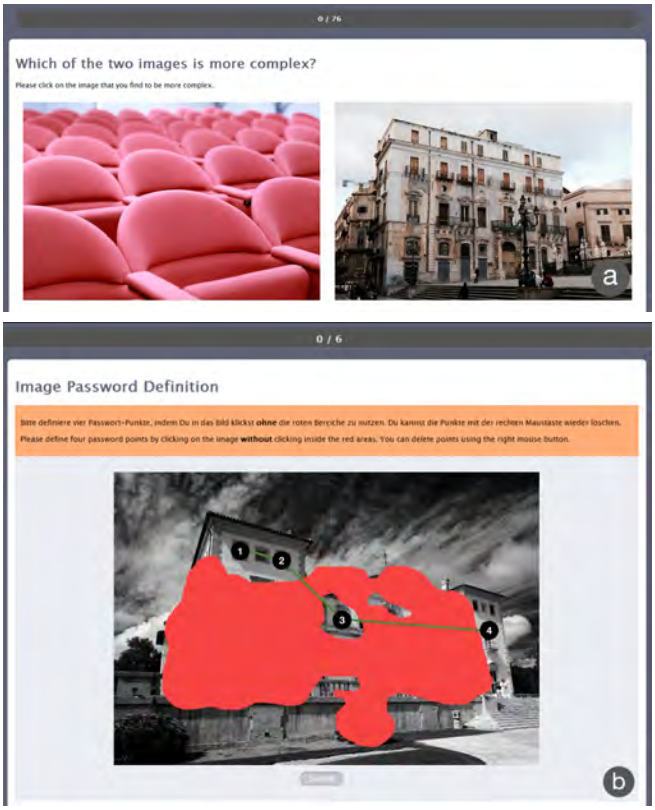

Figure 2: Web interfaces to (a) rate pairs of images for image complexity, and (b) define graphical passwords consisting of four password points in a single image on the example of an image with saliency mask (red area).

The website was implemented to ensure that all image combinations were presented in randomized order (380 in total). Each image pair was shown twice with image positions swapped. Each participant had to make 76 comparisons (20 % of the entire data set), which took them 5 to 10 minutes. We randomly selected three image pairs to be compared twice during the study to be able to evaluate a-posteriori how consistent each participant had been in rating the complexity. Participants were recruited from University's mailing lists, and were not compensated for taking part in the study.

43 participants (23 female and 20 male) aged between 20 and 68 years ($M = 31.0$, $SD = 8.9$) completed the study. This resulted in a total of 3268 image comparisons. To determine the overall complexity of each image we calculated how often participants considered an image to be more complex than another. The most complex image was selected 276 times, whereas the least complex image was selected only 48 times.

Out of the 43 participants, 26 rated all control image pairs consistently, 15 participants rated two out of three consistently, two participants rated one out of three consistently, and no participant rated all image pairs inconsistently. We selected three images for the main study: the most complex image (high complexity), the image ranked tenth (medium complexity), and the least complex image (low complexity). A statistical analysis of the ratings for these images confirmed that the image rated as most complex was significantly more complex than the other two, $\chi^2(1) = 17.00$, $p < .05$, $\chi^2(1) = 13.24$, $p < .05$. The analysis also confirmed that the image with medium complexity was significantly more complex than the one rated as least complex, $\chi^2(1) = 5.94$, $p < .05$.

Pre-Study 2: Password Selection

We opted to not allow participants in the main study to select passwords themselves to prevent any potential influence of the *generation effect*, i.e. the fact that memorability for passwords is better if participants are allowed to generate them instead of just reading them [26]. This is common practice also in the real world, for example, for credit cards for which PINs are assigned by the bank. Because no established method exists for creating graphical passwords in a similar fashion as random alphanumeric password generators, we again followed a user-centred approach for password selection.

We needed three types of real passwords: 4-digit PINs as well as graphical passwords consisting of four password points in an image with and without a saliency mask. For the PINs, we asked 10 people on the university campus to secretly write down a non-trivial 4-digit PIN that they were not currently using. For the graphical passwords, we created a website and asked participants from the first pre-study to define passwords consisting of four password points by clicking on the images selected in the first pre-study. Each participant had to define passwords for each of the three images as well as with and without saliency mask (Figure 2b). The saliency masks were overlaid in red onto the original images and it was not possible for participants to click into these areas. It was also impossible to select password points that overlapped. There was no time limit for creating the passwords and participants were allowed to modify the password as often as needed. To ensure that participants selected passwords they could remember, they had to validate the password directly prior to finally submitting it. If they were not able to validate the passwords within three attempts, they were asked to redefine the password.

Nine people (three female, six male) aged 20 – 48 years ($M = 30.0$, $SD = 7.5$) participated in the study. Each participant defined six graphical passwords on the three images obtained in the first pre-study. This resulted in a total of 54 real passwords. We had to exclude two trivial PINs from the 10 PINs that we initially collected ("0000" and "1234").

Main Study

The pre-studies yielded a set of images with different complexities as well as real passwords chosen by users. The goal of the main study was to investigate the impact of saliency masks on short- and long-term password memorability. To this end, we conducted a long-term study in which each participant had to remember and login to a custom web interface using three types of passwords: one 4-digit PIN (baseline) and two of the graphical passwords, one of which was defined on an image with saliency mask (*GPs*) and one of which was defined on an image without a saliency mask (*GP*).

Figure 3: Example login attempt. Blue dots and lines denote the password to be remembered, while green dots indicate correctly and red dots wrongly entered password points. White circles visualize the accepted input area around each password point (not shown to participants).

We created a web interface where participants could login using their PIN and graphical passwords. PIN entry was implemented using a standard HTML text form. To enter their graphical passwords, participants could select password points by clicking on the image with the mouse.

Participants were recruited from lectures, University mailing lists, and Facebook. As compensation we raffled three 30 EUR Amazon vouchers among all participants. The study was conducted as a controlled study following a within-subjects repeated measures design. It consisted of an initial session in which participants remotely obtained and validated their passwords using the web interface (*validation* session) and three remote login sessions using the same web interface. Login sessions were performed over several weeks: one week, two weeks, and four weeks after the validation session. Note, that the saliency masks were never shown to the participants and they therefore did not know which passwords had originally been defined with and without a saliency mask.

During the recruiting process, participants were provided with a link to the study website. As participants accessed the link for the first time, the initial session was started. Participants were first introduced to the overall study procedure and asked to complete a short demographic questionnaire. Afterwards, participants were given a brief tutorial in which they were explained the interface and how to enter passwords using the mouse. Participants were then shown the first password and asked to immediately validate it by using it to log into the system. After three failed login attempts they were shown the password again. This procedure was repeated for all three passwords. Following the validation, participants were asked for their opinion on whether they would be able to remember their passwords (5-point Likert item, 1: won't be able to remember at all; 5: will definitely be able to remember).

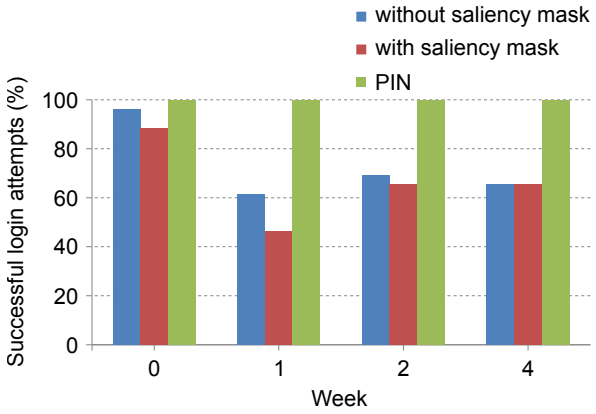

Figure 4: Percentages of successful logins (three attempts to correctly enter the password) for the different types of passwords in the validation session as well as after one, two and four weeks.

After one, two, and four weeks, participants automatically received an email reminding them to log in using the graphical passwords and PIN they had received in the initial session. In order to minimize differences in the time intervals between logins across all participants – that might have influenced memorability – participants received another reminder email after 24 hours. If they did not login within a maximum of 48 hours they were excluded from the study. The presentation order of the passwords was randomized for each session. Participants had three login attempts for each of the three passwords. Similar to earlier long-term studies on password memorability [14, 29, 31], participants were shown the password again if they could not remember one of their passwords.

Finally, all participants who finished the fourth week of the study were asked to complete a short online questionnaire on the perceived complexity of all three images used in the main study and the perceived difficulty of remembering their own graphical passwords and PIN (5-point Likert item, 1: not complex/difficult to remember; 5: very complex/difficult to remember). Given that writing down passwords is common practice [1], we also asked for any password aids participants had used, as well as the daily context in which they had participated (for example, at home, at work, while traveling).

Similar to previous studies on password memorability we compared the three types of passwords in terms of the number of failed login attempts. A login attempt was considered failed if any Euclidean distance between selected and actual password points was larger than a fixed threshold of 50 pixels (Figure 3). We selected this threshold for easier comparison of our results with an earlier study on the security of cued-recall graphical passwords using saliency masks [10].

The downside of this approach is that, in a real-world system, the success of a login attempt, and in turn password memorability, directly depends on the specific value of this threshold: If the threshold is too small, users won't be able to login successfully anymore; if it is too large, login success would increase but security would decrease considerably as well given that any reasonably close password point would be accepted by the authentication system. We therefore also evaluated memorability directly on the Euclidean distances calculated between all

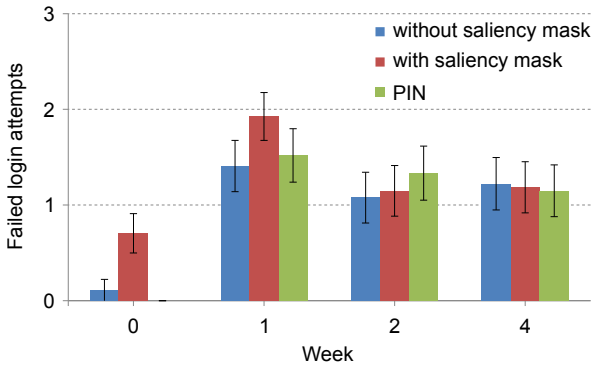

Figure 5: Number of failed login attempts for PINs and graphical password with / without saliency mask over all four sessions (four weeks), averaged over all participants. Error bars indicate the standard error.

four selected and actual password points. This approach does not depend on any specific threshold and allows us to decouple password memorability from potential security issues caused by a "wrong" threshold.

RESULTS

69 participants (14 female, 55 male) aged between 20 and 48 years ($M = 25.3$, $SD = 5.1$) registered for the main study and completed the validation session. Overall, 42 participants dropped out over the course of the following four weeks. In the end, 26 participants (six female, 20 male) aged between 21 and 48 years ($M = 26.0$, $SD = 5.9$) completed all login sessions until after the fourth week. Table 1 provides an overview of all results discussed in the following sections.

Overall Password Memorability

First, we evaluated overall password memorability for the three types of passwords considered in this work: 4-digit PIN, graphical password with saliency mask, and graphical password without saliency mask. We analyzed the percentages of successful login attempts (Figure 4) and the number of failed login attempts (Figure 5) for all four sessions (validation and week 1, 2, 4) and across all 26 participants. The analysis showed that participants could remember all their passwords best in the validation session ($M = 0.27$, $SE = 0.106$). While the average number of failed login attempts across all password types increased considerably after one week ($M = 1.62$, $SE = 0.27$), it decreased again after two ($M = 1.19$, $SE = 0.27$), indicating a learning effect, and remained stable after four weeks ($M = 1.19$, $SE = 0.27$). We tested whether the type of password had any effect on memorability using a one-way repeated measures analysis of variance (ANOVA). The ANOVA showed no statistically significant difference between the three types of passwords, $F(2,50) = 0.756$, $p = .475$. Mauchly's test of sphericity could not show a statistically significant differences for the variances, $\chi^2(2) = 1.586$, $p = .452$. The effect size was very small, $\eta^2 = .029$.

To analyse password memorability in more detail, we followed the multi-store memory model by Atkinson and Shiffrin [4]. According to their model, short-term memory lasts for approximately 15 to 30 seconds, while long-term memory provides the lasting retention of information and lasts from several minutes to a lifetime. Our participants performed the initial password

	Baseline	Saliency mask	Image complexity
Overall	$F(2,50) = 0.756$	N/A	N/A
Long-term	$F(2,50) = 0.123$	$F(1,25) = 1.995$	N/A
Short-term	$F(2,66) = 29.132^{*}$	$t(68) = -2.160^{*}$	$t(22) = 1.719$, $t(21) = -1.356$, $t(23) = 1.169$

Table 1: Overview of the main study results. Statistically significant results are marked with a *. Note that the measure for the baseline tests is the number of failed login attempts, whereas for analyzing saliency masks and image complexity the new measure based on Euclidean distance between password points was used.

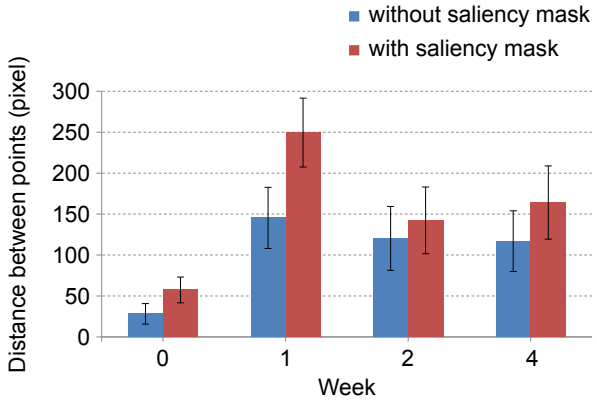

Figure 6: Euclidean distances between the selected and actual password points for graphical passwords with and without saliency mask over all four sessions (four weeks), averaged over all participants. Error bars indicate the standard error.

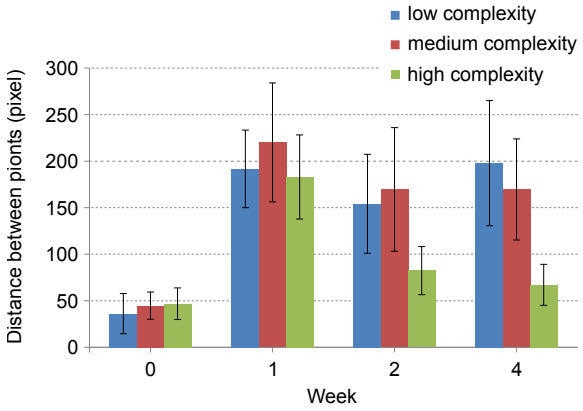

Figure 7: Euclidean distances between the selected and actual password points for graphical passwords defined on images with low, medium, and high complexity over all four sessions (four weeks), averaged over all participants. Error bars indicate the mean, whiskers the standard error.

validation about 30 seconds after receiving their passwords. Thus, in the following analysis, we distinguish between short-term password memorability, which only includes the initial validation session, as well as long-term memorability, which includes all remaining sessions after one, two, and four weeks.

Long-term Memorability

To analyze long-term password memorability for the different types of passwords we again performed a one-way repeated measures ANOVA on the number of failed login attempts of all 26 participants after one, two, and four weeks (Table 2). The analysis did not show any statistically significant difference in memorability between password types, $F(2,50) = 0.123$, $p = .885$. Mauchly's test of sphericity could not show a statistically significant difference, $\chi^2(2) = 1.492$, $p = .474$. The effect size was small, $\eta^2 = .005$.

Impact of Saliency Masks on Graphical Passwords

Furthermore, we investigated the impact of saliency masks on memorability of graphical passwords by using the distance measure described before. As this measure is not meaningful for analyzing PINs, we excluded PINs from this analysis. We first calculated the average Euclidean distance between all selected and actual password points for each participant (see Figure 6). We then performed a one-way repeated measures ANOVA on the mean distances for each graphical password type across all participants. Despite the average distance between selected and actual password point being smaller for passwords defined on images without saliency mask, the ANOVA could not show any statistically significant difference, $F(1,25) = 1.995$, $p = .170$. The effect size was small, $\eta^2 = .074$.

Impact of Image Complexity on Memorability

Finally, we analyzed the impact of image complexity on the memorability of graphical passwords with and without saliency mask. We compared the mean distances between the actual and the selected password points for the three image complexities (low, medium, and high complexity). The mean distances for the three complexities were different (Figure 7). Interestingly, more complex images led to smaller distances between selected and actual password points in the long-term. We explain this by complex images providing more opportunities to select memorable password points not masked out.

Given that images were randomly assigned to participants during registration, we had no influence on how often each image combination was used. Due to the high drop-out rate, we did not conduct further statistical tests.

Short-term Memorability

To analyze short-term password memorability for different password types we again used a one-way repeated measures ANOVA. The ANOVA was performed on the number of failed login attempts of all 69 participants who completed the validation session. The mean number of failed login attempts across all participants was $M_{PIN} = 0.073$ ($SD = 0.356$) for the PIN, $M_{GP} = 0.3188$ ($SD = 0.757$) for the graphical password without saliency mask, and $M_{GPs} = 0.5507$ ($SD = 1.007$) for the graphical password with saliency mask. Because Levene's test indicated no homogeneity of variances, $F(2,66) = 29.132$, we performed a Games-Howell's post-hoc test to analyze the differences between the types of passwords in pairwise comparisons. The test showed a statistically significant difference between both graphical passwords and the PIN, $p < .05$.

	GP without Saliency Mask	GP with Saliency Mask	PIN
Week 1	$M = 1.462$, $SD = 1.392$	$M = 2.000$, $SD = 1.265$	$M = 1.577$, $SD = 1.447$
Week 2	$M = 1.077$, $SD = 1.354$	$M = 1.154$, $SD = 1.405$	$M = 1.308$, $SD = 1.490$
Week 4	$M = 1.269$, $SD = 1.430$	$M = 1.115$, $SD = 1.366$	$M = 1.077$, $SD = 1.383$

Table 2: Means and standard deviations of the number of failed login attempts for different password types after first, second, and fourth week of study.

Impact of Saliency Mask on Memorability

Similar to the long-term analysis, we investigated the impact of saliency masks on short-term memorability of graphical passwords using the distance measure described before. We first compared the Euclidean distances between the selected and actual password points. For graphical passwords without saliency mask the mean distance was 33.16 px ($SD = 55.94$ px) while for graphical passwords with saliency mask it was 57.04 px ($SD = 85.45$ px). We then performed a paired t test, which showed a statistically significant effect of saliency masks on password memorability, $t(68) = -2.160$, $p < .05$. The effect size was small, $r = .03$. This result suggests that passwords defined on an image without salience mask are easier to remember in the short-term.

Impact of Image Complexity on Memorability

We finally analyzed the impact of complexity on the short-term memorability of graphical passwords with and without saliency mask. We compared mean distances between the actual and selected password points for the three image complexities (low, medium, and high complexity). The mean distances were $M_{high} = 45.01$ ($SE = 10.93$), $M_{medium} = 45.24$ px ($SE = 9.72$ px), and $M_{low} = 44.74$ px ($SE = 11.28$ px). Because we only compared image complexity for the two graphical passwords, we performed a series of paired t tests with Bonferroni correction of the significance level to correct the p-value to $p = (0.05/3 = .016)$. The results could not reveal statistically significant difference in password memorability between image complexities: low vs. medium: $t(22) = 1.719$, $p = .100$; low vs. high: $t(21) = -1.356$, $p = .189$; medium vs. high: $t(23) = 1.169$, $p = .254$.

Questionnaire

From the questionnaires completed by 24 participants after the fourth week, we received valuable feedback on the perceived difficulty of remembering their own graphical passwords (see Figure 8) as well as on the image complexities. To analyse the perceived difficulty, we performed a Wilcoxon signed ranks test but did not find any statistically significant difference between graphical passwords with saliency mask ($Mdn_{GPs} = 2.5$) and graphical password without saliency mask ($Mdn_{GP} = 3$), $T = 98$, $p > .05$, $r = -.005$. To analyze the perceived image complexity we used a Friedman analysis of variance by ranks on participant responses for the three complexities.

We also asked for the context in which participants had conducted the study. This included the location or situation in which they entered the passwords, any memory aids used, or whether any elements in the images helped them to memorize the passwords. For the memory aids, we explicitly told participants that their answer would not influence their chances in the lottery but that honest answers would be of high importance.

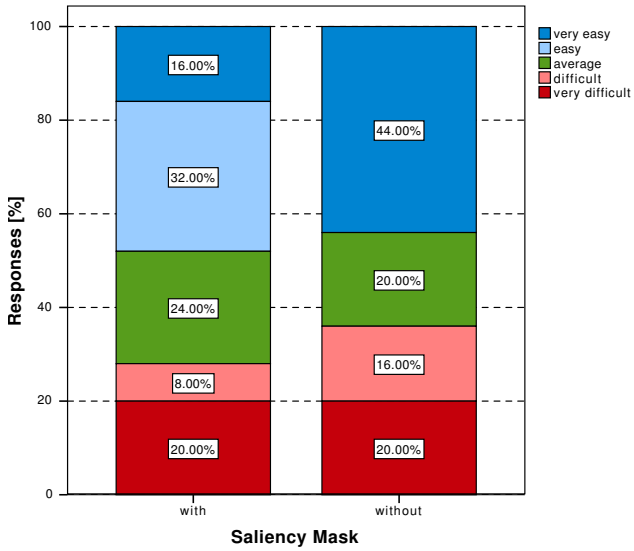

Figure 8: Distribution of responses to question "How difficult was it to remember your graphical passwords?".

We found that 17 out of 24 participants performed the study at home, four did so at work, and three in varying locations (for example, at home, at work, or while commuting). 15 participants reported of not having used any memory aids. Two participants wrote down the passwords, while seven participants used other types of memory aids or mnemonics.

Participants in the second pre-study reported of having used different strategies for *defining* passwords, such as selecting password points that form a square, for example around an object or an animal's eye, or connecting structures of a building (see Figure 9). This was exploited by some participants in the main study for *memorizing* these passwords. Some tried to create stories or rhymes around the password points, memorized the arrangement of the single password points with regard to each other, or tried to fit them into geometrical shapes (*"For the dog, the password was in a clockwise square shape around the eye."*, P4). Image elements that the participant felt attached to also seemed to be useful as memory aids for some (for example, *"I once had a dog – I could instantly remember the password.", P17*). Finally, some participants reported that the symmetric alignment of password points made it easier to remember the complete password.

Similarly, participants in the main study reported of having used different strategies to memorize PINs. Some created equations from the digits (for example, the first digit is the sum of the last two), associated the PIN with important dates (*"The last two digits were my birthday in reverse order."*, P38), created sequences of ascending or descending digits, memorized the arrangement of the digits on a keypad, or fitted the digits into a rhythm (*"eight–one–forty–four"*, P4).

Figure 9: Example strategies to select and memorize graphical passwords: clockwise square shape around objects (in this case the dog's eye) (a), straight line (in this case the four rooftop corners of a building) (b), and prominent points (c).

DISCUSSION

For passwords defined on images with saliency masks we found that at the end of our study, users on average selected points about 30 px further away from the actual password point, compared to passwords defined on images without saliency masks. This suggests a tradeoff between security and usability which needs to be taken into account by designers of authentication schemes based on graphical passwords. Interestingly, this distance constantly became smaller during our study, suggesting that there is indeed a learning effect which may lead to that ultimately password points are remembered more accurately. This needs to be investigated in future work.

Another interesting observation was that images with a higher complexity seem to be able and deal with the aforementioned challenge. It is striking from Figure 7 that for images with high complexity, the distance between selected and actual password point was quite small in week four. We assume the reason for this to be that, apart from the learning effect, complex images provide more opportunities to select memorable password points. This suggests that images of high complexity might be a better choice for authentication schemes based on graphical passwords. Note, however, that the impact of image complexity on security is still to be determined in future work.

Furthermore, we qualitatively assessed heat maps created from passwords which participants of the main study rated to be easy and difficult to remember. Heat maps visually encode image areas in which password points were selected by several participants; the more password points fall in a specific box (edge length 50×50 px) of a regular grid overlaid onto the image, the darker the color. This provided some interesting insights. Saliency masks seem to be effective in masking out many of the preferred password points. At the same time, participants found a number of workarounds. For example, some users defined passwords in a square shape surrounding a masked out object, which aided memorability. Also, users still defined passwords close to prominent points that were, however, masked out. This seemed to make it much more difficult for users to remember these passwords, probably also due to the fact that the saliency mask was missing as a reference in the authentication interface. From this we learn that ways need to be investigated how users can be kept from defining password points that consider the saliency mask as a reference point. An immediate validation upon registration may make this challenge apparent to users.

Finally, users seemed to better memorize passwords defined on images depicting an object users had some kind of connection to, for example, a pet a user owns or once owned. This is similar to a PIN representing the date of a user's birthday or wedding day. Interestingly though, while the latter case might compromise safety (for example, if the attacker knows the user very well), this would not be the case for the graphical passwords, since the image would not support the attacker in guessing the password points. Note, however, that there could be cases where an image could reveal hints to an attacker, for example, if the image showed a number of people, some of which are close friends of the user. As a result, an attacker might guess that password points were defined on the faces of these persons. To exploit the aforementioned phenomenon to increase memorability while not hinting at password points, users could be asked a number of questions upon registration to a graphical password system, for example, about pets they own(ed), cities they live(d) in, or cars they once drove. The system could then suggest a respective image to define the password on (or even with a pre-defined password).

LIMITATIONS AND FUTURE WORK

There are several limitations to our study. First, the size of the tested image set was limited. We cannot exclude further image properties to have any influence on memorability, particularly in cases where users are allowed to select their own images. Second, the size and background of our sample was limited. Hence, we cannot claim that results are generalizable to user groups other than those that participated in the study (for example, children, elderly people, or other cultures). Third, the duration of our study was limited to four weeks. In future work we plan to investigate the impact of specific image properties on password memorability and cases in which users have to remember their passwords for several months. Fourth, we presented pre-defined passwords to the participants. While authentication systems such as ATMs use pre-defined passwords as an additional security enhancing property, many current systems use user-defined passwords. The presented results need to be re-validated with user-defined passwords.

We focused on the mouse as the only input modality. It is important to note that graphical passwords with saliency masks can also be used with other modalities, including but not limited to touch and gaze – or even a combination [21]. This makes the approach potentially useful for any application domain that requires user authentication, from simple online forms, over high-security devices (ATMs), to portable devices

(smartphones, laptops). It is an open question how graphical passwords schemes can be implemented on these platforms. For example, it is not clear whether users should be allowed to choose images and/or passwords themselves or whether certain password patterns or shapes should be enforced or maybe even prohibited to increase security and memorability.

This leads to another potential direction for future research. In contrast to random PIN generators, no established method exists to create secure and memorable graphical passwords. Password points could be randomly distributed over the image but this ignores the potential of constructing password aids based on the selected points, for example, as seen in this work a story that participants make up around these points. It is interesting to think of how a random graphical password generator would look like and how it could consider such aids to make the generated passwords more memorable.

CONCLUSION

In this work we conducted an analysis of the impact of saliency masks on the short and long-term memorability of cued-recall graphical passwords. Results of a user study with 26 participants over four weeks suggest that users remember password points defined on images with saliency masks less accurately. At the same time, images of high complexity seem to be able to address this challenge.

ACKNOWLEDGEMENTS

Note, that for copyright reasons we were not able to include the original images. Images depicted throughout the paper are similar to those used in our studies.

REFERENCES

1. Anne Adams and Martina Angela Sasse. 1999. Users are not the enemy. *Commun. ACM* 42, 12 (1999), 40–46. DOI:http://dx.doi.org/10.1145/322796.322806
2. Florian Alt, Stefan Schneegass, Alireza Sahami Shirazi, Mariam Hassib, and Andreas Bulling. 2015. Graphical Passwords in the Wild: Understanding How Users Choose Pictures and Passwords in Image-based Authentication Schemes. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 316–322. DOI:http://dx.doi.org/10.1145/2785830.2785882
3. Antonella De Angeli, Lynne Coventry, Graham Johnson, and Karen Renaud. 2005. Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. *International Journal of Human-Computer Studies* 63, 1-2 (2005), 128 – 152. DOI:http://dx.doi.org/10.1016/j.ijhcs.2005.04.020
4. R.C. Atkinson and R.M. Shiffrin. 1968. Human memory: A proposed system and its control processes. *The psychology of learning and motivation: Advances in research and theory* 2 (1968), 89–195.
5. J. Beard, L. Clark, and V. Velten. 1985. Characterization of ATR Performance in relation to image measurements. *ATRWG Report, AFWAL/AARF, Wright Patterson AFB, OG* 45433 (1985).
6. B. Bhanu. 1986. Automatic Target Recognition: State of the Art Survey. *IEEE Transactions on Aerospace and Electronic Systems,* AES-22, 4 (1986), 364 –379. DOI: http://dx.doi.org/10.1109/TAES.1986.310772
7. Robert Biddle, Sonia Chiasson, and Paul C Van Oorschot. 2012. Graphical passwords: Learning from the first twelve years. *Comput. Surveys* 44, 4 (2012), 19.
8. Sacha Brostoff and M Angela Sasse. 2000. Are Passfaces more usable than passwords? A field trial investigation. In *People and Computers XIV – Usability or Else!* Springer, Berlin-Heidelberg, 405–424.
9. Alan S. Brown, Elisabeth Bracken, Sandy Zoccoli, and King Douglas. 2004. Generating and remembering passwords. *Applied Cognitive Psychology* 18, 6 (2004), 641–651. http://dx.doi.org/10.1002/acp.1014
10. Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2012. Increasing the Security of Gaze-Based Cued-Recall Graphical Passwords Using Saliency Masks. In *Proceedings of the 30th SIGCHI International Conference on Human Factors in Computing Systems (CHI'12)*. 3011–3020. DOI: http://dx.doi.org/10.1145/2207676.2208712
11. Sonia Chiasson, Robert Biddle, and P. C. van Oorschot. 2007a. A Second Look at the Usability of Click-based Graphical Passwords. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*. ACM, New York, NY, USA, 1–12. DOI: http://dx.doi.org/10.1145/1280680.1280682
12. Sonia Chiasson, P. C. van Oorschot, and Robert Biddle. 2007b. Graphical Password Authentication Using Cued Click Points. In *Proceedings of the 12th European Symposium On Research In Computer Security*. Springer, Berlin-Heidelberg, 359–374.
13. Darren Davis, Fabian Monrose, and Michael K. Reiter. 2004. On User Choice in Graphical Password Schemes. In *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13 (SSYM'04)*. USENIX Association, Berkeley, CA, USA, 11–11. http://dl.acm.org/citation.cfm?id=1251375.1251386
14. Alexander De Luca, Martin Denzel, and Heinrich Hussmann. 2009. Look into My Eyes!: Can You Guess My Password?. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*. ACM, New York, NY, USA, Article 7, 12 pages. DOI: http://dx.doi.org/10.1145/1572532.1572542
15. Rachna Dhamija and Adrian Perrig. 2000. DéJà Vu: A User Study Using Images for Authentication. In *Proceedings of the 9th Conference on USENIX Security Symposium - Volume 9 (SSYM'00)*. USENIX Association, Berkeley, CA, USA, 4–4. http://dl.acm.org/citation.cfm?id=1251306.1251310

16. Katherine M. Everitt, Tanya Bragin, James Fogarty, and Tadayoshi Kohno. 2009. A Comprehensive Study of Frequency, Interference, and Training of Multiple Graphical Passwords. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 889–898. DOI: http://dx.doi.org/10.1145/1518701.1518837

17. Jonathan Harel. 2006. Graph-Based Visual Saliency Toolbox for MATLAB, http://www.klab.caltech.edu/ harel/share/ gbvs.php. (2006). http://www.klab.caltech.edu/~harel/share/gbvs.php

18. Jonathan Harel, Christof Koch, and Pietro Perona. 2006. Graph-Based Visual Saliency. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*. 545–552.

19. Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (1998), 1254–1259.

20. Ian Jermyn, Alain J Mayer, Fabian Monrose, Michael K Reiter, Aviel D Rubin, and others. 1999. The Design and Analysis of Graphical Passwords.. In *Usenix Security*.

21. Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016. GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices. In *Proceedings of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 6. DOI:http://dx.doi.org/10.1145/2851581.2892314

22. Di Lin, Paul Dunphy, Patrick Olivier, and Jeff Yan. 2007. Graphical Passwords & Qualitative Spatial Relations. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*. ACM, New York, NY, USA, 161–162. DOI: http://dx.doi.org/10.1145/1280680.1280708

23. Wendy Moncur and Grégory Leplâtre. 2007. Pictures at the ATM: Exploring the Usability of Multiple Graphical Passwords. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 887–894. DOI: http://dx.doi.org/10.1145/1240624.1240758

24. R.A. Peters and R.N. Strickland. 1990. Image complexity metrics for automatic target recognizers. In *Proc. of the Automatic Target Recognizer System and Technology Conference*. 1–17.

25. Stefan Schneegass, Frank Steimle, Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2014. SmudgeSafe: Geometric Image Transformations for Smudge-resistant User Authentication. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 775–786. DOI: http://dx.doi.org/10.1145/2632048.2636090

26. Norman J. Slamecka and Peter Graf. 1978. The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory* 4, 6 (1978), 592–604. DOI: http://dx.doi.org/10.1037/0278-7393.4.6.592

27. Xiaoyuan Suo, Ying Zhu, and G. Scott. Owen. 2005. Graphical Passwords: A Survey. In *Proceedings of the 21st Annual Computer Security Applications Conference (ACSAC'05)*. IEEE, 463–472. DOI: http://dx.doi.org/10.1109/csac.2005.27

28. Kim-Phuong L. Vu, Robert W. Proctor, Abhilasha Bhargav-Spantzel, Bik-Lam (Belin) Tai, Joshua Cook, and E. Eugene Schultz. 2007. Improving password security and memorability to protect personal and organizational information. *International Journal of Human-Computer Studies* 65, 8 (2007), 744 – 757. DOI: http://dx.doi.org/10.1016/j.ijhcs.2007.03.007

29. Roman Weiss and Alexander De Luca. 2008. PassShapes: Utilizing Stroke Based Authentication to Increase Password Memorability. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges (NordiCHI '08)*. ACM, New York, NY, USA, 383–392. DOI: http://dx.doi.org/10.1145/1463160.1463202

30. Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. 2005a. Authentication Using Graphical Passwords: Effects of Tolerance and Image Choice. In *Proceedings of the 2005 Symposium on Usable Privacy and Security (SOUPS '05)*. ACM, New York, NY, USA, 1–12. DOI: http://dx.doi.org/10.1145/1073001.1073002

31. Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. 2005b. PassPoints: design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies* 63, 1-2 (2005), 102–127. DOI: http://dx.doi.org/10.1016/j.ijhcs.2005.04.010

32. J. Yan, A. Blackwell, R. Anderson, and A. Grant. 2004. Password memorability and security: empirical results. *IEEE Security Privacy* 2, 5 (2004), 25 –31. DOI: http://dx.doi.org/10.1109/MSP.2004.81

33. Jie Zhang, Xin Luo, Somasheker Akkaladevi, and Jennifer Ziegelmayer. 2009. Improving multiple-password recall: an empirical study. *European Journal of Information Systems* 18, 2 (2009), 165–176. DOI:http://dx.doi.org/10.1057/ejis.2009.9