Sweet-Spotting Security and Usability for Intelligent Graphical Authentication Mechanisms

Marios Belk
Cognitive UX GmbH
Germany &
University of Cyprus
Department of Computer Science
Nicosia, Cyprus
belk@cognitiveux.de

Andreas Pamboris
University of Central Lancashire
School of Sciences &
University of Cyprus
Department of Computer Science
Nicosia, Cyprus
apamboris@uclan.ac.uk

Christos Fidas
University of Patras
Department of Cultural Heritage
Management and New Technologies
Patras, Greece
fidas@upatras.gr

Christina Katsini
University of Patras
HCI Group, Department of Electrical and Computer Engineering
Patras, Greece
katsinic@upnet.gr

Nikolaos Avouris
University of Patras
HCI Group, Department of Electrical and Computer Engineering
Patras, Greece
avouris@upatras.gr

George Samaras
Dipartimento di Fisica e Geologia
University of Cyprus
Department of Computer Science
Nicosia, Cyprus
cssamara@cs.ucy.ac.cy

ABSTRACT[1]

This paper investigates the trade-off between security and usability in recognition-based graphical authentication mechanisms. Through a user study (N=103) based on a real usage scenario, it draws insights about the security strength and memorability of a chosen password with respect to the amount of images presented to users during sign-up. In particular, it reveals the users' predisposition in following predictable patterns when selecting graphical passwords, and its effect on practical security strength. It also demonstrates that a *"sweet-spot"* exists between security and usability in graphical authentication approaches on the basis of adjusting accordingly the image grid size presented to users when creating passwords. The results of the study can be leveraged by researchers and practitioners engaged in designing intelligent graphical authentication user interfaces for striking an appropriate balance between security and usability.

CCS CONCEPTS

• **Human-centered computing** → Empirical studies in HCI • **Security and privacy** → Usability in security and privacy.

KEYWORDS

Recognition-based Graphical Authentication; Security; Usability, User Study, Eye-tracking.

ACM Reference format:

Marios Belk, Andreas Pamboris, Christos Fidas, Christina Katsini, Nikolaos Avouris, George Samaras. 2017. Sweet-Spotting Security and Usability for Intelligent Graphical Authentication Mechanisms. In *Proceedings of WI '17, Leipzig, Germany, August 23-26, 2017*, 8 pages.
http://dx.doi.org/10.1145/3106426.3106488

WI '17, August 23-26, 2017, Leipzig, Germany

ACM ISBN 978-1-4503-4951-2/17/08…$15.00
http://dx.doi.org/10.1145/3106426.3106488

1 INTRODUCTION

A user authentication mechanism has two symbiotic pillars at its core, *security* and *usability*. Secure mechanisms force users to set up passwords that are resilient to authentication attacks, by typically constraining the types of passwords accepted. On the other hand, better usability is typically reflected by the users' ability to construct and remember their secret passwords easily. These two pillars are often contradicting in nature, since more secure passwords are also inherently less intuitive and memorable. As a result, security experts are currently struggling to come up with new authentication approaches that strike a balance between security and usability [1].

Graphical User Authentication (GUA) is one such approach, which is gradually making its way into high-security systems, either as the primary or supplementary mechanism for authenticating users. Some examples include PassFaces [4], Android Pattern Unlock, Windows 10 Picture Gesture Authentication, etc. Different types of GUA approaches (see Biddle et al. [2] for a recent review) require users to either sketch a secret image or pattern on the screen (known as *recall-based mechanisms*) [18, 19, 20], select different positions on a background picture (known as *cued-recall-based mechanisms*)

[21, 22], or select pictures from a pool of alternatives, which are presented to users as a grid of images, during the sign-up process (known as *recognition-based mechanisms*) [4, 5, 3].

Research on *recognition-based graphical authentication* is primarily motivated by two factors that mostly relate to usability [2, 17]: *i)* they leverage the picture superiority effect, claiming that pictures are better recognized and recalled by the human brain than textual information [2, 12, 13]; and *ii)* they offer more usable interaction capabilities (*e.g.* quickly selecting images through finger touch on the screen) [2, 5].

With regards to security, an important countermeasure against password hijacking attacks is to increase the set of images offered as choices during password selection. In theory, this reduces the likelihood of successful attacks, since an attacker's workload increases in accordance to the number of possible image permutations that need to be checked. This paper observes, however, that increasing the amount of images beyond a certain threshold could lead to degradation in usability, which may potentially counteract entirely the benefits of GUA.

In light of this observation, it is important for GUA designers to understand precisely the interplay of usability and security in such alternative authentication mechanisms. This paper postulates three hypotheses regarding the correlation of the number of images presented to users at sign-up, with the security and usability properties of GUA:

- Increasing the amount of image alternatives beyond a certain threshold yields fewer benefits in terms of security than what is initially expected. This is because, although in theory having more images to choose from implies a larger search space for attackers, in practice, users are often influenced by the presentation layout of the corresponding images. *They therefore tend to follow predictable patterns when choosing a password.* Attackers can exploit such patterns in order to crack GUA passwords much faster than what is expected due to the asymptotic complexity of the attacking algorithms.
- Increasing the number of image alternatives during the password selection phase inadvertently impacts the usability of GUA mechanisms: *users are more likely to forget their password when asked to choose from an increasing number of possible alternatives.*
- Limiting the number of alternatives beyond a certain value may cause the opposite effect: *users are not able to attach semantic meaning to their choice of images, therefore leading to less memorable secret passwords.*

Through a user study based on a real usage scenario, we provide evidence to back up the abovementioned hypotheses. Our analysis then draws novel insights to drive the design of future GUA mechanisms that properly account for both the stringent security and usability requirements in today's user authentication landscape.

2 METHOD OF STUDY

The research instruments utilized in the study include: *i)* a recognition-based GUA mechanism; *ii)* a traditional desktop computer; and *iii)* a wearable eye-tracking device.

Recognition-based Graphical Authentication Mechanism. A user authentication mechanism (Figure 1) was designed and developed following guidelines of state-of-the-art graphical authentication approaches: DejaVu [3], Passfaces [4] and ImagePass [5]. During sign-up, users created a graphical password by selecting a fixed number of five unique images (in a particular order) from a static grid of images (image sizes were 100x100 pixels). The choice of password length was based on existing works and is typical in recognition-based graphical authentication [2, 6].

Interaction Device. The graphical authentication mechanism was deployed on traditional desktop computers (Intel core i7 with 8GB RAM, 21-inch screen size monitor, standard keyboard and mouse).

Eye-Tracking Device. The Tobii Pro Glasses 2 [7] wearable eye-tracking device was used to capture the users' eye gaze during password selection. The device has the following specifications: 100 Hz gaze sampling frequency, 4 eye cameras, H.264 1920x1080 pixels at 25 fps.

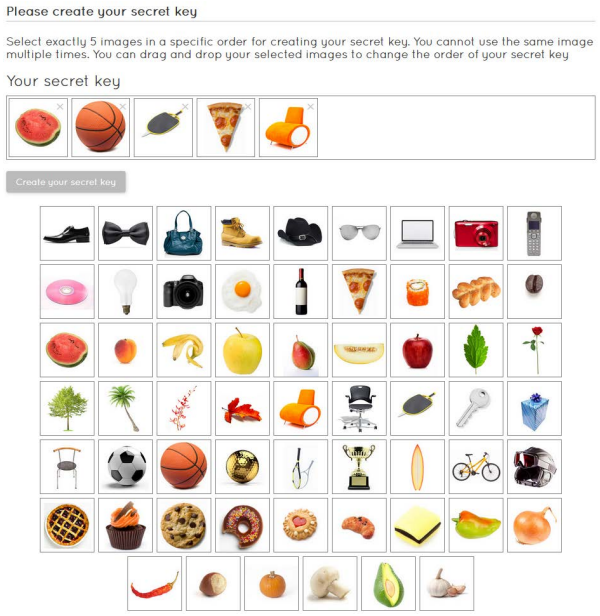

Figure 1: Graphical authentication mechanism during login.

2.1 Procedure

The GUA mechanism was used in the context of an enrolment/registration process for a real-life service in order to increase the ecological validity of our study. The study was conducted in a quiet room in a lab, and each participant was asked to sit in front of a computer at about a 40-cm distance from the screen. Initially, the participants were introduced to the

procedure of the study, having fully familiarized themselves with how the GUA mechanism and eye-tracking equipment work. Participants wearing glasses were allowed to wear the eye-tracking equipment on top of their glasses.

Following a between-subject study design, we split users into four different groups based on the number of images contained in the initial grid of images (*i.e.* 30, 60, 90, and 120). Our choice of a baseline image grid size was influenced by common practice [2, 6], *i.e.* we initially used a grid of 120 images. We then gradually reduced the size (by decrements of 30 images) in order to investigate its effect on usability and security aspects of GUA. Images were grouped in 10 semantically meaningful categories (*e.g.* food, fruit, sport objects, etc.), initially containing 12 images each. Images were not randomized across users aiming to control security and usability aspects beyond the grid size. Starting from the 120- to the 30-grid, we randomly removed the same amount of images (*i.e.* 3 images) from each category, therefore keeping the relative positions of images, and the distribution of image types, stable across image grids.

The participants first signed up to the service by creating a graphical password and subsequently logged in to access the service right after the creation of their password. The reasoning behind this was to control the effect of grid size on login efficiency by avoiding interference of repeat-login effects on the dependent variable. During login, users were required to identify (in the correct order) the five selected images from a grid of twenty-five images (Figure 1).

At the end of the experiment we asked the participants to comment on their experience with the GUA mechanism, their perceived usability, and the strategies they followed when selecting their graphical key.

2.2 Data Collection

For analyzing the *security strength* of the generated graphical passwords, the images selected for each password were saved in a database. To analyze the *GUA mechanism's usability*, we used two metrics: *i) graphical password creation time*, which was measured from page load (after training) until the user successfully created the graphical password; and *ii) login time*, which was measured by keeping track of the time elapsed from first to last image selection (for successful attempts only).

2.3 Participants

A total of 103 individuals participated (43 females, 60 males), ranging in age from 18 to 47 (m = 24.61; sd = 6.02). The ages of the participants were normally distributed according to Shapiro-Wilk's test ($p > .05$). All participants had prior interaction experience with desktop computers. No participant was familiar with the utilized recognition-based graphical authentication mechanism beforehand.

Participation was voluntary and all users provided consent to record their interactions with the GUA-supported system. Also, the participants could opt out of the study any time they liked. Details about the study were not provided until the end to avoid bias effects.

3 SECURITY ANALYSIS

3.1 Theoretical *vs.* Practical Strength of GUA Passwords

The security analysis focuses on: (1) the theoretical security strength of GUA passwords, which is measured by the average combinations required to crack a password selected at random; and (2) the practical security strength of GUA passwords, which is calculated by measuring the resistance of user-selected passwords to an offline brute-force attack. For calculating practical strength, we implemented a brute-force attack that checks all possible permutations of GUA passwords comprising five images, starting from the upper left corner of the image grid and traversing it row-by-row. *(Our choice of attacking method was not entirely based on the best possible way to exploit the user patterns revealed by our work. We rather opted for a mainstream attacking method, to be used as a baseline security metric. As a result, any reported compromises to security are conservative, given that more sophisticated attacks could be used to exploit vulnerabilities associated to user behavior during password creation, which are revealed in this paper. In future work, we plan to explore other attacks that are more tailored to the identified patterns.)*

We measure practical strength by calculating the average "guesses" performed per user until each corresponding password is guessed correctly. The intuition behind this experiment is that, more often than not, theoretical strength is not a representative measure of the practical security strength of user-chosen passwords [11], since users tend to choose their passwords lightly, sacrificing security for convenience during password selection.

Table 1 summarizes the theoretical and practical security strength of graphical passwords for all four types of grids considered. The results suggest that the practical security strength of grid sizes of up to 90 images closely matches the corresponding theoretical strength. This shows that, in practice, the GUA approach under investigation is effective in terms of security, with respect to the theoretical limits of an attacking algorithm.

For a grid size of 120 images, however, the same observation does not hold. In practice, the security strength of the GUA mechanism was found to be a significant 22% less than what is actually backed up by theory.

Table 1: Theoretical *vs.* practical security strength per grid.

Grid	Theoretical strength	Practical strength	Variation
120	**11,434,681,440**	**8,938,727,736**	**-22%**
90	2,636,956,080	2,708,735,227	+3%
60	327,690,720	322,632,269	-2%
30	8,550,360	8,907,618	+4%

3.2 User Selections

The aforementioned results can be explained by investigating how user selections are subconsciously affected by the size of the image grid during sign-up. Figure 2 is an intensity map that shows the frequency of image selections when users were asked to create a password. In particular, it shows the layout of each of the grids considered, with each cell representing a unique image. Cell colors indicate the frequency at which each image was selected as either the first, second, third, fourth or fifth image (darker colors indicate higher frequencies). Based on the color intensity distribution in each grid, at each step of the password selection process, we infer that for grid sizes of up to 90 images, user choices are more evenly distributed across the entire grid. Analysis of users' feedback at the end of the study (see Section 5) further suggests that users chose their password based on the semantics of the images, rather than the order they are presented to them. However, for the 120-image grid, we observe that users tend to first choose images that appear earlier (upper part), and gradually expand to the rest of the grid. This flow closely follows the order at which a typical brute force attack checks all possible permutations of GUA passwords, which explains why in practice passwords are guessed much faster than what is actually expected in theory.

Figure 2. Frequencies of user image selections in each grid.

A chi-square goodness-of-fit test was further conducted to determine whether the selected images are evenly distributed across the entire image grid. We ran the test for each grid group. The results reveal that in the case of the 120-image grid, the number of selected images in the upper part *vs.* lower part of the grid are statistically significantly different, $\chi^2(1) = 41.4$, $p < .001$; 92 images (80%) were selected from the upper part, whereas only 23 images (20%) were selected from the lower part. For the 90-, 60-, and 30-image grids, the chi-square goodness-of-fit test indicates that the selected images are more evenly distributed in the upper and lower parts of the grid ($p > .05$).

Finally, a qualitative analysis based on the heat map shown in Figure 3 further supports the aforementioned behavior of all users who participated in the study that utilized the initial grid of 120 images. It illustrates that the majority of users conveniently focused mostly on the upper left corner of the grid, as a result of having too many images to choose from.

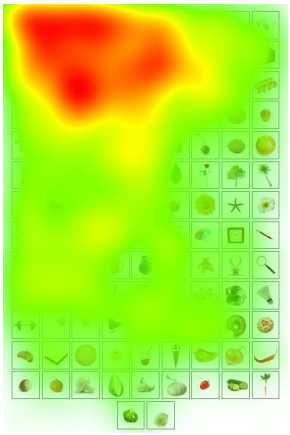

Figure 3. Heat maps based on fixation for the 120-image grid.

4 USABILITY ANALYSIS

4.1 Time to Sign-Up

This analysis examines the effects of grid size on time to sign-up. There were no outliers in the data, as assessed by inspection of a boxplot. The assumption of normality for time to sign-up was violated for all image grids, as assessed by Shapiro-Wilk's test ($p < .05$). Therefore, we conducted the alternative non-parametric Kruskal-Wallis H test to determine if there are differences in time to sign-up between groups of users that interacted with different image grids. Distributions of sign-up times are similar for all groups, as assessed by visual inspection of a boxplot. Figure 4 illustrates the sign-up times per image grid.

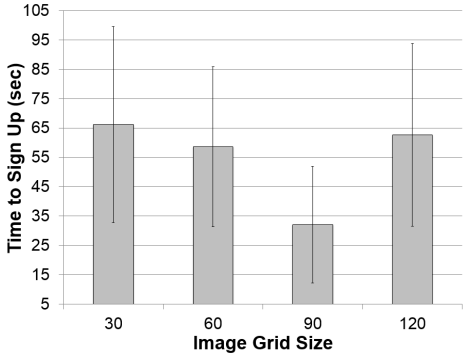

Figure 4. Time to sign-up (sec) per image grid.

Median times are statistically significantly different between the different grids, $\chi_2(3) = 1638.63$, $p < .001$. Results reveal that the 90-image grid is the most usable in terms of time to sign-up. Accordingly, pairwise comparisons were performed using Dunn's procedure [9] with a Bonferroni correction for multiple comparisons between the 90-image grid and all other grids. Adjusted p-values are presented. This post hoc analysis reveals statistically significant differences in median sign-up times between the 30 grid (2367) and 90 grid (1313) ($p < .001$), the 60 grid (2148) and 90 grid (1273) ($p < .001$), and the 120 grid (3110) and 90 grid (1630) ($p < .001$).

A qualitative analysis of the users' feedback on their experiences with the GUA mechanism (see Section 5) explains the aforementioned results; the 90-image grid allows enough flexibility for users to choose semantically meaningful graphical passwords. Below this threshold, the options are limited, whereas above the threshold, users are overwhelmed by the high number of alternatives presented to them, therefore they need more time to choose the images that will form their graphical password.

4.2 Time to Login

We also analyze user login times for each image grid. Inspection of boxplots revealed two significant outliers in the 30 and 120 grids, which were therefore removed. Normality of login times was assessed by Shapiro-Wilk's test indicating that the data were not normally distributed ($p < .05$). Hence, the non-parametric Kruskal-Wallis H test was conducted to examine if the different image grids during sign-up have a main effect on login time. Visual inspection of boxplot distributions revealed that login times are similar for all groups. Figure 5 illustrates the login times per image grid.

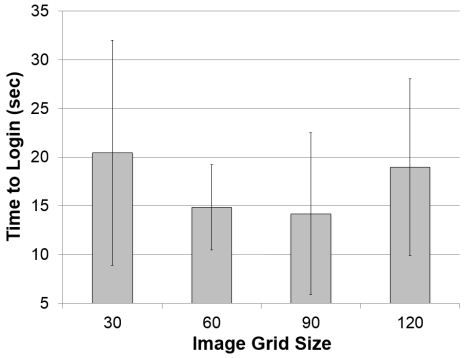

Figure 5. Time to login (sec) per image grid.

Median login times are statistically significantly different between the different grids, $\chi_2(3) = 594.92$, $p < .001$. Furthermore, pairwise comparisons reveal statistically significant differences in median login times between the 30 grid (1104) and 60 grid (835) ($p < .001$), the 30 grid (2034) and 90 grid (1369) ($p < .001$), and the 30 grid (3639) and 120 grid (3485) ($p = .018$).

Results reveal that the 60- and 90-image grids were the most usable in terms of login time. When employing the two extremes (30- and 120-image grid), users required significantly longer login times. As indicated by the qualitative analysis in Section 5, in the case of the 30-image grid, users had a very limited number of choices, preventing them from creating meaningful, and therefore memorable, graphical passwords. On the other hand, using a 120-image grid overwhelmed users with too many options to choose from, who therefore made their choices based on convenience and without attaching semantic meaning to their password. This explains why users required significantly longer login times, compared to the 60- and 90-image grids where they were inclined to create more memorable graphical passwords.

5 QUALITATIVE ANALYSIS

In order to verify the conclusions drawn from our quantitative analysis (Section 4), we also conducted a qualitative analysis on the users' feedback concerning the strategies they followed in creating their graphical password, as well as their perceived experience and usability.

A considerable number of participants of the 120-image grid reported that they created their graphical password from images that were positioned in the upper part of the grid, primarily because they didn't want to scroll to the lower part of the screen. For example, a user stated, *"I focused on the images of the upper part that were presented in the grid as initially shown. I didn't want to lose time scrolling"*.

Participants of the 30-image grid responded that they rather randomly selected images from the grid because they didn't have many options to choose from, which disallowed them from

attaching a meaningful story to the selected graphical key. A user stated, *"I randomly selected images from the grid. I would have preferred more image alternatives to choose from, which would have allowed me to pick a graphical password that I could personally better reason about"*. This explains the increased difficulties in recalling their graphical key, compared to the other user groups, as they required more time to login. In addition, when the same user group was asked whether they would remember their graphical password after one month, the majority answered negatively.

Participants of the 60- and 90-image browsed through most of the available images and considered the entire image grid during password selection, as it was easily visible/accessible on screen. The strategy they followed for selection was based on their preference towards specific categories (*e.g.* hobbies such as football, tennis, etc., or favorite food such as pizza, sushi, etc.), while others created stories based on their unique experiences. A user stated, "*I love sweets so I selected my favorite sweets such as Haribos"*, while another stated, *"I am hungry now so I selected my favorite food, starting from pizza"*. This further supports the quantitative security results since users selected images evenly from the entire grid based on their preference, hence the practical entropy was fairly similar to the theoretical entropy. In addition, the users perceived the usability of the GUA mechanism since the majority stated that they did not face any difficulties during login, and they positively responded when asked whether they would remember their graphical password after one month.

6 MAIN FINDINGS

The main findings of the paper reveal how the image grid size of GUA mechanisms significantly affects: (1) graphical password guessability; (2) time to sign-up; and (3) time to login.

Finding 1 – Mismatch between theoretical and practical security for the 120-image grid. The quantitative and qualitative security analysis revealed that the practical entropy of the 120-image grid was significantly less than the perceived theoretical entropy, since participants did not create random graphical passwords and mostly selected images from the upper part of the image grid. Based on the users' feedback, the participants were overwhelmed by the large number of choices presented to them, and thus they chose images in accordance to the way they were presented to them (focusing mostly on the upper part of the screen, without scrolling down to consider additional choices included in the lower part).

Finding 2 – Out of the image grid sizes considered in this study, the 90-image grid is the most usable in terms of graphical password creation. Results reveal that the two extremes (30- and 120-image grids) are the least usable in terms of time to create a graphical password. For the 30-image grid, this can be explained by the fact that users are not provided with an adequate number of images to choose from, which prevents them from creating meaningful graphical passwords (as also indicated by the qualitative analysis). As a result, it takes longer to choose images to create a meaningful graphical password. On the other hand, the 120-image grid overwhelms users with an abundance of choices, which leads to an increase in the time spent creating a graphical password. Overall, the 60- and 90-image grids are found to be more usable, with the 90-image grid significantly outperforming the rest in terms of time to create the graphical key.

Finding 3 – Out of the image grid sizes considered in this study, the 60- and 90-image grids are the most usable in terms of graphical login. Similar to Finding 2, the 30- and 120-image grids are the least usable in terms of time to login. The limited number of choices in the 30-image grid prevented users from creating meaningful, and therefore memorable, graphical passwords. Hence, login times increased as participants couldn't memorize well their graphical passwords. Regarding the 120-image grid, users were overwhelmed by an increasing number of possible choices and therefore tended to ignore the semantic meaning of images when constructing their graphical passwords. This resulted in less memorable passwords and therefore higher login times, since users tend to pay less attention to the actual semantic meaning of images, and are influenced more by the presentation layout of images (as shown in the heat maps). Again, the 90-image grid was found to be more usable than the rest.

Overall, GUA designers should carefully consider the implications of the image grid size on security and usability. We have shown that a "sweet-spot" exists (this was found to be the 90-image grid), which strikes a balance between the two.

7 IMPLICATIONS

This paper reveals implications of variations in GUA interface designs on the security and usability of GUA mechanisms. Such findings could drive the design of dynamic GUA policies, which will bootstrap design aspects of GUA based on security and usability guidelines.

Building on existing works [15], which address an optimization problem of assigning "best-fit" security mechanisms based on security and usability attributes, the findings of this paper further contribute to the development of an extendable framework that would allow service providers to specify precisely the levels of security and usability they desire. As shown in Figure 6, such an approach would support multiple policies, aligned with different individual user models, to allow service providers customize security and usability aspects of GUA mechanisms in order to benefit end-users. For example, by specifying baseline levels for theoretical and practical security, such a framework could help optimize usability factors of GUA mechanisms. By further considering human and technology factors (*e.g.* research has shown that cognitive decline of older adults [16], cognitive abilities and styles [6, 23, 25, 26], or device characteristics [24] affect GUA usability and security), such a framework would generate a personalized GUA design considering the service provider's desired security and usability

levels, as well as the user's individual and technology-related characteristics.

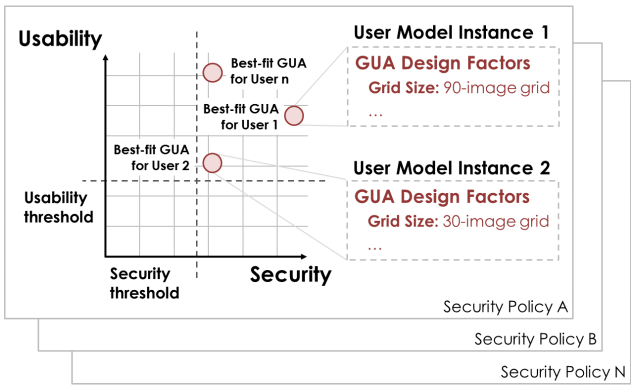

Figure 6. Instances of "best-fit" GUA mechanisms.

We further emphasize the requirement for intelligent mechanisms to guide users towards creating GUA passwords that, on average and across users, cover the entire image grid. Such mechanisms would ensure that the theoretical security of a particular GUA mechanism is not sacrificed due to predictable user behaviors, which can be potentially exploited by attackers. For example, given a large image grid size, strong GUA passwords can be enforced by intelligently positioning images in a grid when presenting them to different users, as opposed to having a static layout that can be leveraged easily by attackers to improve the efficiency of their attacking algorithms. Alternatively, GUA interfaces could prompt users to choose their password incrementally, *i.e.* by presenting only a subset of the possible images to users during each step of the password selection phase. Such approaches would ensure a better coverage of the entire grid, making the most of the attainable theoretical security guarantees of GUA.

Finally, the findings of this study underpin the necessity for providing feedback on the strength of user-selected GUA passwords. Users are often unaware of the security strength of their chosen passwords. This leaves them vulnerable to even the simplest of access control and user authentication attacks. A user authentication mechanism could take into account the results of our security analysis to provide support for a GUA password strength meter. This will indicate the level of "randomness" of a selected graphical password, in light of all possible image permutations, as well as previously selected graphical passwords by other users. The strength meter would therefore ensure that, overall, the images selected by users are evenly distributed across the entire image grid.

8 CONCLUSIONS

This paper provides empirical knowledge on the users' behavior when creating a graphical password by showing the correlation between image grid sizes and the levels of security and usability in GUA mechanisms. Based on this knowledge, it provides valuable insights for designing intelligent GUA mechanisms, which account for the inherent tradeoff between usability and security.

The contribution of this work spans both *theory* and *application*. Regarding *theory*, the study provides evidence that varying equilibriums exist among security and usability aspects in different GUA designs. Regarding *application*, findings underpin the value for versatility in the design and development of GUA. Such knowledge is currently missing, and could be leveraged to provide adaptive and personalized solutions to service providers seeking for configurable levels of security and usability, depending on custom requirements, application and user constraints.

Limitations of the study are related to its ecological and external validity. First, controlling the policy (*e.g.* requiring fixed-size graphical passwords) may not fully reflect the behavior of users in real-life situations. For example, previous work shows that users tend to use more than the minimum length of a given policy [10]. Another limitation affecting ecological validity relates to the nature of in-lab experiments, which however was required in order to be able to use the eye-tracking setup. Furthermore, we only investigated a particular GUA mechanism, although other alternatives exist [2, 14], which may not be affected in the same way. For example, the type of images used (faces *vs.* objects), or the authentication process followed (*e.g.* showing images in one or multiple screens) may influence usability and security in different ways, which is something we plan to explore in future work.

ACKNOWLEDGMENTS

This paper was partially supported by the European H2020 project GrowMeUp (#643647), and the project ADVisE, in the frame of the University of Cyprus' internal funded research projects.

REFERENCES

[1] L. Koved, and E. Stobert. 2016. Who are you?! Adventures in authentication (WAY). *Workshop at the Symposium on USENIX Usable Privacy and Security (SOUPS 2016)*. USENIX Assoc.
[2] R. Biddle, S. Chiasson, and P. van Oorschot. 2012. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys* 44, 4, Article 19 (2012).
[3] R. Dhamija, and A. Perrig. 2000. DejaVu: A user study using images for authentication. In *Proc. of the USENIX Security Symposium*. USENIX Assoc.
[4] Passfaces Corporation. 2009. The Science Behind Passfaces. Retrieved from www.passfaces.com/enterprise/resources/white_papers.htm.
[5] M. Mihajlov, and B. Jerman-Blazic. 2011. On designing usable and secure recognition-based graphical authentication mechanisms. *Interacting with Computers* 23, 6 (2011), 582-593.
[6] Y. Ma, J. Feng, L. Kumin, and J. Lazar. 2013. Investigating user behavior for authentication methods: A comparison between individuals with Down syndrome and neurotypical users. *ACM Transactions on Accessible Computing* 4, 4, Article 15 (2013).
[7] Tobii AB. 2017. Tobii Pro Glasses 2. Retrieved from http://www.tobiipro.com/product-listing/tobii-pro-glasses-2/#Specifications.
[8] D. Davis, F. Monrose, and M. Reiter. 2004. On user choice in graphical password schemes. In *Proc. of USENIX SSYM 2004*. USENIX Assoc.
[9] O. J. Dunn. 1964. Multiple comparisons using rank sums. *Technometrics* 6 (1964), 241-252.
[10] R. Shay, S. Komanduri, P. Kelley, P. Leon, M. Mazurek, L. Bauer, N. Christin, and L. Cranor. 2010. Encountering stronger password requirements: user attitudes and behaviors. In *Proc. of ACM SOUPS 2010*. ACM Press, article 2.
[11] S. Komanduri, R. Shay, P. Kelley, M. Mazurek, L. Bauer, N. Christin, L. Cranor, and S. Egelman. 2011. Of passwords and people: Measuring the effect of password-composition policies. In *Proc. of ACM CHI 2011*. ACM Press, 2595-

2604.

[12] A. Paivio, and K. Csapo. 1973. Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology* 5, 2 (1973), 176-206.

[13] D. L. Nelson, U. S. Reed, and J. R. Walling. 1976. Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning & Memory* 2 (1976), 523-528.

[14] P. Verma. 2012. icAuth: Image-color based authentication system. *In Proc. of ACM IUI 2012.* ACM Press, 329-330.

[15] C. Fidas, H. Hussmann, M. Belk, G. Samaras. 2015. iHIP: Towards a user centric individual human interaction proof framework. In *Proc. of the ACM Conference on Human Factors in Computing Systems (CHI 2015).* ACM Press, 2235-2240.

[16] J. Nicholson, L. Coventry, and P. Briggs. 2013. Age-related performance issues for PIN and face-based authentication systems. In *Proc. of ACM Conference on Human Factors in Computing Systems (CHI 2013).* ACM Press, 323-332.

[17] E. Stobert, and R. Biddle. 2013. Memory retrieval and graphical passwords. In *Proc. of the ACM Symposium on Usable Privacy and Security (SOUPS 2013).* ACM Press, article 15, 14 pages.

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

[19] H. Gao, X. Guo, X. Chen, L. Wang, and X. Liu. 2008. YAGP: Yet Another Graphical Password Strategy. In *Proc. of the Conference on Computer Security Applications.* IEEE Computer Society, 121-129.

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

[21] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. 2005. Authentication using graphical passwords: Effects of tolerance and image choice. In *Proc. of the ACM Symposium on Usable Privacy and Security (SOUPS 2005).* ACM Press, 1-12.

[22] S. Chiasson, A. Forget, R. Biddle, and P. van Oorschot. 2008. Influencing users towards better passwords: Persuasive cued click-points. In *Proc. of the BCS Conference on People and Computers.* British Computer Society, 121-130.

[23] M. Belk, C. Fidas, P. Germanakos, G. Samaras. 2013. Security for diversity: Studying the effects of verbal and imagery processes on user authentication mechanisms. In *Proc. of the IFIP TC13 Conference on Human-Computer Interaction (INTERACT 2013).* Springer-Verlag, 442-459.

[24] R. Schlöglhofer, and J. Sametinger. 2012. Secure and usable authentication on mobile devices. In *Proc. of the ACM Conference on Advances in Mobile Computing & Multimedia (MoMM 2012).* ACM Press, 257-262.

[25] Katsini, C., Fidas, C., Belk, M., Avouris, N., & Samaras, G. (2017). Influences of users' cognitive strategies on graphical password composition. In Extended Abstracts of the ACM Conference on Human Factors in Computing Systems (CHI 2017), ACM Press, 2698-2705.

[26] M. Belk, C. Fidas, P. Germanakos, and G. Samaras. 2017. The interplay between humans, technology and user authentication: a cognitive processing perspective. *Computers in Human Behavior.* (to appear).