Understanding Insider Attacks in Personalized Picture Password Schemes

Argyris Constantinides^{1,2}, Marios Belk^{3,1}, Christos Fidas⁴, Andreas Pitsillides¹

¹ Department of Computer Science, University of Cyprus, Nicosia, Cyprus aconst12@cs.ucy.ac.cy, cspitsil@cs.ucy.ac.cy ² Cognitive UX LTD, Nicosia, Cyprus argyris@cognitiveux.com ³ Cognitive UX GmbH, Heidelberg, Germany belk@cognitiveux.de ⁴ Department of Electrical and Computer Engineering, University of Patras, Patras, Greece

fidas@upatras.gr

Abstract. Picture passwords, which require users to complete a picture-based task to login, are increasingly being embraced by researchers as they offer a better tradeoff between security and memorability. Recent works proposed the use of personalized familiar pictures, which are bootstrapped to the users' prior sociocultural activities and experiences. However, such personalized approaches might entail guessing vulnerabilities by people close to the user (*e.g.*, family members, acquaintances) with whom they share common experiences within the depicted familiar sceneries. To shed light on this aspect, we conducted a controlled in-lab eye-tracking user study (n=18) focusing on human attack vulnerabilities among people sharing common sociocultural experiences. Results revealed that insider attackers, who share common experiences with the legitimate users, can easily identify regions of their selected secrets. The extra knowledge possessed by people close to the user was also reflected on their visual behavior during the human attack phase. Such findings can drive the design of assistive security mechanisms within personalized picture password schemes.

Keywords: Picture Passwords, Security, Eye-Tracking, User Study.

1 Introduction

Computer security systems encompass concepts and methods for the protection of sensitive information. In this context, user authentication is an essential security task performed daily by millions of users. Traditional solutions employ text-based passwords, which require users to memorize a sequence of alphanumeric characters. However, memorizing strong text-based passwords results in increased cognitive load and often leads to poor usability and limited security [1]. To offer a better tradeoff between security and usability, prior works proposed various picture password schemes [3], which require users to complete a picture-based task to authenticate.

An important interface design factor that affects both the security [4, 5, 7] and usability [9-11, 13] of picture password schemes is the background picture(s) used [4, 5, 15]. Several studies have investigated various picture content types, which can be broadly categorized as generic (i.e., not familiar to the users, e.g., stock, landscapes, abstract, etc.) and personal (i.e., highly familiar to the users, e.g., depicting scenes, people, or objects highly familiar to users), and reported their effects on the security and memorability of the user-chosen picture passwords. In particular, the use of generic pictures impacts negatively both the security and memorability of picture passwords [4, 15], while the use of personal pictures impacts negatively the security but leads to increased memorability of picture passwords [15, 16]. In an attempt to achieve a better tradeoff between security and memorability, more recent works investigated and proposed the use of personalized familiar pictures, which are bootstrapped to the users' prior sociocultural activities, experiences and explicit memories [18, 19, 34, 35], revealing a positive impact on the security without hampering the memorability of picture passwords [18, 19]. Nevertheless, such personalized picture delivery approaches might be susceptible to attacks performed by insiders [21, 22] (i.e., people close to the user, such as, family members, acquaintances) with whom they share common experiences within the depicted familiar pictures.

Given that the process of picture password authentication is a visual search task, eye-tracking technology could be used to shed light on how a legitimate user's gaze path relates to an insider attacker's gaze path, and eventually infer whether the person attempting to login is a legitimate user or an insider attacker close to the legitimate user. While attempts have been made towards improving and estimating the security of authentication schemes using eye-tracking technology [15, 23-25, 27, 28], to the best of the authors' knowledge, no research attempts have been made to estimate the legitimacy of the user authenticating in a personalized picture password scheme that leverages on users' prior sociocultural activities, experiences and explicit memories. This work presents the initial findings of applying an eye gaze-driven metric for unobtrusively estimating the legitimacy of the users' eye gaze behavior during login.

2 Related Work

2.1 Picture Content in Picture Passwords

Prior works investigated the use of picture semantics and their effects on the security and memorability of user-chosen picture passwords. Pictures can be broadly categorized as generic (*i.e.*, not directly relevant nor familiar to the users, *e.g.*, abstract, nature, landscapes, etc.) or personal (*i.e.*, directly relevant and highly familiar to the users, *e.g.*, depicting people, objects, or scenes highly personal to users). The use of generic picture content has a negative impact on both the security and memorability of the user-chosen passwords. Studies in [4, 15] revealed that various generic pictures are susceptible to hotspots (*i.e.*, certain points on a picture that are more likely to be selected by users), which leads to the creation of predictable passwords that are prone to automated attacks [30]. From the memorability perspective, generic picture content leads to decreased memorability since users experience difficulties in creating strong connections between their episodic memories and the depicted content [16, 31]. The use of personal picture content also impacts the security and memorability of picture passwords. From the security perspective, the use of pictures that are familiar to the user increases the likelihood of certain areas on the picture to be chosen as part of the password [15]. However, from the memorability perspective, the use of personal pictures leads to increased memorability possibly due to familiarity of users with the depicted picture content [16]. More recent works investigated the use of personalized familiar pictures, which are bootstrapped to the users' prior sociocultural activities, experiences and explicit memories, revealing a positive impact on the security without hampering the memorability of picture passwords [18, 19].

2.2 Eye Gaze in User Authentication

Eye-tracking technology has been widely used in the context of user authentication. Darrell and Duchowski [23] proposed a rotary interface for gaze-based PIN code entry during user authentication, while Bulling et al. [15] proposed to hide potential picture hotspots using saliency maps. A study conducted by Sluganovic et al. [27] revealed that the reflexive physiological behavior of human eyes can be used to build fast and reliable biometric authentication systems. More recent works employed eye gaze data for predicting image content familiarity in picture password schemes [36], as well as for understanding how individuals make their picture password selections [26]. Moreover, works in [24, 28] proposed eye gaze-driven security metrics for estimating the strength of picture passwords.

3 Eye-Tracking Study

Bearing in mind that when using the personalized picture password approach, the password selections are based on the users' existing sociocultural experiences, it is probable that such personalized approaches might be susceptible to attacks performed by insiders [21, 22] (*i.e.*, people close to the user, such as, family members, acquaint-ances) with whom they share common experiences. In order to shed light on this aspect, we conducted an in-lab eye-tracking human attack study focusing on attacks performed by insiders among people sharing common sociocultural experiences. Each session of the study embraced pairs of participants that were closely related (*e.g.*, friends, couples, relatives, etc.) and who shared common experiences. In each session, both participants were first requested to create a picture password, and then each participant was requested to guess the password selections of the other participant.

3.1 Research Question

RQ. Is there a significant difference in users' visual behavior between legitimate users and insider attackers when authenticating in a picture password scheme that employs personalized picture content?



Fig. 1. A subset of pictures used in the human attack study illustrating sceneries in which participants share common experiences.

3.2 Study Instruments and Metrics

Picture Password Authentication Scheme. We implemented a Web-based picture password scheme, similar to Windows 10TM PGA [32], in which users can create picture passwords consisting of three gestures (any combination of taps, lines, and circles). The picture is divided in a grid containing 100 segments on the longest side and scaled accordingly on the shortest side. The mechanism allows for a tolerance distance in terms of the coordinates on the grid (36 segments around each initial selected segment are acceptable¹ [13], thus, building a circle of 3 segments radius). This tolerance allows for better accuracy of users' selections during login. However, there is no tolerance regarding ordering, type, and directionality of the gestures.

Picture Content. To control participants' sociocultural familiarity with the picture semantics and thus investigate the research question, we adjusted the picture semantics to reflect participants' shared, individual and common sociocultural experiences from their daily life context (*i.e.*, working places in the case of colleagues, café/bars in which couples or close friends usually hang out, etc.), as depicted in **Figure 1**. For doing so, prior to the study, we asked each pair of participants to provide a set of pictures from places in which they share common experiences. To avoid bias effects, we did not inform the participants about the reason they were providing us the pictures until the end of the study. The sets of pictures were based on existing research that has shown that users tend to select pictures illustrating sceneries [5, 8, 33].

Considering that the number of hotspots and the picture complexity affect the password strength [6, 13], we chose pictures of similar number of hotspots and complexity. For doing so, we followed a semi-automated approach to detect the hotspots regions through a combination of computer vision techniques for object detection^{2,3} and saliency filters [12]. Furthermore, we assessed the equivalence of the two picture sets by calculating the picture complexity using entropy estimators [29].

¹ MicrosoftTM Picture Password blog - bit.ly/2SajCDO

² Google Cloud Vision - bit.ly/21xSsUV

³ Tensorflow - bit.ly/IMWEhkH

Equipment and Eye Gaze Metrics. An All-in-One HP computer with a 24" monitor was used (1920x1080 pixels, 16:9 aspect ratio). To capture eye movements, we used Gazepoint GP3⁴ eye tracker, which captures data at 60Hz and was calibrated following the manufacturer's guidelines. No equipment was attached to the participants. Following existing approaches for capturing the variability of users' eye movement characteristics within picture password schemes [24, 28], we relied on the gaze transition entropy proposed by Krejtz et al. [14]. In particular, we estimated the stationary entropy H_s , which captures the distribution of fixations over the stimulus (*i.e.*, areas of interest (AOIs) in which the eye-tracking metrics are applied). Greater values of H_s occur when the visual attention is distributed more equally among AOIs, while lower values of H_s indicate that fixations tend to be concentrated on certain AOIs. Stationary entropy H_s was conducted using Shannon's entropy equation:

$$H_{s}(X) = \sum_{i=1}^{N} p_{i} * \log_{2}\left(\frac{1}{p_{i}}\right)$$
(1)

where *X* is the set of fixations for each user, *N* is the number of the available AOIs, and *p* is the probability of a user to fixate on AOI *i*. Considering that fixation duration correlates with cognitive processing [17, 20], and that users who exhibit longer fixations on AOIs tend to select them [2], the probability p_i is computed as follows:

$$p_i = \frac{d_i}{N}, \sum_{i=1}^{N} = 1$$
 (2)

where d_i is the distribution of p_i across N, representing the total fixation duration on AOI *i*. By applying equation (2) to equation (1), the entropy of fixations is computed as follows:

$$H_{s}(X) = \sum_{i=1}^{N} \frac{d_{i}}{N} * \log_{2}(N)$$
(3)

N=3: the picture is divided into three vertical AOIs [14].

3.3 Sampling and Procedure

Participants. A total of 18 individuals (9 females) participated in the study, ranging in age between 25-60 years old (m=41.43, sd=11.88). Since the purpose of this study was to understand whether there are differences between legitimate users' and insiders' visual behavior, we intentionally recruited pairs of participants that are close to each other (3 couples, 3 pairs of close friends, 3 pairs of colleagues). To increase the internal validity of the study, we recruited participants that had no prior experience with picture password authentication mechanisms, as assessed by a post-study interview in order to exclude any participants with prior knowledge on picture passwords.

Experimental Design and Procedure. Participation in the study was anonymized to ensure privacy compliance according to the EU General Data Protection Regulation. Participants were informed that the collected data will be analyzed for research pur-

⁴ GP3 Eye Tracker - bit.ly/3g8rDWq

poses only. Also, we took all the necessary measures against Covid-19 to ensure the participants' safety. The study was conducted in a quiet lab room with only the researcher present and was split in two phases as follows: i) Phase A - Password Creation: Each pair of closely related participants (e.g., friends, couples, colleagues, etc.) visited the laboratory in a pre-scheduled time within the Covid-19 safety regulations. First, the eye calibration process started, and then participants were requested independently to create a picture password by drawing 3 gestures on the picture (any combination of taps, lines, circles) in order to access an online service. To avoid bias effects during *Phase B* (Human Guessing Attack), each participant created a password on a different picture that depicted places in which they share common experiences; ii) Phase B – Human Guessing Attack: We switched the picture of the pairs and each participant was requested to guess the other participant's secrets by indicating 3 areas (i.e., 3 (x, y) segments on the grid) on the picture for which they believe that the other participant made their selections around them. Also, we adopted the think-aloud protocol aiming to elicit whether the rationale behind the attacker's selections is related to the shared memories and experiences with the other participant from the same pair. Finally, both participants completed a questionnaire on demographics.

3.4 Analysis of Results

Visual behavior differences between legitimate users and insider attackers during login. To investigate our RQ, we ran a paired-samples t-test with the entropy from equation (3) as the dependent variable tested under two different conditions (*i.e.*, during legitimate user login and during insider attacker login). The analysis revealed that insider attackers exhibited higher stationary entropy H_s (8.70 ± 2.02 bits) than legitimate users (1.55 ± 0.78 bits), a statistically significant difference of 7.15 ± 1.24 bits (95% CI, 3.35 to 10.94 bits), t(8)=4.04, p=.001. Figure 2 shows the stationary entropy H_s of both legitimate users and insider attackers.

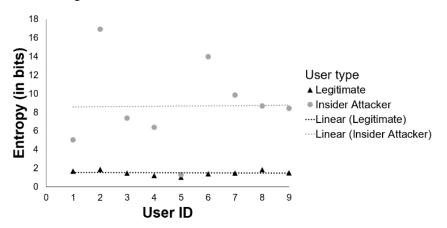


Fig. 2. Stationary entropy H_s of both legitimate users and insider attackers.

6

Revealing the insider attacker's strategy when guessing a picture password. To get further insights about the approach followed by the insider attackers, at the end of *Phase B (Human Guessing Attack)* we asked each participant to show us the picture password selections they made on the screen, and we labelled them as either *H* (Hotspot), *E* (Experience spot; provided by the user), or *O* (Other; non-hotspot, non-experience spot). In order to understand the similarities in terms of areas correctly matched on the picture grid between legitimate users' password selections and insider attackers' guessing selections, we disregarded the order and the type of the gestures and rather focused on the positions of the password selections as follows: For circles, we disregarded the radius and the directionality, and kept only the center of the circle as a (x, y) segment, while for lines, we considered only the (x, y) segment of the starting point of the line. **Table 1** summarizes the approach followed by the legitimate users and the areas correctly matched by the insider attackers.

 Table 1. Summarization of the approach followed by the legitimate users and areas correctly matched by the insider attackers. H denotes hotspot selection; E denotes experience spot selection; and O denotes other (non-hotspot, non-experience spot) selection. The insider attackers' areas matched are highlighted in gray color.

	Gesture 1	Gesture 2	Gesture 3	# of insider attacker's areas matched (out of 3)	
	Н	Е	0	1	
	Н	0	О	0	
	Н	Ο	0	0	
	0	Ο	Н	0	
	Е	Н	0	1	
	0	Н	0	0	
	0	0	0	0	
	0	0	Е	1	
	Н	Н	0	1	
	Е	Н	Е	2	
	0	0	Е	0	
	Е	0	0	1	
	Е	Е	Н	2	
	Е	0	0	1	
	Е	Е	0	2	
	Е	Н	Н	1	
	Н	Н	Н	1 1	
	Е	Н	Н		
				# areas	# insider
Labels				matched	attackers
H	5	7	5	0	6
E	8	3	3	1	9
0	5	8	10	2	3
				3	0

4 Conclusions and Future Work

In this work, we conducted a controlled in-lab eye-tracking user study focusing on human attack vulnerabilities among people sharing common sociocultural experiences within personalized picture password schemes. Results revealed that insider attackers who share common experiences with the legitimate users can easily identify regions of their selected secrets, as shown in **Table 1**. The extra knowledge possessed by people who are close to the legitimate user was also reflected on their visual behavior during the human guessing attack phase. In particular, we found that the insider attackers exhibited higher stationary entropy H_s than the legitimate users. As stated previously, greater values of H_s occur when the visual attention is distributed more equally among AOIs, which might occur in cases of insider attackers who use extra knowledge to guess the user's picture password, while lower values of H_s indicate that fixations tend to be concentrated on certain AOIs, which might occur in cases of legitimate users who know their passwords and make fixations on certain AOIs.

Such findings can be used for the estimation of the legitimacy of the user authenticating in a personalized picture password scheme that leverages on users' prior sociocultural activities, experiences and explicit memories, and drive the design of assistive security mechanisms. We envision that such visual behavior differences in personalized picture password schemes can be used for the creation of multi-class classifiers for predicting the legitimacy of the individual during authentication (*i.e.*, legitimate user, insider attacker, other attacker). Such a classifier will notify the legitimate users about the type of attacker attempting to login to their account, as well as limit the account lockout threshold accordingly (*e.g.*, apply a more strict policy in cases of insider attackers). Expansion of our research will consider the feasibility of building such a multi-class classifier for predicting the legitimacy of the user authenticating, as well as conducting additional user studies to triangulate findings with diverse user communities and sociocultural experiences.

Acknowledgements

The work has been partially supported by the EU Horizon 2020 Grant 826278 "Securing Medical Data in Smart Patient-Centric Healthcare Systems" (Serums), the Research and Innovation Foundation (Project DiversePass: COMPLEMENTARY/0916/0182), and the European project TRUSTID - Intelligent and Continuous Online Student Identity Management for Improving Security and Trust in European Higher Education Institutions (Grant Agreement No: 2020-1-EL01-KA226-HE-094869), which is funded by the European Commission within the Erasmus+ 2020 Programme and the Greek State Scholarships Foundation I.K.Y.

8

References

- Sasse, M.A., Brostoff, S., & Weirich, D. (2001). Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. BT technology journal, 19(3), 122-131.
- Raptis, G.E., Katsini, C., Belk, M., Fidas, C., Samaras, G., & Avouris, N. (2017). Using Eye Gaze Data and Visual Activities to Infer Human Cognitive Styles: Method and Feasibility Studies. ACM UMAP '17, ACM Press, 164-173.
- Biddle, R., Chiasson, S., & Van Oorschot, P.C. (2012). Graphical passwords: Learning from the first twelve years. ACM Comput. Surv. 44, 4, Article 19, 41 pages.
- 4. Thorpe, J. and van Oorschot, P.C. (2007). Human-seeded attacks and exploiting hot-spots in graphical passwords. USENIX Security Symposium (SS'07), Article 8, 1-16.
- Alt, F., Schneegass, S., Shirazi, A.S., Hassib, M., & Bulling, A. (2015). Graphical Passwords in the Wild: Understanding How Users Choose Pictures and Passwords in Imagebased Authentication Schemes. ACM MobileHCI '15, ACM Press, 316-322.
- Wiedenbeck, S., Waters, J., Birget, J.C., Brodskiy, A., & Memon, N. (2005). Authentication using graphical passwords: Effects of tolerance and image choice. Symposium on Usable privacy and security (SOUPS '05). ACM Press, 1-12.
- Zhao, Z., Ahn, G.J., Seo, J.J., & Hu, H. (2013). On the security of picture gesture authentication. USENIX conference on Security (SEC '13), 383-398.
- Zhao, Z., Ahn, G.J., & Hu, H. (2015). Picture gesture authentication: Empirical analysis, automated attacks, and scheme evaluation. ACM TISSEC '15, 17(4), pp.1-37.
- Mihajlov, M., Jerman-Blažič, B., & Ciunova Shuleska, A. (2016). Why that Picture? Discovering Password Properties in Recognition-Based Graphical Authentication. Elsevier IJHCS, 32(12), 975-988.
- Mihajlov, M. and Jerman-Blažič, B. (2011). On designing usable and secure recognitionbased graphical authentication mechanisms. Interact. Comput. 23, 6, 582-593.
- Everitt, K.M., Bragin, T., Fogarty, J., & Kohno, T. (2009). A comprehensive study of frequency, interference, and training of multiple graphical passwords. ACM SIGCHI '09, ACM Press, 889–898.
- Perazzi, F., Krähenbühl, P., Pritch, Y., & Hornung, A. (2012). Saliency filters: Contrast based filtering for salient region detection. IEEE conference on computer vision and pattern recognition. IEEE, 733-740.
- Katsini, C., Fidas, C., Raptis, G.E., Belk, M., Samaras, G., & Avouris, N. (2018). Influences of Human Cognition and Visual Behavior on Password Strength during Picture Password Composition. ACM CHI '18, ACM Press, Paper 87, 1-14.
- 14. Krejtz, K. et al. (2015). Gaze Transition Entropy. ACM TAP '15, 13, 1, 1-20.
- Bulling, A., Alt, F., & Schmidt, A. (2012). Increasing the security of gaze-based cuedrecall graphical passwords using saliency masks. ACM SIGCHI '12, ACM Press, 3011-3020.
- Tullis, T.S. and Tedesco, D.P. (2005). Using personal photos as pictorial passwords. ACM CHI EA '05, ACM Press, 1841-1844.
- Fidas, C., Belk, M., Hadjidemetriou, G., Pitsillides, A. (2019). Influences of mixed reality and human cognition on picture passwords: An eye tracking study. IFIP TC13 Human-Computer Interaction (INTERACT 2019), Springer-Verlag, 304-313
- Constantinides, A., Fidas, C., Belk, M., Pietron, A., Han, T., & Pitsillides, A. (2021). From hot-spots towards experience-spots: Leveraging on users' sociocultural experiences to enhance security in cued-recall graphical authentication. Elsevier IJHCS 149 (2021): 102602

- Constantinides, A., Pietron, A., Belk, M., Fidas, C., Han, T., & Pitsillides, A. (2020). A Cross-cultural Perspective for Personalizing Picture Passwords. ACM UMAP '20, ACM Press, 43-52.
- David E. Irwin. (2004). Fixation Location and Fixation Duration as Indices of Cognitive Processing. In J. M. Henderson & F. Ferreira (Eds.), The interface of language, vision, and action: Eye movements and the visual world (p. 105–133). Psychology Press.
- Aljahdali, H.M. and Poet, R. (2014). Educated Guessing Attacks on Culturally Familiar Graphical Passwords Using Personal Information on Social Networks. ACM SIN '14, ACM Press, 272–278.
- Muslukhov, I., Boshmaf, Y., Kuo, C., Lester, J. and Beznosov, K. (2013). Know your enemy: the risk of unauthorized access in smartphones by insiders. ACM MobileHCI '13, ACM Press, 271-280.
- 23. Best, D.S. and Duchowski, A.T. (2016). A Rotary Dial for Gaze-based PIN Entry. ACM ETRA '16, ACM Press, 69-76.
- Katsini, C., Raptis, G.E., Fidas, C., & Avouris, N. (2018). Towards gaze-based quantification of the security of graphical authentication schemes. ACM ETRA '18, ACM Press, Article 17, 5 pages.
- De Luca, A., Denzel, M., & Hussmann, H. (2009). Look into my eyes!: can you guess my password?. ACM SOUPS '09, ACM Press, Article 7, 12 pages.
- Constantinides, A., Fidas, C., Belk, M., & Pitsillides, A. (2019). "I Recall this Picture": Understanding Picture Password Selections based on Users' Sociocultural Experiences. In IEEE/WIC/ACM WI '19, ACM Press, 408–412.
- Sluganovic, I., Roeschlin, M., Rasmussen, K.B., & Martinovic, I. (2016). Using Reflexive Eye Movements for Fast Challenge-Response Authentication. ACM SIGSAC CCS '16, ACM Press, 1056-1067.
- Constantinides, A., Belk, M., Fidas, C., & Pitsillides, A. (2020). An eye gaze-driven metric for estimating the strength of graphical passwords based on image hotspots. ACM IUI '20, ACM Press, 33-37.
- Cardaci, M., Di Gesù, V., Petrou, M., &Tabacchi, M.E. (2009). A fuzzy approach to the evaluation of image complexity. Fuzzy Sets and Systems, 160(10), pp.1474-1484.
- Salehi-Abari, A., Thorpe, J., & Van Oorschot, P. C. (2008). On purely automated attacks and click-based graphical passwords. IEEE ACSAC '08, pp. 111-120.
- Renaud, K. (2009). On user involvement in production of images used in visual authentication. Journal of Visual Languages & Computing, 20(1), pp.1-15.
- Johnson, J.J., Seixeiro, S., Pace, Z., van der Bogert, G., Gilmour, S., Siebens, L., & Tubbs, K. (2014). Picture Gesture Authentication. Retrieved from https://www.google.com/patents/US8910253, last accessed 2021/06/10
- Dunphy, P. and Yan, J. (2007). Do background images improve "draw a secret" graphical passwords?. ACM CCS '07, ACM Press, pp. 36-47.
- Constantinides, A., Belk, M., Fidas, C., & Samaras, G. (2018). On Cultural-centered Graphical Passwords: Leveraging on Users' Cultural Experiences for Improving Password Memorability. ACM UMAP '18, ACM Press, 245–249.
- Constantinides, A., Fidas, C., Belk, M., & Samaras, G. (2018). On sociocultural-centered graphical passwords: an initial framework. ACM MobileHCI '18 Adjunct, ACM Press, 277–284.
- Constantinides, A., Belk, M., Fidas, C., & Pitsillides, A. (2019). On the Accuracy of Eye Gaze-driven Classifiers for Predicting Image Content Familiarity in Graphical Passwords. ACM UMAP '19, ACM Press, 201–205.

10