Shoulder Surfing on Mobile Authentication: Perception vis-a-vis Performance from the Attacker's Perspective

Kanlun Wang
Department of Business Information Systems and Operations Management
The University of North Carolina at Charlotte
Charlotte, USA
kwang17@charlotte.edu

Lina Zhou
Department of Business Information Systems and Operations Management
The University of North Carolina at Charlotte
Charlotte, USA
lzhou8@charlotte.edu

Dongsong Zhang
Department of Business Information Systems and Operations Management
The University of North Carolina at Charlotte
Charlotte, USA
dzhang15@charlotte.edu

Jianwei Lai
School of Information Technology
Illinois State University
Normal, USA
jlai12@ilstu.edu

2023 IEEE International Conference on Intelligence and Security Informatics (ISI) | 979-8-3503-3773-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISI58743.2023.10297219

***Abstract*—Shoulder-surfing studies in the context of mobile user authentication have focused on evaluating the attackers' performance, yet have paid much less attention to their perception of the shoulder-surfing process. Whether and how the shoulder-surfing setting might affect the attackers' perception remains under-explored. This study aims to investigate the perception of shoulder surfers with two different password-based mobile user authentication methods and three different observation angles. Moreover, this work examines the relationship between the attackers' perception and performance in shoulder surfing and the possible moderating effect of the authentication method for the first time. Based on the data collected from an online experiment, our analysis results reveal the effects of authentication methods and observation angles on the attackers' perception in terms of cognitive workload, observation clarity, and repetitive learning advantage. In addition, the results also show that the relationship between the attackers' cognitive workload and performance in shoulder surfing varies with the mobile user authentication method. Our findings not only deepen the understanding of shoulder-surfing attacks from an attacker's perspective, but also facilitate developing countermeasures for shoulder-surfing attacks.**

Keywords—shoulder surfing attacks, mobile user authentication, perception, performance, cognitive workload

I. INTRODUCTION

Shoulder surfing is a common security attack that intends to obtain a victim's personal information via direct or indirect observation without a victim's knowledge [1]. Shoulder surfing over a mobile user authentication session can expose the target user to the risk of losing his/her login credential such as a password and provide the attacker an opportunity to gain access to the mobile device. To gain a comprehensive understanding of the effectiveness and/or vulnerabilities of mobile user authentication methods against shoulder-surfing attacks, it is important to consider both subjective (e.g., user perception) and objective measures (e.g., shoulder-surfing performance) from the attacker's perspective [2]. For instance, a mobile user authentication method may give an attacker a false perception of a 'successful' attack, while in fact, it was a failure. Nevertheless, the relationship between user perception and user performance measures remains under-explored, particularly when considering different shoulder-surfing settings (e.g., authentication methods and observation angles).

The subjective measures used in shoulder-surfing research are commonly focused on the perspective of the victim or target user such as the perceived threat of shoulder-surfing attacks (e.g., [3], [4]). The few studies from the attacker's perspective have considered the perceived confidence in password identification (e.g., [5], [6]), which is outcome-oriented. However, they have not explored the attackers' perception of the shoulder-surfing process. To this end, cognitive load (the working memory load experienced when performing a specific task) becomes relevant [7]. For instance, despite that an attacker achieved similar levels of performance in shoulder surfing with two different mobile user authentication methods *A* and *B*, the attacker exerted a lower level of cognitive effort while attacking using method *A* than *B*; and accordingly, method *A* is an easier target for attack than *B*. To measure the cognitive load associated with completing a specific work, NASA Task Load Index (TLX) [7] is a comprehensive scale that measures multidimensional demands in cognitive workload, which has been commonly used to evaluate the usability of mobile user authentication methods (e.g., [6], [8]), but not yet been used to evaluate shoulder-surfing attacks.

Prior shoulder-surfing studies employed a wide range of objective measures to evaluate the actual performance of password identification through shoulder surfing, including characteristics-based metrics (e.g., the percentage of correctly identified characters [5], [9], and the percentage of correctly identified character positions [10]), and distance-based metrics [22]. The inconsistent employment of different objective measures makes the comparison of the shoulder-surfing performance between different authentication methods extremely difficult, if not entirely impossible. In addition, empirical shoulder-surfing studies have mainly focused on the comparisons between traditional authentication methods, such as PINs vs. pattern locks [11] and textual passwords vs. graphical passwords [12]. As an emerging augmentation to traditional authentication methods, behavioral biometrics leverages an individual's unique behavioral characteristics [13]. Among the various behavioral biometrics that have been incorporated into mobile user authentication methods, keystrokes and touch gestures have received the most attention [14]. The former is based on a user's tapping behaviors on the soft keyboard of a mobile device while using it to enter a password [15], while the latter

is based on a user's touch gestures (e.g., swiping and stroking) on the touch screen of a mobile device [16]. Yet, empirical investigations of shoulder-surfing attacks with the keystroke- and touch gesture-based authentication methods remain scarce.

To fill the aforementioned research gaps, this study aims to investigate attackers' perception (subjective measures) and their relationship with performance (objective measures) in shoulder-surfing attacks over keystroke- and touch gesture-based authentication methods under different settings. Specifically, this research answers the following questions by conducting an online experiment: RQ1) How do shoulder-surfing settings affect attackers' perception of the shoulder-surfing process, if any? RQ2) Is there a relationship between attackers' cognitive workload and their performance in shoulder surfing? If so, does the relationship vary with the mobile user authentication method?

II. Related Work

In comparison with shoulder surfing through direct observation, indirect observation via video recordings poses a greater risk to mobile user authentication. This is due to the ability of attackers to review the authentication credentials multiple times [1], significantly increasing the likelihood of a successful attack. In this study, our focus is on shoulder surfing through indirect observation.

To counter shoulder surfing, a host of studies have proposed novel designs of mobile user authentication methods, including password-based methods (e.g., [3], [10], [13]). In addition, different design strategies have been proposed for shoulder-surfing countermeasures, including visibility reduction, action segmentation, knowledge transformation, and characteristics verification [17]. For instance, a touch gesture-based authentication method [13] offers keypress-free interaction that reduces the saliency of viewable areas on the touch screen of a mobile device; supports using two or fewer directional touch gestures to enter characters that temporally or spatially divides an input action into sub-actions that will be performed sequentially or concurrently; transforms consecutive touch gestures to characters to increase the perplexity of password identification; and incorporates both password matching and behavioral biometrics for authentication.

Shoulder-surfing performance can be evaluated in terms of both objective and/or subject measures. Sample measures are summarized in Table I.

TABLE I. Subjective and Objective Measures from Shoulder-surfing Studies

Category	Perspective	Sample Measures
Subjective	Victim	Perceived threat [3], [4]
	Attacker	Confidence in password identification [5], [6]
	Attacker	NASA TLX [18]
	Both	Positive and negative feelings [19]
Objective	Attacker	Success rate [11], [20], [21]
	Attacker	Correctly identified characters [9], [17]
	Attacker	Correctly identified character positions [10]
	Attacker	Levenshtein Distance [17], [22]

A. Subjective Measures

Subjective measures have been deployed to assess participants' perception of shoulder surfing mainly from a victim's perspective (e.g., the perceived threat of shoulder-surfing attack [3], [4]). Although a few studies (e.g., [5], [6]) have considered the subjective measures of shoulder surfing from an attacker's perspective, they only measured perceived confidence in password identification. Some others reported that attackers have negative feelings about shoulder surfing, with a few exceptions [19]. However, whether and how those negative feelings are associated with the attackers' subsequent performance in shoulder surfing remains unclear. To the best of our knowledge, only one study [18] has considered the attackers' cognitive workload in shoulder surfing textual content on a mobile device with different visual privacy protection methods. While both mobile user authentication and visual privacy methods share the common goal of safeguarding user privacy and enhancing security on mobile devices, the design of their methods differs significantly from that of ours. In addition, mobile user authentication primarily focuses on the process of verifying user identity, while visual privacy methods are centered around safeguarding the content displayed on the screen of a mobile device. Therefore, attackers' cognitive workload is associated with shoulder surfing performance in mobile user authentication remains significantly under-explored.

B. Objective Measures

Success rate (e.g., [11], [17], [20], [21]) is the most common measure of shoulder-surfing performance. In view that the dichotomous representation might be too coarse to measure the performance of shoulder-surfing attacks, ranging from partially or totally correct, fine-grained metrics have been developed, including the percentage of correctly identified characters [9], the percentage of correctly identified character positions (right spot) [10], Levenshtein distance between the identified password and the actual password [22], and a decrease in password unpredictability (guessing order) [23].

III. Method

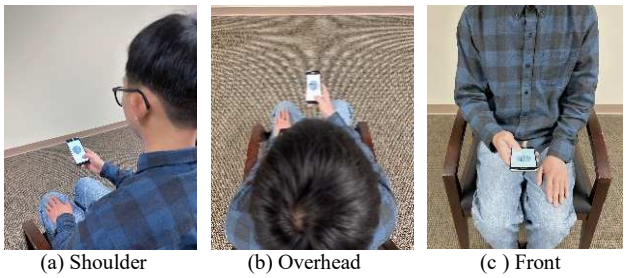
We collected data by conducting an online experiment to answer the research questions. The study was approved by the Institutional Review Board of the first author's institute.

A. Shoulder-surfing Settings

Authentication methods: Textual passwords remain among the most common types of passwords for mobile user authentication [24]. We selected two types of authentication methods supporting textual passwords. One is a keystroke-based authentication method that allows a user to enter a password by tapping the corresponding letters/keys on a soft keyboard (e.g., QWERTY keyboard) of a mobile device (e.g., [3], [25]). Another is a touch gesture-based authentication method that allows a user to enter a password via various touch gestures on the mobile screen [16]. Specifically, we selected the ThumbStroke keypad [13], where a user performs one or two consecutive directional touch gestures (i.e., up, down, left, right, upper-left, lower-left, upper-right, and lower-right) to enter a password character.

Observation angle: An adversary may observe a password entry from different angles and distances. We selected three observation angles (Fig. 1), namely *shoulder*, where an

adversary is positioned behind the target user's left shoulder, approximately 60 centimeters away from the mobile screen. *overhead* is right over the target user's head with an observation distance of 95 centimeters, and *front* is facing the target user straight with an observation distance of 45 centimeters. Among them, the front angle has never been examined in previous shoulder-surfing studies. The orders of shoulder surfing settings were randomized across different participants.



(a) Shoulder (b) Overhead (c) Front

Fig. 1. Shoulder-surfing Observation Angles

B. Password Authentication Setup

The passwords used in the mobile authentication sessions were 8-character long. The choice of password length was motivated by the information processing theory [26] that most people can only keep 7 ± 2 elements in their short-term memory. We prepared 6 different non-dictionary-word passwords and randomly assigned one password to each shoulder-surfing setting. The password entry sessions were video-recorded in advance by an expert user to eliminate possible confounding factors introduced by using different participants. In addition, the expert user was instructed to use a Samsung Galaxy S6 smartphone with a 5.1" touchscreen to simulate an average user's behavior during password authentication.

C. Tasks and Procedure

The pre-experiment procedure consisted of password entry training, a familiarity test, and a pre-experiment survey. Password entry training aimed to get participants familiarized with the two different keyboards. Assuming that all the participants were familiar with the QWERTY keyboard, the training of the keystroke-based authentication method asked participants to recall adjacent letters of three different letters on a QWERTY keyboard. The training of the touch gesture-based authentication method required the participants to watch a tutorial and demonstration video first and then complete two rounds of practice with a web-based prototype. In each round of the practice, the participants were required to enter 26 letters one by one into their designated empty text fields accurately. The familiarity test was focused on the memorization of all letters in each of the keyboards used by the two authentication methods separately. Those participants who could not memorize more than 10 letters and those who were not able to recall the touch gestures of a letter within three attempts would be disqualified. The pre-experiment survey collected participants' demographic information.

The formal experiment consisted of shoulder surfing training and test sessions. During the training, the participants were asked to watch video recordings of two password authentication sessions, one for each authentication method first, and then identify and enter the passwords. The above process was repeated three times in a row and participants got a chance to modify the password they identified in an earlier attempt. The tasks of the test sessions were identical to those of the training sessions, where the participants were first asked to watch six pre-recorded videos of password authentication sessions under different settings. Next, the participants were asked to respond to a shoulder-surfing survey questionnaire (See Section III - E). The video sequence was randomized for each participant, and all of the participants' interaction behaviors were recorded in the system log, such as identified passwords and the start and end time of video playing/replaying.

D. Participants

The experiment was conducted on Amazon Mechanical Turk (MTurk), which has been widely used for shoulder-surfing studies (e.g., [11], [27]). To ensure data quality, we adopted screening measures, such as the MTurkers must have a 95% HIT approval rate (i.e., the percentage of work requests a worker has successfully completed out of the total number of work requests they have attempted), and have completed more than 500 approved HITs as part of prescreening. In addition, attention questions were included in the survey questionnaire. Among the 818 MTurkers who passed the pre-screening survey and completed the pre-experiment, 108 were eligible to proceed with the formal experiment, and 66 participants successfully completed the entire study and each received $5 as compensation. Based on the participants' performance of shoulder surfing, each of the top 10 performers also received $3 as a bonus.

Among those 66 participants, seven were aged between 19 and 24 years old; thirty-five between 25 and 34 years old; seventeen between 35 and 44 years old; five between 45 and 54 years old; two between 55 to 64 years old, and thirty participants were female. Six had a high school degree or equivalent; five had some college yet without a degree; two had an associate degree; thirty-eight had a bachelor's degree; twelve had a master's degree; and three had a doctorate degree.

E. Measures and Variables

We designed three variables to measure the subjective perception of the shoulder-surfing process: cognitive workload, observation clarity, and repetitive learning advantage. NASA TLX [7] is an established research instrument for measuring cognitive workload. NASA TLX consists of multiple dimensions, including mental demand, temporal demand, overall performance, frustration level, and effort. Each dimension is scaled from 0 to 6, and then aggregated into a single dimension, scaled from 0 to 30 (i.e., a higher score indicates a higher cognitive workload). The latter two variables are novel, which we designed specifically for shoulder-surfing tasks, as described below. The variables are measured with a 7-point Likert scale, ranging from strongly disagree to strongly agree.

- *Observation clarity*: "I could clearly observe the password."
- *Repetitive learning advantage*: "My performance in identifying passwords would be significantly improved if I were given additional opportunities to observe the same mobile user authentication session."

The objective measures include accuracy and distance. Accuracy is defined as the percentage of correctly identified characters in a password, and distance is defined as the difference between an identified password and the actual one in terms of the minimum number of single-character edits [17],

[22]. More specifically, we operationalized accuracy as the maximum accuracy and distance as the minimum distance across the three observation attempts under the same shoulder-surfing settings.

To answer RQ1, we performed repeated-measures ANOVA using two authentication methods (touch gesture-based vs. keystroke-based) and observation angle (shoulder, overhead, vs. front) as the independent variables, and cognitive workload, observation clarity, and repetitive learning advantage as the dependent variables. To answer RQ2, we conducted a curvilinear regression analysis using a quadratic model and a moderation analysis, by treating each of the cognitive workload and performance measures as the dependent and independent variables separately, and the authentication method as the moderator.

IV. Results

A. Shoulder-surfing Perception

The descriptive statistics of shoulder-surfing perception are reported in Table II. The results of repeated-measures ANOVA are reported in Table III. The results of post-hoc multiple comparisons of observation angle with Bonferroni adjustments are reported in Table IV.

The analyses of cognitive workload yield significant main effects for both the authentication method (p<.001) and observation angle (p<.001). Specifically, the participants' cognitive workload was higher for the touch gesture-based authentication method (mean=21.04) than for the keystroke-based method (mean=18.40). The multiple-comparison results of the observation angle show that the front observation angle resulted in higher cognitive workload (mean=20.87) than the shoulder (mean=18.94; p<.01) and overhead (mean=19.35; p<.01) observation angles, but no difference was detected between the overhead and shoulder observation angles (p=n.s.).

The analyses of observation clarity yield significant main effects for the authentication method (p<.001) and observation angle (p<.001). Specifically, the touch gesture-based authentication method (mean=2.67) had a lower observation clarity than the keystroke-based counterpart (mean=3.72). The multiple comparisons results of observation angle show that the front observation resulted in a lower observation clarity (mean=2.65) than that of the shoulder (mean=3.61; p<.001) and overhead (mean=3.33; p<.001) observation angles, yet no difference between the overhead and shoulder observation angles (p=n.s.).

The analyses of repetitive learning advantage yield significant main effects of the authentication method (p<.001) and observation angle (p<.001). Specifically, the touch gesture-based authentication method (mean=3.4) has a lower repetitive learning advantage than the keystroke-based authentication method (mean=4.28). The results of multiple comparisons of observation angle show that the front observation angle resulted in lower repetitive learning advantage (mean=3.42) than the shoulder (mean=4.20; p<.001) and overhead (mean=3.89; p<.01) observation angles, and the repetitive learning advantage with the overhead angle was marginally lower than that with the shoulder observation angle (p=.068).

TABLE II. Descriptive Statistics (mean (standard deviation)) of Shoulder-surfing Perception

Authentication Method	Observation Angle	Cognitive Workload	Observation Clarity	Repetitive Learning Advantage
Touch Gesture	Shoulder	20.80	3.02 (2.06)	3.77 (1.79)
	Overhead	21.14	2.67 (2.07)	3.36 (2.13)
	Front	21.17	2.33 (2.14)	3.06 (2.20)
Keystroke	Shoulder	17.08	4.21 (1.67)	4.64 (1.37)
	Overhead	17.56	3.98 (1.69)	4.42 (1.63)
	Front	20.58	2.97 (2.07)	3.77 (1.84)

TABLE III. Results of Repeated ANOVA on Shoulder-surfing Perception

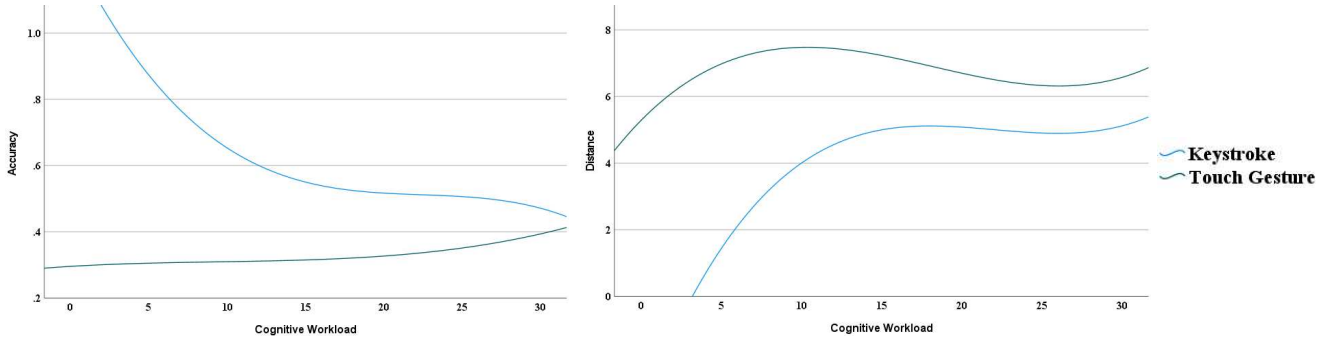
Settings	Cognitive Workload		Observation Clarity		Repetitive Learning Advantage	
	F-value	Sig.	F-value	Sig.	F-value	Sig.
Authentication Method	19.21	<.001***	48.173	<.001***	21.904	<.001***
Observation Angle	8.288	<.001***	16.854	<.001***	12.824	<.001***

a:mean difference; ***: p<.001

TABLE IV. Results of Multiple-Comparisons of Shoulder-surfing Perception from Three Observation Angles

Observation Angle		Cognitive Workload			Observation Clarity			Repetitive Learning Advantage		
(I)	(J)	(I-J)[a]	Std. Error	Sig.	(I-J)[a]	Std. Error	Sig.	(I-J)[a]	Std. Error	Sig.
Shoulder	Overhead	-.409	.457	1	.288	.134	.108	.311	.133	.068
	Front	-1.932	.583	.005**	.962	.204	<.001***	.788	.186	<.001***
Overhead	Front	-1.523	.449	.004**	.674	.164	<.001***	.477	.146	.005**

[a]:mean difference; ***: p<.001; **: p<.01



(a) Accuracy vs. Cognitive Workload (b) Distance vs. Cognitive Workload

Fig. 2. Relationships between Cognitive Workload and Performance Measures in Two Authentication Methods

B. *Correlation and Moderation Effect*

We performed curve fitting between cognitive workload and each of the performance measures for each password authentication method separately. The linear and quadratic relationships were both significant for the two authentication methods, which are plotted in Fig 2. The results of curvilinear regression analysis show that accuracy is positively correlated with cognitive workload (β=.584, t=2.838, p<.01), and distance is positively correlated with cognitive workload (β=1.330, t=7.072, p<.001).

The results of moderation analysis reveal a significant interaction effect between cognitive workload and authentication methods on accuracy (F(3,392)=14.269, p<.001), suggesting that the relationship between accuracy and cognitive workload is moderated by the authentication method. For the keystroke-based authentication, the cognitive workload is negatively associated with accuracy (β=-6.408, t=-4.277, p<.001). However, the cognitive workload is positively associated with accuracy (β=4.969, t=2.015, p<.05) for the touch gesture-based authentication.

In addition, the moderation analysis results also show a significant interaction effect between cognitive workload and authentication methods on distance (F(3,392)=17.676, p<.001), suggesting that the relationship between cognitive workload and distance is moderated by the authentication method. For the keystroke-based authentication, the cognitive workload is positively associated with distance (β=.508, t=3.381, p<.001), whereas cognitive workload is negatively associated with distance for the touch gesture-based authentication method (β=-.927, t=-3.025, p<.01).

V. Discussion

In this study, we investigated the subjective perception of the shoulder-surfing process with two different mobile authentication methods and examined the relationships between the perception and performance in shoulder surfing from an attacker's perspective. There are several major findings of this study. First, the participants perceived higher cognitive workload and lower observation clarity and repetitive learning advantage with the touch gesture-based authentication method than with its keystroke-based counterpart. In addition, the participants perceived higher cognitive workload and lower observation clarity and repetitive learning advantage with the front observation angle shoulder-surfing attacks than with the shoulder and overhead observation angles. Second, this study demonstrates significant correlations between cognitive workload and shoulder-surfing performance. However, the relationships between the perception and performance measures varied with the mobile authentication methods. The shoulder surfing with the keystroke-based authentication method reaches optimal performance when the cognitive load is low. However, the association of cognitive workload and performance measures reveals a contradictory pattern for the touch gesture-based authentication method. Some possible explanations are that the complexity involved in observing and interpreting users' interaction with the touch gesture-based authentication method might have elevated the overall cognitive load of the attackers (see Table II), making additional cognitive effort ineffective.

This study makes multi-fold contributions to mobile security research. First, our study compares the attackers' perception of shoulder surfing between touch gesture- and keystroke-based authentication methods with different observation angles. In particular, this is the first study that includes the front observation angle, which provides a novel and valuable aspect to the field of shoulder-surfing research. Second, this study elucidates the relationship between the subjective and objective measures of shoulder-surfing performance. Moreover, it reveals that the above relationship varies with the mobile authentication method for the first time. Third, we not only used a multidimensional subjective construct but also proposed two new variables to measure the attackers' perception of the shoulder-surfing process.

In view of the relationship between perception and performance in shoulder surfing, increasing the attackers' cognitive workload could help mitigate the success of shoulder surfing attacks. To this end, the designers of mobile user authentication methods could employ strategies to increase the cognitive workload for shoulder surfers. For instance, distraction techniques that introduce distractions during the authentication process can divert the attackers' attention away from critical information. visibility reduction that creates visual obfuscation or pattern masking can obscure the authentication inputs from shoulder surfers, and knowledge transformation that maps a combination of touch gestures to letters. It's important to note that no single countermeasure is foolproof, and a combination of these strategies may provide more effective protection against shoulder-surfing attacks. Since the front observation angle appears to result in the highest cognitive workload, mobile users are recommended to try to face others while being

cognizant of who is behind their back or above their head in a public space to minimize shoulder surfing attacks.

This research has several limitations that warrant future research. Correct password identification through shoulder surfing does not guarantee successful access to a mobile device when state-of-the-art authentication systems are in place. It is imperative for future research to examine both shoulder surfing and authentication as integral components of a holistic security framework. In addition, the password length was fixed across different shoulder-surfing settings in this study. Whether our findings are generalizable to passwords with varying lengths and complexity should be investigated in future studies.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation [Award #s: CNS 1917537]. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the above funding agency.

REFERENCES

[1] F. Binbeshr, M. L. Mat Kiah, L. Y. Por, and A. A. Zaidan, "A systematic review of PIN-entry methods resistant to shoulder-surfing attacks," *Computers & Security*, vol. 101, p. 102116, Feb. 2021, doi: 10.1016/j.cose.2020.102116.

[2] F. Tari, A. A. Ozok, and S. H. Holden, "A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords," in *Proceedings of the second symposium on Usable privacy and security - SOUPS '06*, Pittsburgh, Pennsylvania: ACM Press, 2006, p. 56. doi: 10.1145/1143120.1143128.

[3] J. Lai and E. Arko, "A Shoulder-Surfing Resistant Scheme Embedded in Traditional Passwords," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021. doi: 10.24251/HICSS.2021.860.

[4] S. Schneegass, A. Saad, R. Heger, S. Delgado Rodriguez, R. Poguntke, and F. Alt, "An Investigation of Shoulder Surfing Attacks on Touch-Based Unlock Events," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. MHCI, p. 207:1-207:14, Sep. 2022, doi: 10.1145/3546742.

[5] M. Bâce, A. Saad, M. Khamis, S. Schneegass, and A. Bulling, "PrivacyScout: Assessing Vulnerability to Shoulder Surfing on Mobile Devices," *Proceedings on Privacy Enhancing Technologies*, 2022, Accessed: May 07, 2023. [Online]. Available: https://petsymposium.org/popets/2022/popets-2022-0090.php

[6] Y. Abdrabou, M. Khamis, R. M. Eisa, S. Ismail, and A. Elmougy, "Just gaze and wave: exploring the use of gaze and gestures for shoulder-surfing resilient authentication," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, in ETRA '19. New York, NY, USA: Association for Computing Machinery, Jun. 2019, pp. 1–10. doi: 10.1145/3314111.3319837.

[7] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*, P. A. Hancock and N. Meshkati, Eds., in Human Mental Workload, vol. 52. North-Holland, 1988, pp. 139–183. doi: 10.1016/S0166-4115(08)62386-9.

[8] A. Kumar *et al.*, "PassWalk: Spatial Authentication Leveraging Lateral Shift and Gaze on Mobile Headsets," in *Proceedings of the 30th ACM International Conference on Multimedia*, in MM '22. New York, NY, USA: Association for Computing Machinery, Oct. 2022, pp. 952–960. doi: 10.1145/3503161.3548252.

[9] L. Bošnjak and B. Brumen, "Improving the Evaluation of Shoulder Surfing Attacks," in *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*, in WIMS '18. New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 1–2. doi: 10.1145/3227609.3227687.

[10] W. I. Khedr, "Improved keylogging and shoulder-surfing resistant visual two-factor authentication protocol," *Journal of Information Security and Applications*, vol. 39, pp. 41–57, Apr. 2018, doi: 10.1016/j.jisa.2018.02.003.[11] A. J. Aviv, J. T. Davin, F. Wolf, and R. Kuber, "Towards Baselines for Shoulder Surfing on Mobile Authentication," *Proceedings of the 33rd Annual Computer Security Applications Conference on - ACSAC 2017*, pp. 486–498, 2017, doi: 10.1145/3134600.3134609.

[12] L. Bošnjak and B. Brumen, "Shoulder surfing: From an experimental study to a comparative framework," *International Journal of Human-Computer Studies*, vol. 130, pp. 1–20, Oct. 2019, doi: 10.1016/j.ijhcs.2019.04.003.

[13] L. Zhou, Y. Kang, D. Zhang, and J. Lai, "Harmonized authentication based on ThumbStroke dynamics on touch screen mobile phones," *Decision Support Systems*, vol. 92, pp. 14–24, Dec. 2016, doi: 10.1016/j.dss.2016.09.007.

[14] K. Sundararajan and D. L. Woodard, "Deep Learning for Biometrics: A Survey," *ACM Comput. Surv.*, vol. 51, no. 3, p. 65:1-65:34, May 2018, doi: 10.1145/3190618.

[15] H. Lee, J. Y. Hwang, D. I. Kim, S. Lee, S.-H. Lee, and J. S. Shin, "Understanding Keystroke Dynamics for Smartphone Users Authentication and Keystroke Dynamics on Smartphones Built-In Motion Sensors," *Security and Communication Networks*, vol. 2018, p. 2567463, Mar. 2018, doi: 10.1155/2018/2567463.

[16] D. Zhang, L. Zhou, and S. Pisupati, "Tracing One's Touches: Continuous Mobile User Authentication Based on Touch Dynamics," *AMCIS 2019 Proceedings*, Jul. 2019, [Online]. Available: https://aisel.aisnet.org/amcis2019/human_computer_interact/human_computer_interact/3

[17] L. Zhou, K. Wang, J. Lai, and D. Zhang, "A Comparison of a Touch-Gesture- and a Keystroke-Based Password Method: Toward Shoulder-Surfing Resistant Mobile User Authentication," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 2, pp. 303–314, Apr. 2023, doi: 10.1109/THMS.2023.3236328.

[18] M. Khamis, M. Eiband, M. Zürn, and H. Hussmann, "EyeSpot: Leveraging Gaze to Protect Private Text Content on Mobile Devices from Shoulder Surfing," *Multimodal Technologies and Interaction*, vol. 2, no. 3, Art. no. 3, Sep. 2018, doi: 10.3390/mti2030045.

[19] M. Eiband, M. Khamis, E. von Zezschwitz, H. Hussmann, and F. Alt, "Understanding Shoulder Surfing in the Wild: Stories from Users and Observers," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, in CHI '17. New York, NY, USA: ACM, 2017, pp. 4254–4265. doi: 10.1145/3025453.3025636.

[20] J.-N. Luo, M.-H. Yang, and C.-L. Tsai, "An Anti-shoulder-surfing Authentication Scheme of Mobile Device," *Journal of Internet Technology*, vol. 19, no. 4, p. 10, 2018, doi: 10.3966/160792642018081904028.

[21] M. Khamis, M. Hassib, E. von Zezschwitz, A. Bulling, and F. Alt, "GazeTouchPIN: protecting sensitive data on mobile devices using secure multimodal authentication," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, in ICMI '17. New York, NY, USA: Association for Computing Machinery, Nov. 2017, pp. 446–450. doi: 10.1145/3136755.3136809.

[22] G. Dhandapani, J. Ferguson, and E. Freeman, "HapticLock: Eyes-Free Authentication for Mobile Devices," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, in ICMI '21. New York, NY, USA: Association for Computing Machinery, Oct. 2021, pp. 195–202. doi: 10.1145/3462244.3481001.

[23] O. Osunade, I. A. Oloyede, and T. O. Azeez, "Graphical User Authentication System Resistant to Shoulder Surfing Attack," *Advances in Research*, pp. 1–8, Jun. 2019, doi: 10.9734/air/2019/v19i430126.

[24] K. Wang, L. Zhou, and D. Zhang, "User Preferences and Situational Needs of Mobile User Authentication Methods," in *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, Jul. 2019, pp. 18–23. doi: 10.1109/ISI.2019.8823274.

[25] L. Zhou, K. Wang, J. Lai, and D. Zhang, "Behaviors of Unwarranted Password Identification via Shoulder-Surfing during Mobile Authentication," in *2021 IEEE International Conference on Intelligence and Security Informatics (ISI)*, Nov. 2021, pp. 1–3. doi: 10.1109/ISI53945.2021.9624730.

[26] G. A. Miller, "The magical number seven plus or minus two: some limits on our capacity for processing information.," *Psychological review*, 1956, doi: 10.1037/h0043158.

[27] P. Bekaert *et al.*, "Are Thermal Attacks a Realistic Threat? Investigating the Preconditions of Thermal Attacks in Users' Daily Lives," in *Nordic Human-Computer Interaction Conference*, in NordiCHI '22. New York, NY, USA: Association for Computing Machinery, Oct. 2022, pp. 1–9. doi: 10.1145/3546155.3546706.