Behaviors of Unwarranted Password Identification via Shoulder-Surfing during Mobile Authentication

Lina Zhou
Department of Business Information Systems and Operations Management
UNC Charlotte
Charlotte, USA
lzhou8@uncc.edu

Kanlun Wang
Department of Business Information Systems and Operations Management
UNC Charlotte
Charlotte, USA
kwang17@uncc.edu

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

Dongsong Zhang
Department of Business Information Systems and Operations Management
UNC Charlotte
Charlotte, USA
dzhang15@uncc.edu

2021 IEEE International Conference on Intelligence and Security Informatics (ISI) | 978-1-6654-3838-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISI53945.2021.9624730

***Abstract*—Password-based mobile user authentication is vulnerable to shoulder-surfing. Despite the increasing research on user password entry behavior and mobile security, there is limited understanding of how an adversary identifies a password through shoulder-surfing during mobile authentication. This study empirically examines the behaviors and strategies of password identification through shoulder-surfing with multiple observation attempts and from different observation distances. The results of analyzing data collected from a user study reveal the strategies and dynamics of password identification behaviors. The findings have implications for enhancing users' password security and improving the design of mobile authentication methods.**

Keywords—password, mobile authentication, password identification, shoulder-surfing, adversarial behaviors

I. Introduction

Passwords remain one of the most common methods for mobile user authentication [1]. Although a 'legacy' password offers benefits such as usability and deployability, it receives a poor rating on security [2]. For instance, shoulder-surfing attacks, if successful, can lead to illegal access to a mobile device or app, opening the door to malicious activities.

Despite bodies of research on user password behaviors (e.g., [3], [4]), empirical studies of shoulder-surfing on password-based authentication remain scarce. A few studies on shoulder-surfing susceptibility of password-based authentication (e.g., [5], [6]) have not examined shoulder-surfing behaviors. More importantly, previous studies have overlooked the adversaries' shoulder-surfing strategies across multiple observation attempts.

This study aims to fill the literature gap by investigating the strategies and behavior dynamics of password identification in shoulder-surfing on mobile authentication across multiple observations. The findings have the potential to increase user awareness of security threats to password-based mobile authentication, and guide the design and development of anti-shoulder-surfing authentication methods.

II. Background and Related Work

A password-based authentication matches a user-entered password against a pre-set secret password that typically consists of a string of letters, digits, graphics, and/or symbols [7]. Among others, textual passwords are the most common [8]. In addition, textual passwords are still widely used in accessing mobile apps [9] and are not excluded in advanced biometrics-based mobile authentication design (e.g., [10], [11]). However, passwords are vulnerable to various types of threats such as shoulder-surfing attacks.

During shoulder-surfing, adversaries may observe the live or recorded authentication sessions [12]. Existing empirical studies on shoulder-surfing behavior in mobile authentication have focused on the comparisons between PIN and pattern lock variations [6], PIN and ForcePIN [13], or alphanumeric and graphical passwords [14]. While a previous study investigated the shoulder-surfing susceptibility of password-based authentication [5], the participants were allowed to observe each password entering process only once. Similarly, the majority of shoulder-surfing studies only allowed for less than two observations from an adversary in decoding the victim's authentication credentials (e.g., [12], [13]). However, empirical evidence has shown that an adversary needs to observe a login attempt more than three times on average in order to reproduce low-entropy passwords, and the number is even much higher for high-entropy passwords [15]. More importantly, those studies overlooked adversaries' behavioral changes across shoulder-surfing attempts.

This study aims to address the above limitations by answering three research questions: 1) What are strategies of password identification in shoulder-surfing on mobile authentication?, 2) Do password identification behaviors change over multiple attempts?, and 3) How does the observation context affect password identification behaviors?

III. Methods

To answer the research questions, we conducted a longitudinal user study.

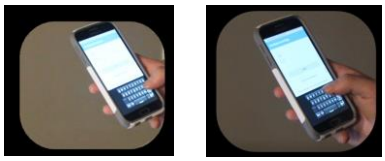
A. Study Design

The study followed a within-subject design. During the preparation phase, the participants started with a lab-based training on password-based authentication using QWERTY and another keyboard, followed by a two-week-long daily practice and ended with a lab-based test. Only those participants who were able to keep up with the daily practice and achieved a sufficient level of accuracy in the lab-based test were eligible for participation in the shoulder-surfing study.

During the formal study, the participants were asked to play the role of an adversary whose task was to identify passwords based on their observations of mobile authentication sessions three times in a row under a variety of settings and to enter their identified passwords after each

978-1-6654-3838-4/21/$31.00 ©2021 IEEE

observation. In this study, we focused on the settings of shoulder-surfing on QWERTY-based password authentication from different distances: near and far. Screenshots of the password authentication sessions are shown in Fig. 1.



a. Far-observation b. Near-observation

Fig.1. Screenshots of the shoulder-surfing settings

We pre-recorded videos of password-based mobile authentication sessions of an expert user. The passwords were set to be 8-character long and non-dictionary-words (e.g., "mkxcrdqu") to achieve a balance between security and complexity. We developed a web-based system that allowed the participants to watch videos of the authentication sessions and enter and/or modify passwords. The system recorded participants' shoulder-surfing behaviors (e.g., entering or deleting characters) in log files. The participants first received training on how to use the experiment system with sample video clips of password authentication sessions identical to those used in the formal study. They did not receive any feedback on the accuracy of their entered passwords.

We recruited participants via a university mailing list. A total of 13 participants successfully completed the experiment and each received a $45 gift card. Among the participants, five were male and eight were female, and nine were aged from 19 to 24 years and four from 25 to 36 years. The study was approved by the university's Institutional Review Board.

B. Data Analyses

To quantify should-surfing behaviors and their dynamics, we introduce several novel measures:1) *Revision:* a binary variable indicating whether an adversary identifies a password from scratch (0) or through revising the result of the previous attempt (1); 2) *Deletion*: the number of deletion operations used in entering a password; 3) *Character-change*: the minimum number of single-character edits needed to transform the password identified in the previous attempt to that of the current attempt; and 4) *Length-change*: the difference in the number of characters of the identified passwords between the current and the previous attempt. The first two variables reflect inter-attempt behaviors directly, and the last two indirectly through measurement on an inter-attempt basis. We performed two-way repeated-measures ANOVA to test the effects of inter-attempt and observation distance (near vs. far) on shoulder-surfing behaviors.

To identify the strategies of password identification behaviors, we applied the open coding method [16] to analyze the content of server logs. Three researchers who have been actively involved in mobile user authentication research first came up with an initial coding scheme as a group, then they used the scheme as a guide to analyzing the data independently while allowing for the creation of new categories, and finally they consolidated the coding results through group meetings.

IV. Results

The descriptive statistics of shoulder-surfing behaviors are reported in Table I. The results of revision show that the majority of the participants leveraged their identification results from earlier attempts by making further modifications. The results of two-way repeated-measures ANOVA are reported in Table II. The results show a significant effect of observation distance on character-change (p<0.05). Specifically, the character-change of the far observation (mean=3.69) was greater than that of the near observation (mean=2.54). In addition, the results show a significant effect of inter-attempt on length-change (p<0.05), with more characters being modified from the first to the second attempt (mean=1.58) than between the last two attempts (mean=0.50).

TABLE I. DESCRIPTIVE STATISTICS OF SHOULDER-SURFING BEHAVIORS

Measurements	Observation Distance	1-to-2 Attempt	2-to-3 Attempt
Revision	Near	77%	92%
	Far	69%	92%
Character-change	Near	3.154 (1.214)	1.923 (1.801)
	Far	3.923 (2.178)	3.462 (2.222)
Deletion	Near	1.462 (1.898)	1.538 (2.727)
	Far	2.615 (2.567)	2.308 (2.287)
Length-change	Near	1.385 (1.325)	0.846 (1.144)
	Far	1.769 (1.964)	0.154 (1.519)

TABLE II. MEAN DIFFERENCE (EFFECT SIZE) OF ANOVA RESULTS

Measurements	Character-change	Deletion	Length-change
Observation Distance	1.154 (0.091) *	0.962 (0.042)	0.154 (0.003)
Inter-attempt	0.846 (0.051)	0.115 (0.001)	1.077 (0.120) *

***: p<0.001; **: p<0.01; *: p<0.05

Based on the content analysis results, we identify several strategies of password identification behaviors. One is *divide-and-conquer*. The participants coordinated their efforts across multiple attempts to identify the password in a shoulder-surfing attack. For instance, one participant identified subsets of a password in separate attempts and finally combined them. Another strategy is *layout-dependent replacement*. The keypress-based design along with the fixed letter arrangement in the keyboard layout allowed the participants to limit the scope of their identification efforts within a sub-area of high likelihood by exploring different alternatives. For instance, one participant replaced the letter 's' with 'd', another replaced 'j' with 'k', and a third replaced 's' with 'x', which are adjacent to each other on the keyboard. A third strategy is *modification vs. fresh-start*. Most of the participants chose to modify the results of their previous attempts, and only four opted to start from scratch. The fresh-start strategy did not appear to work well in most cases. In contrast, building on the previous results through modification was a more effective strategy. Further, the modification occurred to any characters in a password.

V. Discussion and Conclusion

Our findings show that divide-and-conquer is an effective strategy for adversaries to leverage multiple attempts in shoulder-surfing on password-based authentication. Given that shoulder-surfing is a cognitively demanding task, divide-and-conquer can help reduce cognitive load for each of the attempts while achieving optimal overall outcomes. The shoulder-surfing behavior on password-based authentication seems to be keyboard-dependent, where adversaries tend to replace one character with an adjacent one in any of the four directions of the selected keyboard in revising their previous results. Despite the option of modifying the passwords from their previous attempts, a non-negligible percentage of

participants chose to start from scratch. One explanation is that a fresh-start can be a more efficient strategy than incremental changes when the extent of change is large. Another explanation is that password identification is a complex task, which requires both identifying a series of characters and putting them into the correct order. Identifying which character to replace or delete, and where to insert a new character can be demanding given that shoulder-surfing is time-sensitive. Nevertheless, our observations show that fresh-start may not be effective for password identification partly because it completely discounts the previous efforts.

The observation distance had a significant effect on character-change. The far observation introduces a higher level of difficulty to shoulder-surfing attacks than the near-observation. Inter-attempt had a significant effect on length-change suggesting that shoulder-surfers tend to observe more characters to the identified password between the first two attempts than between the last two attempts. Nevertheless, the degree of change to the password remains relatively stable across attempts. Thus, an additional attempt would allow the adversaries to improve their performance continuously. These findings are also consistent with the cognitive load theory that humans have limited working memory for processing information [17]. These findings elucidate the importance of studying shoulder-surfing behavior beyond two attempts.

The findings of this study have two-fold implications. First, mobile users of password-based authentication should consider keeping their distance from others while minimizing the number of authentication sessions to deter shoulder-surfing attacks. Second, developers should use the adversaries' strategies in behavior change as a guide in the design of shoulder-surfing resistant mobile authentication methods.

This research has some limitations that could invite future research. First, it is not uncommon to use a small sample size in controlled lab experiments for mobile user authentication studies [18,19]. This research could benefit from a larger sample of a diversified population. Second, the participants were informed that they had the opportunity to make three observations at the beginning of the study. The findings on behavioral patterns and strategies may not generalize to other shoulder-surfing settings. Third, the password length was set to eight. It would be fruitful to replicate this study with varying the length of passwords to gain an understanding of the possible influence of password length on shoulder-surfing behaviors.

ACKNOWLEDGMENT

This research was partially supported by the National Science Foundation [Award #s: CNS 1917537 and SES 1912898].

REFERENCES

[1] 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.

[2] J. Bonneau, C. Herley, P. C. van Oorschot, and F. Stajano, "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes," in *2012 IEEE Symposium on Security and Privacy*, May 2012, pp. 553–567. doi: 10.1109/SP.2012.44.

[3] O. Wiese and V. Roth, "Pitfalls of Shoulder Surfing Studies," Jan. 2015. doi: 10.14722/usec.2015.23007.

[4] 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.

[5] F. Schaub, R. Deyhle, and M. Weber, "Password Entry Usability and Shoulder Surfing Susceptibility on Different Smartphone Platforms," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, New York, NY, USA, 2012, p. 13:1-13:10. doi: 10.1145/2406367.2406384.

[6] 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.

[7] D. Wang, Z. Zhang, P. Wang, J. Yan, and X. Huang, "Targeted Online Password Guessing: An Underestimated Threat," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, Oct. 2016, pp. 1242–1254. doi: 10.1145/2976749.2978339.

[8] C. Herley and P. Van Oorschot, "A Research Agenda Acknowledging the Persistence of Passwords," *IEEE Security Privacy*, vol. 10, no. 1, pp. 28–36, Jan. 2012, doi: 10.1109/MSP.2011.150.

[9] E. von Zezschwitz, A. De Luca, and H. Hussmann, "Honey, I shrunk the keys: influences of mobile devices on password composition and authentication performance," in *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, New York, NY, USA, Oct. 2014, pp. 461–470. doi: 10.1145/2639189.2639218.

[10] J. Li, H.-C. Chang, and M. Stamp, "Free-Text Keystroke Dynamics for User Authentication," *arXiv:2107.07009 [cs]*, Jul. 2021, Accessed: Sep. 29, 2021. [Online]. Available: http://arxiv.org/abs/2107.07009

[11] 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.

[12] O. Wiese and V. Roth, "See you next time: a model for modern shoulder surfers," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, USA, Sep. 2016, pp. 453–464. doi: 10.1145/2935334.2935388.

[13] H. Khan, U. Hengartner, and D. Vogel, "Evaluating Attack and Defense Strategies for Smartphone PIN Shoulder Surfing," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, Montreal QC, Canada, 2018, pp. 1–10. doi: 10.1145/3173574.3173738.

[14] 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, 2006, p. 56. doi: 10.1145/1143120.1143128.

[15] P. Dunphy, A. P. Heiner, and N. Asokan, "A closer look at recognition-based graphical passwords on mobile devices," in *Proceedings of the Sixth Symposium on Usable Privacy and Security*, New York, NY, USA, Jul. 2010, pp. 1–12. doi: 10.1145/1837110.1837114.

[16] S. H. Khandkar, "Open Coding," 2009. https://www.coursehero.com/file/12549215/open-coding/ (accessed Aug. 24, 2021).

[17] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behavioral and Brain Sciences*, vol. 24, no. 1, pp. 87–114, Feb. 2001, doi: 10.1017/S0140525X01003922.

[18] M. Khamis, T. Seitz, L. Mertl, A. Nguyen, M. Schneller, and Z. Li, "Passquerade: Improving Error Correction of Text Passwords on Mobile Devices by using Graphic Filters for Password Masking," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–8. Accessed: Jun. 08, 2021. [Online]. Available: https://doi.org/10.1145/3290605.3300916

[19] S. A. Alsuhibany, M. Almushyti, N. Alghasham, and F. Alkhudhayr, "The impact of using different keyboards on free-text keystroke dynamics authentication for Arabic language," *Information & Computer Security*, vol. 27, no. 2, pp. 221–232, Jan. 2019, doi: 10.1108/ICS-09-2017-0062.