Understanding Attacking Behaviors toward Password-based Mobile User Authentication

Lina Zhou, *UNC Charlotte* Kanlun Wang, *UNC Charlotte*
Jianwei Lai, *Illinois State University* Dongsong Zhang, *UNC Charlotte*

Abstract

Password-based mobile user authentication is vulnerable to a variety of security threats. Shoulder-surfing is the key to those security threats. Despite a large body of research on password security with mobile devices, existing studies have focused on shaping the security behavior of mobile users by enhancing the strengths of user passwords or by establishing secure password composition policies. There is little understanding of how an attacker actually goes about observing the password of a target user. This study empirically examines attackers' behaviors in observing password-based mobile user authentication sessions across the three observation attempts. It collects data through a longitudinal user study and analyzes the data collected through a system log. The results reveal several behavioral patterns of attackers. The findings suggest that attackers are strategic in deploying attacks of shoulder-surfing. The findings have implications for enhancing users' password security and refining organizations' password composition policies.

1. Introduction

Password remains one of the most common methods for mobile user authentication [1]. Despite some benefits offered by passwords in terms of usability and deployability, 'legacy' passwords receive a poor rating on security [2]. Targeted online password guessing is an underestimated threat [3], in view of the wealth of personal data ranging from usernames and passwords to social security numbers stored on those devices. In particular, shoulder-surfing attacks, if successful, can lead to illegal access to all kinds of sensitive data and information on a mobile device or system which can potentially lead to malicious activities.

There are separate bodies of research on user password behaviors, password strength, password cracking/guessing, and shoulder-surfing attacks. Unlike shoulder-surfing attacks, password cracking/guessing is focused on automated programs or computer-based solutions [4]. Empirical studies on shoulder surfing attacks on password-based authentication remain scarce. Despite a few relatively comprehensive studies of shoulder surfing susceptibility of password-based authentication methods (e.g., [5,6]), none has examined the shoulder-surfing behavior and strategies across multiple observation attempts. Additionally, one of the studies [6] considered graph passwords instead of text passwords. We aim to fill the literature gap by answering the following research question: What are the behavioral patterns of should-surfing attacks? Do attackers coordinate their shoulder-surfing behaviors over multiple attempts? If so, how?

We answer these research questions by conducting a longitudinal user study where participants were asked to play the role of imposters who observed password-based authentication sessions. To support the investigation of multiple observation attempts, we simulate observation attacks by preparing pre-recorded video clips of password authentication sessions and showing each video three times without interruption. This is the first study that examines the temporal patterns of shoulder-surfing behaviors in observation attacks. The findings can help increase users' awareness of security threats to password-based mobile user authentication, guide mobile authentication developers in developing strategies for combating shoulder-surfing attacks, and enhance the password composition policies of organizations and websites.

2. Background and Related Work

In this section, we first provide background on password-based mobile user authentication and shoulder-surfing security models, and finally discuss shoulder-surfing susceptibility of password-based methods.

2.1. Password-based Authentication

Passwords are one of the most common methods for mobile user authentication [7]. 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 [3,4]. Among others, textual passwords are the most common [10]. Passwords bring some usability and deployability benefits [2], but they are vulnerable to various types of attacks (e.g., [11]), where attackers aim to shoulder surf the target user's passwords by leveraging various sources of information such as observations and personal information. Passwords are vulnerable to security threats, partly because password login attempts can potentially be observed by shoulder-surfers.

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021. August 8 -- 10, 2021, Virtual Conference.

2.2. Shoulder-surfing Security Models

Security models used in shoulder-surfing research can be classified into four categories based on whether authentication sessions were recorded or not and how many times adversaries can observe the authentication sessions [12]. The single recording and multiple recording models both present adversaries with recorded authentication sessions. The difference between the two models is that adversaries are only given a small number of recorded sessions with the single recording model, but a large number of recorded authentication sessions with the latter. With the opportunistic observer and the insider observer models, adversaries observe user authentication sessions live. However, opportunistic observers can only observe a small number of authentication sessions live, but insider observers can observe victims many times. It is important to choose the right model based on the context of attacks [12]. For example, the single recording model is not suitable for research on shoulder-surfing attacks among family members, who can observe their victims repeatedly [13]. Instead, the insider observer model will be more suitable. All types of security models would expose credentials to attackers while increasing the vulnerability of a mobile device.

2.3. Shoulder-surfing Susceptibility of Password-based methods

In case of password-based mobile user authentication, the goal of shoulder-surfing attacks is to steal a victim's password. The limited empirical studies on shoulder-surfing behavior in mobile user 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 on a variety of smartphone platforms [5], it was focused on the security model of opportunistic observers, where the participants were allowed to observe each password entering process once only. Similarly, the majority of shoulder-surfing studies (e.g., [15,16]) only allowed for a single observation from an adversary in decoding the victim's authentication credentials except for a couple of studies that considered two observations [13,17]. Empirical evidence has shown that the observer needs to observe a login attempt more than three times on average in order to reproduce the low-entropy passwords, and the number is even much higher for high-entropy passwords [18]. More importantly, those studies did not consider the attackers' change of behaviors over different attempts.

This study aims to fill the literature gap by investigating the dynamics of shoulder-surfing behaviors across different observations of password-based mobile user authentication. To this end, we adopted the opportunistic observer/multiple recording model in this study.

3. Methods

We first describe the password-based authentication method, and then introduce the details of the user study design.

3.1. Password Method

QWERTY remains the de facto keyboard for mobile users to enter their passwords. Since our user study was conducted in the United States, we adopted the conventional QWERTY-based textual passwords as the authentication method. To enter a password, the user needs to press each key corresponding to each character in a password one by one. A login attempt succeeds if an entered password exactly matches the actual password. We developed a prototype to support mobile user authentication using QWERTY, which logs the user's keystroke activities with timestamps.

Figure 1. A screenshot of the shoulder-surfing setting

3.2. Study Design

The study was approved by the Institutional Review Board of the authors' home institute. The participants were first introduced to the objectives and procedure of the study in a designated lab. After signing the consent form, the participants received training on password authentication methods using QWERTY and another keyboard. To make sure that the participants were familiar with the password authentication methods regardless of their prior experience, the participants were asked to go through a two-week-long daily practice with both password authentication methods by entering a new set of 20 passwords. Finally, the participants were asked to take a test in the lab by entering a new set of 40 passwords. Only those participants who were able to keep up with the daily practice and achieved a sufficient level of accuracy (75% or higher) in the last lab-based test were eligible for participation in the shoulder-surfing study. The accuracy was defined as the percentage of the characters in the passwords that were entered correctly.

The shoulder-surfing session was conducted in a lab, where the participants were asked to play the role of an attacker. The attacker's task was to identify passwords based on their observations of mobile user authentication attempts. To simulate the opportunistic observer model, we pre-recorded videos of password-based sessions of an expert user. The use of an expert user followed previous shoulder-surfing studies [6, 19], which not only allowed us to manipulate the password entry performance to match the level of an average user but also helped to minimize the effects of confounding factors introduced by using participants. We developed a web-based experiment system to support the experiment. The system allowed the participants to watch the videos and enter and/or correct their passwords. Participants' video watching and password-entering behaviors (e.g., letter and deletion operation) were recorded in a server log. The participants did not receive any feedback on the accuracy of their entered passwords throughout the experiment. The password length was set to be 8-character long and did not exist in dictionaries (e.g., "mkxcrdqu") to achieve a balance between security and complexity.

The participants received training with the experiment system and practiced shoulder-surfing attacks with sample video clips of password authentication sessions similar to those used in the formal study before proceeding with the formal tasks. The participants were also informed that they would be observing each of the authentication sessions three times in a row, and making one password entering attempt after each observation, and the password length is 8-character long. The shoulder-surfing attacks in the formal tasks were performed under a variety of settings. In this paper, we focused on the setting where the participants observed the QWERTY-based password authentication method from a near distance when the authentication was performed using the thumb of the same hand that holds the device.

It is common to use student participants in mobile user authentication studies [20,21], we recruited participants via a university mailing list. Among the 17 participants who had participated, 13 were successfully completed all experiment tasks and each of them received a $45 gift card. All the participants were English speakers and had experience with interacting with mobile devices in the past three months. Among the participants, nine were aged from 19 to 24 years and four from 25 to 36 years, and five were male and eight were female.

3.3. Variables and Measurements

Given the novelty of the research problem, we introduce the following variables to measure attackers' shoulder-surfing behavior.

- *Guessed password length* is defined as the number of characters included in each password guess,
- *Specificity* is defined as the percentage of correctly identified characters in a password guess, and
- *Sensitivity* is defined as the percentage of characters in the actual passwords that are guessed correctly.

The following three measures were introduced specifically to understand attackers' possible coordination behavior across multiple observation attempts.

- *Levenshtein Distance* measures the discrepancy between two strings of characters [22], which is defined as the minimum number of single-character edits to transform a password guess into another.
- *Modification* is defined as the percentage of the participants who revise a password guess based on the previous password guess(es).
- *Deletion* refers to the number of deletion operations performed in each password observation attempt.

4. Results

The descriptive statistics of shoulder-surfing behavior and across-attempt coordination behavior are reported in Table 1. Since modification and Levenshtein Distance are measured among attempts, we report them separately in Table 2. We test the effect of multiple observation attempts on shoulder-surfing behavior and performance, we performed one-way repeated-measures ANOVA by using attempt as the independent variable, and each of the shoulder-surfing and coordination behavior as the dependent variable separately. For the Levenshtein Distance, we analyzed the effect of the attempt using a paired-sample t-test.

The analyses of specificity did not yield a significant main effect (p>.1). However, the result shows that specificity gradually increases as the number of observations increases. In addition, four out of the thirteen participants could observe the actual password correctly within the three attempts. Among them, three participants made a correct guess in the first attempt and one participant made it in the second attempt. For those who achieve success in the first attempt, only one participant kept their correct guess in the sequential two observations and the other two participants revised their guesses to incorrect ones in the second and third attempts respectively. There was no change found for the participant who made a correct guess in the second attempt.

Given the significant main effect for sensitivity (p<.05), we followed up with post-hoc multiple comparisons. The results are reported in Table 3. The result shows that there was a significant difference in sensitivity between the first two attempts (p<.05) and between the first and the third attempts (p<.001), but the difference between the last two attempts was not significant (p>.1).

In addition, the analyses yielded a significant main effect for guessed password length (p<.01). The results of post-hoc multiple comparisons reveal that the guess length increased from the first to the third attempt (p<0.01) and from the first to the second attempt (p<0.05), and the increase in the guessed length from the second to the third attempt was marginally significant (p<0.1).

Further, the analysis results on Levenshtein Distance yields a marginally significant effect of attempt (p<.1). Specifically, the distance between the first two attempts is greater than that between the last two attempts. A comparison of modification between attempts shows that modification increases with the number of attempts, however, the analysis results on deletion do not reveal a significant effect of attempt (p>.1).

Table 1: Descriptive statistics (mean (std)) for shoulder-surfing behavior

Variables	Attempt 1	Attempt 2	Attempt 3
Specificity	0.732 (0.182)	0.788 (0.184)	0.812 (0.171)
Sensitivity	0.538 (0.194)	0.712 (0.194)	0.817 (0.141)
Guessed password length	5.923 (2.019)	7.308 (1.312)	8.154 (0.689)
Deletion	0.077 (0.277)	1.462 (1.898)	1.538 (2.727)

Table 2: Descriptive Statistics (mean (std)) of shoulder-surfing coordination behavior

Variables	1 to 2	2 to 3
Levenshtein Distance	3.154 (1.214)	1.923 (1.801)
Modification*	76.9%	92.3%

Note: *: Binary variable is reported with the percentage.

Table 3: Post-hoc Multiple Comparison Results

Variables	1 Vs. 2	2. Vs. 3	1 Vs. 3
Sensitivity	p=0.032 (0.173)	p =.124 (0.106)	p =.0003 (0.278)
Guessed password length	p =.049 (1.385)	p =.051 (0.846)	p =.001 (2.231)

Note: p-values are reported, and mean differences are in parenthesis.

5. Discussion and Conclusion

Although the security vulnerability of password-based mobile user authentication has received widespread attention, there are few research studies of attackers' shoulder-surfing on mobile devices. In this study, we empirically investigate the attackers' behaviors displayed in shoulder-surfing while targeting mobile user authentication. Based on an analysis of server log files, we were able to identify several behavioral patterns of shoulder-surfers across multiple attempts.

First, the sensitivity of password guesses improves and the length of password guesses increases over time. These findings are also in line with the cognitive load theory that humans have limited working memory for processing information. The capacity of short-term memory is estimated to be in the order of four items [23], which is much less than the length of passwords used in our experiments. Studies have shown that repetition is one of the most powerful influencers on memory [24]. The effect of repetition on memory judgments is particularly pronounced [24].

Second, there was a greater amount of modification between the first and the second attempts than between the last two attempts based on the Levenshtein Distance. Similarly, there was a greater amount of changes in the guessed password length between the first two attempts than between the last two attempts. These observations further underline the importance of repetition for shoulder-surfing. Even one additional attempt could make a significant difference in the shoulder-surfing outcomes.

Third, while being provided with the option of modifying their previous guesses, a few participants still chose to redo the entire observation in subsequent attempts. One explanation is that the attacker participants had to make significant modifications to their previous guesses, and it might be more efficient to start from scratch than making incremental changes.

The above findings on the temporal patterns of shoulder-surfing behaviors suggest that attackers are strategic in observing user passwords.

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 [25,26]. This research could benefit from a larger sample of a diversified population. Second, the attacker 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 of shoulder-surfing may not generalize to other types of settings. Third, the password length was set to eight. In reality, the length of user passwords varies significantly. Thus, 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. Fourth, we used video recordings of password-based mobile user authentication sessions as stimuli in the user study. In addition to the observation-based attacks, the attackers could also launch recorded attacks. It is a worthy effort to study the behavioral patterns of recorded attacks separately. Our research paves the way for future research in this area.

Acknowledgment

This research was partially supported by the National Science Foundation [Award #s: CNS 1917537 and SES 1912898). Any opinions, findings, and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the above funding agency.

References

[1] K. Wang, L. Zhou, and D. Zhang, "User Preferences and Situational Needs of Mobile User Authentication Methods," 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 Com-

parative 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] 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.

[4] E. Conrad, "Chapter 9 - Domain 9: Operations Security," in Eleventh Hour CISSP, E. Conrad, Ed. Boston: Syngress, 2011, pp. 147–160. doi: 10.1016/B978-1-59749-566-0.00009-6.

[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] K. Wang, L. Zhou, D. Zhang, Z. Liu, and J. Lim, "What is More Important for Touch Dynamics based Mobile User Authentication?," p. 15, 2020.

[8] C. Castelluccia, M. Durmuth, and D. Perito, "Adaptive Password-Strength Meters from Markov Models," p. 14.

[9] B. Ur et al., Poster: The Art of Password Creation.

[10] 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.

[11] J. Yan, A. Blackwell, R. Anderson, and A. Grant, "Password memorability and security: empirical results," IEEE Security Privacy, vol. 2, no. 5, pp. 25–31, Sep. 2004, doi: 10.1109/MSP.2004.81.

[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] A. De Luca, E. von Zezschwitz, L. Pichler, and H. Hussmann, "Using fake cursors to secure on-screen password entry," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13, Paris, France, 2013, p. 2399. doi: 10.1145/2470654.2481331.

[16] M. Khamis, F. Alt, M. Hassib, E. von Zezschwitz, R. Hasholzner, and A. Bulling, "GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices," in Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA, May 2016, pp. 2156–2164. doi: 10.1145/2851581.2892314.

[17] F. Schaub, M. Walch, B. Könings, and M. Weber, "Exploring the design space of graphical passwords on smartphones," in Proceedings of the Ninth Symposium on Usable Privacy and Security - SOUPS '13, Newcastle, United Kingdom, 2013, p. 1. doi: 10.1145/2501604.2501615.

[18] 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.

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

[20] W. Meng, Y. Wang, D. S. Wong, S. Wen, and Y. Xiang, "TouchWB: Touch behavioral user authentication based on web browsing on smartphones," Journal of Network and Computer Applications, vol. 117, pp. 1–9, Sep. 2018, doi: 10.1016/j.jnca.2018.05.010.

[21] S. Sen and K. Muralidharan, "Putting 'pressure' on mobile authentication," in 2014 Seventh International Conference on Mobile Computing and Ubiquitous Networking (ICMU), Jan. 2014, pp. 56–61. doi: 10.1109/ICMU.2014.6799058.

[22] 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.

[23] 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.

[24] D. L. Hintzman, "Repetition and Memory11Preparation of this chapter was supported by a grant GB-40360 from the National Science Foundation. Special thanks are due to Michael J. Hacker and James V. Hinrichs for making their unpublished data available to the author.," in Psychology of Learning and Motivation, vol. 10, G. H. Bower, Ed. Academic Press, 1976, pp. 47–91. doi: 10.1016/S0079-7421(08)60464-8.

[25] 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.

[26] K. Vertanen and P. O. Kristensson, "Complementing text entry evaluations with a composition task," ACM Trans. Comput.-Hum. Interact., vol. 21, no. 2, p. 8:1-8:33, Feb. 2014, doi: 10.1145/2555691.