Smartphone User Authentication using Swipe Gestures on a Keyboard

Alvin Kuruvilla[1], Usman Sheikh[2], Kiran Balagani[3], Paolo Gasti[3], and Anand Santhanakrishnan[3]
[1] Hofstra University, Department of Computer Science, New York, NY, USA
[2] Stony Brook University, College of Engineering and Applied Sciences, New York, NY, USA
[3] New York Institute of Technology, College of Engineering and Computing Sciences, New York, NY, USA
E-mails: akuruvilla1@pride.hofstra.edu, usman.sheikh@stonybrook.edu, {kbalagan, pgasti, asanthan}@nyit.edu

Abstract—**With the increased use of touchscreen smartphones, users have developed unique swipe behavioral characteristics. This paper focuses on swipe length, velocity, acceleration, and percentile velocity to develop user templates and probes. Analyses were conducted on the swipe data of 100 users consisting of over 10,000 swipes. We discuss our methodology for pre-processing user data for better swipe and feature extraction. We then discuss how we created genuine and impostor scores to generate results. Results are measured in error rates generated by our biometric pipeline. Although we could not achieve suitable equal error rates for user authentication, we still expect strong discriminability between user swipes. We end by discussing future work and other possible approaches to refine error rates.**

I. Introduction

Mobile device security has become a growing area of interest and research in recent years, given how heavily we rely on our mobile devices. Traditional solutions employ password or pin-based authentication mechanisms. However, users prefer picking weaker, easier-to-remember passwords [1]. Several vectors could help determine a user's password [1]. For example, users' passcodes can be inferred through smartphone sensors or touchscreen smudges [1]. Thus, a new authentication method needs to address these concerns.

Of all the various solutions, swiping-based authentication is of particular interest. Typically much of a user's time on a smartphone is spent performing swipes of one form or another. Thus, a swiping-based authentication mechanism provides a unique opportunity to authenticate users non-intrusively.

This paper contributes to the existing literature by creating a biometric pipeline using swipe biometrics. Furthermore, while other papers on swipe biometrics often use a small number of participants, this paper uses the Levia "How We Swipe" dataset containing over 1300 users across both Android and IOS platforms (See Section 3A for more details) [2].

The rest of this paper is organized as follows. Section 2 details background and related work, while Section 3 lays out our preliminary methodology. Sections 4 and 5 discuss the results and avenues for future research respectively.

II. Related Work

A. Swipe-Based Continuous/Active Authentication

Continuous Authentication is an authentication mechanism aimed at providing real-time user identity verification. The prevalence of swipe-based operations in a smartphone environment makes it an interesting choice for continuous authentication because it would allow for non-intrusive user authentication in a natural smartphone environment. Frank et al. investigated whether a classifier can be used for continuous authentication after training on touch operations extracted from users [3]. This paper tested the KNN classifier for its speed, and the SVM classifier for its usefulness for binary classification problems [3]. After collecting 30 touch operations features from 41 participants, Frank et al. found that both the KNN and SVM classifiers achieved EERs between 0 and 4% [3]. These results thus suggest swipe biometrics has practical applications in continuous authentication scenarios.

Shen et al. evaluate the applicability of swipe biometrics in active smartphone authentication scenarios across various touch operation types, varying operation lengths, and different application scenarios [1]. After analyzing 134,900 touch operations across 71 participants, Shen et al.'s results demonstrate that swipe biometrics can be used for active authentication achieving equal error rates between 1.72% and 9.01% for numerous touch operations of length 11 [1]. Furthermore, the authors also found authentication improved in specific task scenarios and when having a long observation period or a shorter gap between training and testing [1].

B. Another Angle - Keystrokes

Swipe-based analysis is not the only proposed trait for use in biometrics pipelines. Another promising technique is keystroke dynamics which is the detailed analysis and timing of typing patterns to discriminate between users and imposters [4]. Killourhy and Maxion implemented and evaluated 14 classifiers from the typing dynamics literature against data collected from 51 participants [4]. They found that the top 3 detectors, the scaled Manhattan detector, the Nearest Neighbor (Mahalanobis) detector, and the Outlier Count (z-score) detector, all had EERs between 9% and 11%. These results indicate that keystroke dynamics are another viable contender for desktop active or continuous authentication scenarios.

III. Methodology

A. Exploring the "How We Swipe" Dataset

As mentioned in the Introduction, the Levia "How We Swipe" dataset published in 2022 is the first large-scale dataset for nonlinear swipe data [2]. The dataset contains data for 647 Android users and 262 IOS users [2]. Previous work like Frank et al. and Shen et al. used much smaller datasets (41 and 71 participants, respectively) [1] [3]. These papers also did not focus on keyboard swipes. Instead, these papers focused on straightforward scrolling operations such as up/down and left/right swipes [1], [3]. Beyond the dataset, the authors also found interesting relationships such as that, on larger screens, swipe trajectories are longer but also faster [2].

B. Extracting Swipes

The Levia dataset provides swipe coordinate data for words. The challenge is to extract quality swipes between 4 to 7 characters, as shown in Figure 2, out of these trajectories suitable for biometric authentication (The trajectories represent touchpoints during the swipe(s), so we need to determine what touchpoints make up a swipe). Figure 1 shows the perfect swipes, running through the center of the letter, and user swipes for the word "delay" on a smartphone keyboard.

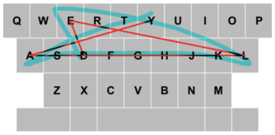

Fig. 1. Perfect and user swipe trajectories glide typing "delay" on a digital keyboard

We postulated the difference between touchpoints would be greater when changing swipe trajectory. For example, when swiping "what" we postulated that the greatest time delta would be the transitions such as from swiping "wh" to "ha". Figure 2 shows the touchpoint time difference for a single user swiping the word "delay".

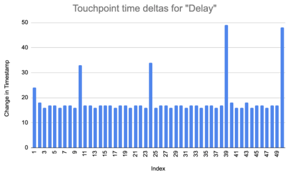

Fig. 2. Time difference (in milliseconds) for touchpoints in "delay" for 1 user

After analyzing the lengths of all the words used in the dataset, we found majority of the lengths to be concentrated around 3 to 7 characters, as shown in Figure 3. Thus when extracting swipes, we only focused on swipes of words in this range to ensure that a substantial amount of the keyboard was swiped through.

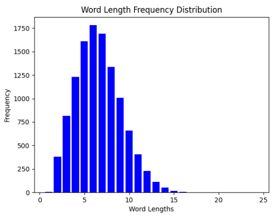

Fig. 3. Dataset Word Length Frequencies

C. Feature Extraction

We based our initial feature set on Shen et al. and Frank et al.'s papers. First, we start with calculating the velocity and acceleration features of the swipe by summing the slope between each point divided once or twice by the difference in the timestamps for velocity and acceleration, respectively. The next feature we used is pairwise percentile velocity. Pairwise percentile velocity is calculated by taking the velocity of the swipe and applying a percentile to it, with each percentile holding some relative mutual information. We employed the 20^{th} percentile pairwise velocity due to Frank et al.'s results showing that this feature held 19.63% of relative mutual information from a list of 31 extracted features. The final feature we used is swipe length which can be derived by summing the Euclidean distance of the coordinates of all the touchpoints that make up a swipe.

D. Generating Genuine and Imposter Scores

Genuine Scores refer to matching user A's template with user A's probe— how similar the biometric data for a given user is to that user's template. The template is the mean feature vector of 70% of swipes of a user. Each probe consists of a feature vector per swipe per impostor user. Impostor Scores refer to the distance between Person A's template and Person B's probe.

E. Biometric Errors

When collecting multiple biometric samples from an individual, likely, these samples would not be identical due to confounding factors such as imperfect imaging or changes in behavioral or physiological characteristics [4]. Consequently, biometric systems calculate a matching score measuring the similarity between the input and the template, with a greater score representing a greater similarity between the two. The system decides whether to accept or reject a user based on a threshold.

In biometric systems, two errors can occur: a false accept and a false reject. A false acceptance could arise when the system mistakes the input of two different people for the same person, while a false rejection is when the system mistakes the inputs of the same person to be those of different users [5]. In such systems, the false accept rate (FAR) and the false reject rate (FRR) are functions of the system threshold [5]. If the system threshold t decreases, the system may become more tolerant to input variation, resulting in a greater FAR[5]. On the other hand, if t increases, the system may become more secure, resulting in a greater FRR [5].

A receiving operating characteristic (ROC) curve can plot FAR against FRR for various threshold values, allowing system performance evaluation [5]. The equal error rate (EER) is the value when the FAR and FRR are equal. Since the FAR and FRR are negatively correlated, it is beneficial for the EER to be a small number. A low EER is required to maintain a low FAR and FRR, resulting in greater accuracy and less sensitivity to data noise from the system.

IV. Results

We chose to run this experiment with 100 users rather than 600 to generate genuine and imposter scores quicker. After calculating the genuine and impostor scores for 100 android users, we generated EER graphs for both individual and multiple features. As shown all the figures converge at around 65% EER.

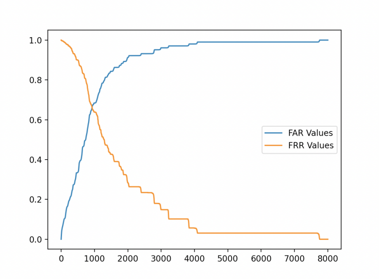

Fig. 4. The EER curve for the Velocity feature for 100 android users

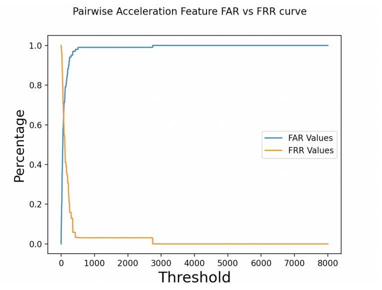

Fig. 5. The EER curve for the Pairwise Acceleration feature for 100 android users

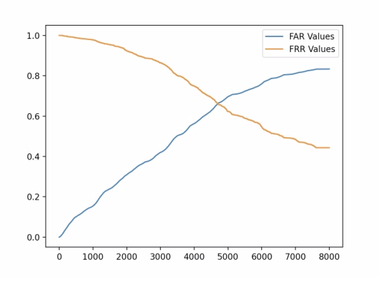

Fig. 6. The EER curve for the length, velocity, acceleration, and percentile velocity features for 100 android users

V. Conclusion and Future Work

While an EER of 65% does not make swipe typing viable for authentication, this paper has shown some discriminability between user swiping patterns and the potential for improvement in swipe authentication. In the future, we plan to improve the EER to make glide typing a more viable form of authentication. Furthermore, we will investigate other potential features that could be used, such as the mid-stroke area covered and the swipe angle. Mid-stroke area covered indicates how much area the finger occupies and, according to Frank et al., provides about 20% of relative mutual information [3]. Finally, we would look into improving the swipe extraction methodology.

VI. Acknowledgements

This research was supported by the National Science Foundation (NSF) Research Experiences for Undergraduates (REU) program. We would like to thank all of the mentors and research fellows at the New York Institute of Technology who have provided their helpful insight and expertise that greatly assisted with our research. This project is funded by National Science Foundation Grant No. CNS-1852316 and New York Institute of Technology.

References

[1] C. Shen, Y. Zhang, X. Guan, and R. A. Maxion, "Performance analysis of touch-interaction behavior for active smartphone authentication," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 3, pp. 498–513, 2016.

[2] L. A. Leiva, S. Kim, W. Cui, X. Bi, and A. Oulasvirta, *How We Swipe: A Large-Scale Shape-Writing Dataset and Empirical Findings*. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3447526.3472059

[3] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 136–148, 2013.

[4] K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *2009 IEEE/IFIP International Conference on Dependable Systems Networks*, 2009, pp. 125–134.

[5] A. K. Jain, A. A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 4–20, 2004.