Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861
Volume XVIII, Issue II, July-December 2024

PRIVACY-PRESERVING KEYSTROKE AUTHENTICATION ON SMARTPHONES

[1]Brijesh Kumar Singh, [2]Meghansh Agarwal, [3]Rachit Koolwal, [4]Aman Goyanka, [5]Piyush Gupta

[1]Assistant Professor, Department of Information Technology, JECRC College
[2]B.Tech Student, Department of Information Technology, JECRC College
[3]B.Tech Student, Department of Information Technology, JECRC College
[4]B.Tech Student, Department of Information Technology, JECRC College
[5]B.Tech Student, Department of Information Technology, JECRC College

ABSTRACT:

This study examines the feasibility of utilizing keystroke dynamics for continuous authentication of smartphone users, addressing the rising need for enhanced security measures. Traditional password-based authentication methods serve as the primary defense against unauthorized access to sensitive smartphone and online data. keystroke dynamics authentication framework seamlessly integrated into modern Identity and Access Management (IAM) systems to provide continuous authentication. To protect user privacy, we incorporate privacy-preserving techniques including permutation, substitution, and suppression, preventing service providers from reconstructing the originally typed text. involves assessing authentication accuracy using real user data, with additional analysis of the impact of privacy preserving techniques on accuracy. Our findings indicate a consistent Equal Error Rate(EER) of approximately for user classification and for user clustering when applying the permutation technique, like results obtained without this technique. Conversely, employing the substitution technique leads to an EER increase of for user classification for user clustering. Additionally, the suppression technique results in a proportional escalation of EER with the number of suppressed keystrokes. This study provides insights into the practical implications of implementing keystroke dynamics for continuous smartphone authentication.

Keywords- Biometrics Technology, Fingerprint Reader, Mobile devices, Laptop, fingerprint or Digital Security.

[1] INTRODUCTION

In today's daily life, the ubiquitous presence of smartphones and mobile devices underscores the need for robust security measures. These devices, once primarily used for calls, now store a wealth of personal information, and facilitate access to sensitive online data such as financial accounts and social networks. However, the predominant reliance on password based authentication methods for device access and online services presents significant vulnerabilities. Attackers need only obtain a single piece of information, whether it's a passcode, PIN, locking pattern, or password, to gain illicit access to a device and compromise sensitive information. Moreover, certain authentication methods may not be suitable for devices with limited interaction capabilities, such as To address these challenges, there's a burgeoning interest in alternative authentication strategies, particularly active or continuous authentication. This paradigm involves the ongoing verification of a user's identity through beta mimetic techniques that monitor their interactions with the device. This study aims to augment existing authentication systems with continuous authentication capabilities by integrating keystroke dynamics—a technique that identifies users based on their typing rhythm. While keystroke dynamics have shown promise in authenticating desktop and laptop users, their efficacy for smartphones, given the virtual nature of the keyboard, remains a subject of investigation. Our research evaluates the suitability of keystroke dynamics for smartphone authentication using three distinct datasets. Furthermore, privacy concerns arise when implementing continuous authentication techniques, as service providers may misuse collected keystrokes to reconstruct users' original input. To mitigate these privacy risks, we incorporate privacy- preserving techniques into our solution. The structure of the paper is as follows: Section II provides an overview of related work on keystroke dynamics and continuous authentication, offering in sights into existing techniques. In Section III, weel aborate on our proposed solution. Section IV presents the results of experiments conducted to assess the effectiveness of our solution. Finally, Section V concludes the paper.

[2] RELATEDWORK

Behavior metrics are metrics that quantify human behavior to recognize or verify a person's identity. Numerous studies have explored the application of behavior metrics to achieve continuous and non-intrusive authentication, analyzing user interactions throughout a session. Various behavior metrics, including keystroke dynamics, mouse movements (along with display resolution), CPU and RAM usage, stylometry, and web browsing behavior, have been investigated for user identification and verification. Keystroke dynamics, which identify users based on their typing behavior, have shown superior results compared to other behavior metrics in previous research. Despite being studied since 1980, keystroke dynamics remains a compelling research area, with recent focus shifting towards their application in While many

algorithmsutilizingmachinelearningtechniqueshavebeenproposedforanalyzingkeystroke dynamics on smartphones, studies have also demonstrated good accuracy using statistical techniques. Notably, Ginette etal. Introduced a novel measure called the "degree of disorder" to discern whether two typing samples likely belong to the same person. This measure serves as a distance metric, enabling the development of various authentication strategies. Although their techniques were not specifically tailored for smartphone authentication, our solution builds upon their work [12]. This study aims to assess the efficacy of utilizing keystroke dynamics as a behavior metric for continuously authenticating smartphone users. To gather and analyze behavioral data, both a client and a server component are necessary. The client component records keystrokes whenever the user types on the smartphone keyboard, while the server component collects and conducts authentication operations continuously. Our design approach facilitates seamless integration with an Identity Access Management (IAM) system, particularly ForgeRock's Open AM authentication system, which offers pre-built authentication modules and supports external authentication services. With integration with Open AM in mind, we developed a continuous authentication framework featuring RESTful APIs accessible by the Open AM system[13]. The authentication strategies employed by this framework hinge on a distance measure proposed by Ginette et al. This measure assesses the similarity between typing samples generated by users, thereby underpinning the authentication process. Understanding how this distance is computed is crucial for comprehending the authentication mechanism. Additionally, we introduce three privacy-preserving techniques aimed at preventing untrusted service providers from reconstructing the text typed by users [15]. To compute the distance between two typing samples, we adopt the methodology proposed by Ginette et al. Users' typing samples consist of sequences of keystrokes, which can be reorganized into sequences of n-graphs for a chosen value of n. Each n-graph represents a sequence of n keys and the latency between them, reflecting the time duration between the pressure of the first and the nth key. Once the value of n is determined, we calculate a distance, denoted as d, between two typing samples, E1 and E2. This distance is computed by assessing both the degree of disorder and the similarity between the two samples. Initially, the n-graphs of the two typing samples are arranged in ascending order based on their time duration and filtered to retain only common n-graphs shared by both samples. The disorder is then determined as the sum of the distances between the position of each element in the two arrays. For instance, considering arrays A=[2,5,1,4,3]and A'=[3,1,2,5,4],the disorder would be computed as(2+2+1+1+4)=10.Thismeasureconsiders that while a user may vary in typing speed, certain key combinations are consistently typed faster than others. The resulting value is normalized by dividing it by the maximum disorder value two n-graphs, if d1 and d2 represent the time durations of the same n-graph, they are considered similar if 1< max (d1, d2)/min (d1, d2) for a constant t> 1. The similarity between two samples is then calculated as 1 minus the ratio of similar n-graphs between S1 and S2 to thetotal number of n-graphs shared by E1 and E2. The final distance is determined by combining these two measures: disorder and similarity [9].To facilitate authentication operations, the framework needs to generate and retain user typing models. Following an initial training

phase, where these models are constructed, the verification of user identities involves comparing typed keystrokes against their respective reference models. Each user model is represented as an NxM matrix, where each row corresponds to a user typing sample, composed of M digraphs [10]. During the training phase, as N typing samples are collected fora user, the system computes the mean distance between all training typing samples. In this scenario, the mean distance for user A'styping samples would be calculated by averaging the distances between all pairs of typing samples. The value m(A) serves as a representation of how a specific user A interacts with a keyboard. Following the model's training phase, subsequent keystrokes received from user A are utilized for authentication purposes. To verify a user's identity, the framework implements two modes of operation:smartphone authentication [1].

[3] COMPARATIVE ANALYSIS

A. To validate the results of the experiments, the k-fold cross validation method was employed due to the absence of a predefined distinction between training and testing data in the datasets utilized. Specifically, a3-fold cross-validation approach was chosen, as it necessitated enough keystrokes to construct an adequate number of typing samples for the testing phase. This method partitions the dataset into three subsets, using two subsets for training and one for testing in each iteration, ensuring robust evaluation of the system's performance across various data samples [8]. The outcomes of the experiments conducted utilizing in [12]]the third dataset, which involved users freely texting on smartphones are presented in this subsection. These findings are subsequently contrasted with those obtained from the other two datasets. Both modes of operation of the proposed framework were examined and evaluated.

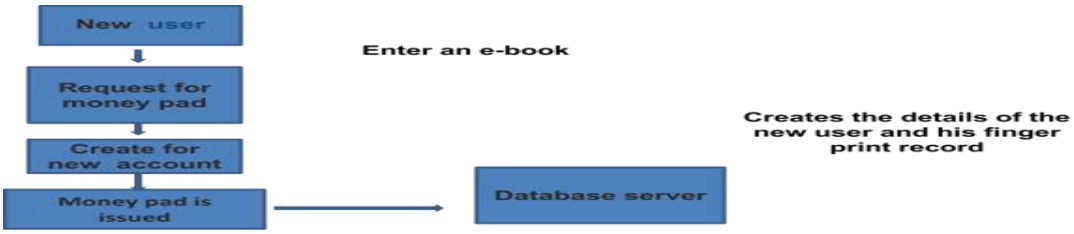

Fig1. Process of Issuing Money Pad

B. The initial findings for the classification mode are depicted in FigureIV-B. The graph illustrates that as the number of digraphs comprising a typing sample increase, there is a notable enhancement in overall performance. The most substantial improvement in performance is observed when typing samples consist of40-50 digraphs, resulting in False Acceptance Rates (FAR) and False Rejection Rates (FRR) ranging between 15% and 20% [3].

A threshold parameter was introduced to determine the proximity of a typing sample to the user model, dictating whether it is recognized as belonging to that user. Adjusting this parameter allows for obtaining different values of False Acceptance Rates (FAR) and False Rejection Rates (FRR). this experiment, the EER was identified as the intersection point of the FAR and FRR in Figure IV-B, yielding a value of 16%.

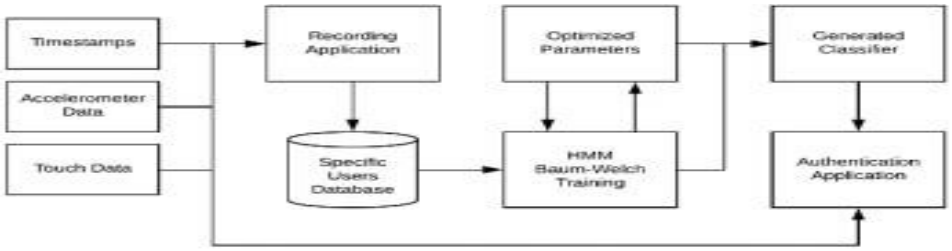

Fig 1 Continuous mobile authentication on various devices

[4] CONCLUSION AND FUTURE WORK

In conclusion, the comparative analysis of studies on continuous user authentication on mobile devices reveals a diverse range of approaches and advancements in the field. From the integration of biometric and PIN authentication for enhanced security in internet banking to the exploration of federated machine learning-based active authentication systems, researchers have contributed significantly to addressing the evolving challenges of securing mobile devices. The introduction of novel frameworks like Gait2vec for continuous authentication based on gait behavior and warmup and transfer knowledge- based federated learning approaches demonstrates the innovation driving this field forward. Additionally, the development of benchmarks like LEAF and FedML facilitates benchmarking and comparison of different authentication methods. Overall, these studies underscore the importance of continuous innovation and collaboration in ensuring robust and effective authentication mechanisms for mobile devices, addressing both security and user privacy concerns. link.

[5] REFERENCES

[1] Cheri nor,UmaruBah1,*,AfzaalHussainSeyal1,UmarYahya2"combiningpinand biometricidentificationsas enhancement to authentication in internet banking" IEEE [2023]

[2] Chen Wang, Yan Wang, Yingying Chen "User Authentication on Mobile Devices: Approaches, Threats, and Trends "IEEE CNS [2020]

[3] Mohamad Wazzeh1, Hakima Ould-Slimane2, Chamseddine Talhi1, Azzam Mourad "Warmup and Transfer Knowledge-Based Federated Learning Approach for IoT Continuous Authentication" [2022]

[4] C.He,S.Li,J.So,M.Zhang,H.Wang,X.Wang,P.Vepakomma,A.Singh,H.Qiu,

[5] L. Shen, P. Zhao, Y. Kang, Y Liu, R. Raskar, Q. Yang, M. Ann avaram, and S. Avestimehr, "Fedml: A research library and benchmark for federated machine learning," preprintarXiv:2007.13518, 2020

[6] S. Caldas, S. M. K. Duddu, P. Wu, T. Li, J. Konecnˇ y, H. B. McMahan, `V. Smith, and A. Talwalkar, "Leaf: A benchmark for federated settings,"arXiv preprint arXiv:1812.01097, 2019

[7] L. He, C. Ma, C. Tu, and Y. Zhang, "Gait2vec: Continuous authentication of smartphone users based on gait behavior," in 2022 IEEE 25thInternational Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2022, pp. 280–285.

[8] M. Abuhamad, T. Abuhmed, D. Mohaisen, and D. Nyang, "Autosen: Deep-learning-basedimplicit continuousauthenticationusingsmartphonesensors,"IEEE Internet of Things Journal, vol. 7, no. 6, pp.5008–5020, 2020

[9] L. Sun, B. Cao, J. Wang, W. Srisa-an, S. Y. Philip, A. D. Leow, andS. Checkoway, "Kollector: Detecting fraudulent activities on mobiledevices using deep learning," IEEE Transactions on Mobile Computing,vol. 20, no. 4, pp. 1465–1476, 2020

A. Monschein, J. A. P. Perez, T. Piotrowski, Z. Nochta, O. P. Wald horst, ´and C. Zirpins, "Towards a peer-to-peer learning environment for continuous authentication," in 2021 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2021, pp. 1–6.

[10] P. Oza and V. M. Patel, "Federated learning-based active authentication on mobile devices," arXiv preprint arXiv:2104.07158, 2021.

[11] O. A. Wahab, A. Mourad, H. Otrok, and T. Taleb, "Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems," IEEE Communications Surveys Tutorials, vol. 23, no. 2, pp. 1342–1397, 2021.

[12] T. Zhu, Z. Qu, H. Xu, J. Zhang, Z. Shao, Y. Chen, S. Prabhakar, and J. Yang, "Riskcog: Unobtrusive real-time user authentication on mobile devices in the wild," IEEE Transactions on Mobile Computing, vol. 19, no. 2, pp. 466–483, 2019.

[13] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," ACMTransactionsonIntelligent SystemsandTechnology(TIST),vol. 10, no. 2, pp. 1–19, 2019.

[14] J. M. J. Valero, P. M. S. Sanchez, A. H. Celdran, and G. M. P ´ erez, "Ma- ´chine learningasanenablerofcontinuousandadaptiveauthenticationinmultimediamobile devices," inHandbook ofResearch on Multimedia Cyber Security.IGI Global, 2020,

[15] pp.21–47

A. Mourad, H. Tout, O. A. Wahab, H. Otrok, and T. Dbouk, "Adhoc vehicular fog enabling cooperative low latency intrusion detection," IEEE Internet of Things Journal, vol. 8, no. 2, pp. 829–843, 2020.