Siamese Neural Network for Keystroke Dynamics-Based Authentication

Subjects: Computer Science, Artificial Intelligence

Contributor: Kamila Lis , Ewa Niewiadomska-Szynkiewicz , Katarzyna Dziewulska

User-specific behavioral biometrics is widely used to increase login security. The usage of behavioral biometrics can support verification without bothering the user with a requirement of an additional interaction.

keystroke dynamics behavioral biometry keyboard

1. Introduction

User authentication enforces secure access to computer systems, networks, and devices. The motivation to authenticate users ranges from access control reasons to business development purposes such as adding e-commerce, e-banking, etc. In recent years, the challenges related to identifying users and ensuring secure access to their accounts have increased due to the increasing number of cyber threats. Therefore, it is recommended to extend passwords to make them more challenging to crack [1]. Meanwhile, most people need to take precautions to secure their accounts. Persuading users to set strong passwords is needed. They often use simple passwords and do not use two-factor authentication for their accounts. Many people use the same password for most of their accounts. In contrast, attackers launch more and more phishing URLs every year. Phishing targets a user's authentication rights and identity.

Nowadays, it is clear that passwords are no longer the only approach to user authentication. There is a growing demand for different types of technologies for user identification, both online and in physical systems. As technology advances, authentication mechanisms are constantly being improved. Many authentication technologies and an even greater range of activities require authentication methods.

A standard security mechanism is the addition of authentication steps, most often requiring the rewriting of a secret transmitted through another communication channel. With frequent logins, such an extended process can persist for the user. However, it is possible to conduct additional verification without introducing additional interaction with the user, but only by observing their behavior. It has been proven that the keystroke pattern of an individual is unique, in a similar manner as a handwritten signature [2]. In this way, when logging in, the typing dynamics can also be analyzed, in addition to verifying the correctness of the password. In addition, in the event of password compromise, device takeover, or remote control, behavioral biometrics allows abnormalities in user behavior to be observed. It is also worth mentioning that implementing algorithms based on behavioral biometrics is simple, as they do not require any additional hardware components.

Although the potential uses of keystroke dynamics remain strong, it still needs to be widely deployed in many security applications. Unlike the stability offered by iris or retinal recognition, there is no permanence in the specific typing pattern of an individual. Typing rhythm may change daily due to several factors, such as physical fatigue, lack of attention when typing, or even using a different computer keyboard. Data acquisition is challenging in an uncontrolled environment. Users can use different machines, virtual keyboards, smartphones, etc. For each device, the collected data are slightly different. Furthermore, poorer quality equipment can degrade the accuracy of the result. Maxion et al. [3] claim that artifacts injected by a USB keyboard can change an algorithm's decision by nearly 20 percentage points. In addition, there may also be some legal issues or user concerns about the security of their data. These concerns are analogous to other biometrics that are nevertheless widely accepted (e.g., unlocking a phone with a fingerprint).

As noted in [4], most methods based on behavioral biometrics usually have three main drawbacks: (i) they need lengthy interactions (minutes), or a very long time to collect the passphrase samples, to learn the user behavior, (ii) they require ad-hoc interaction challenges or (iii) need a model per user to improve model accuracy.

Despite the disadvantages and challenges mentioned above, biometric methods are increasingly used in authentication systems. It is due to their numerous advantages. These include:

- Uniqueness—everyone writes differently;
- Low implementation and deployment costs—no need to provide the user with additional equipment and easy integration of additional modules into commonly-used authentication systems;
- Transparency;
- Provision of additional security mechanism;
- Possibility of continuous monitoring.

It is also worth mentioning that The New York State DMV and the European Banking Authority approve typing biometrics as a compliant identity authentication method [5].

2. Siamese Neural Network for Keystroke Dynamics-Based Authentication

Analysis of keystroke dynamics, similarly to other biometrics such as voice or written text analysis, can be broadly classified into two types—*static* (structured text) and *dynamic* (free text) [6]. Static analysis involves analyzing the keystroke behavior of an individual on a predetermined phrase at certain points in the system. In authentication systems based on login and password, such predefined phrases are the user's login data. It can also involve the use of a particular phrase that is common to all users. The user's typing pattern is analyzed only at this stage of interaction with the system. Static text analysis can be deployed, especially in systems without further text entries.

Dynamic analysis involves continuous or periodic monitoring of keystroke behavior. The analysis starts at login and continues the entire time one uses the system. Compared with fixed-text keystroke dynamics, the free-text case presents some additional challenges. First, the number of valuable features may differ among input sequences. Second, the optimal length of a keystroke sequence for analysis is a factor that must be considered—a more extended sequence is slower to process and might include more noise. In comparison, a shorter sequence may need more features to be considered.

The idea of using information about typing rhythm on a keyboard to authenticate a user has been developing for many years. Its basis lay in the 1897 observation when it was noticed that telegraph operators have distinctive patterns of keying messages over telegraph lines and can recognize their fellow workers based on their typing rhythms [7]. This method of identifying the sender of the telegraph by using the rhythm, pace, and syncopation of the telegraph keys, known as the “Fist of the Sender”, was also valuable during World War II [8].

Over the past 40 years, it has been used in many methods for static analysis of user keystroke dynamics. The first attempts were based on a simple statistical approach. The mean and the standard deviation were computed for comparison using hypothesis testing by authors of [9][10]. In other research papers, various distance measures were used [11][12][13]. In the 1990s, widely used solutions employed statistical estimation of the distance between vectors of delays between keyboard events. The topic of keystroke dynamics authentication gained significant popularity in the 2000s [14]. In subsequent years, techniques using clustering methods, different distance estimations, measures of the randomness of intervals (for free text), decision trees, and artificial neural networks have been developed and tested. Since 2010 measures of disorder, Hidden Markov models, probability density estimates, and machine learning (SVM, Random Forest) have also been used.

Siamese neural network, a unique artificial neural network, has received much attention recently. In [15], authors investigated the feasibility of using a Siamese network for keystroke authentication on full passwords. They developed and trained the Siamese network model on 200 samples per user taken from the CMU dataset [13], collecting data on the typing dynamics of full passwords by users who entered the same password on the same machine. They experimented with various network architectures, pre-trained models, and attention obtaining satisfactory results. The authors claim they obtained 90.8% user authentication accuracy after 30 logins. Another tool using the Siamese network for user authentication based on his few previous logins is described and evaluated in [4]. Logins on both workstations and smartphones were considered. The authors considered different human-computer and human-smartphone interaction features, i.e., keystroke and mouse dynamics, holding patterns, and touch patterns. The method was tested on a database that contained over 100 K different web interactions collected in the wild. Unfortunately, this dataset is not publicly available online. The user authentication accuracy achieved after 30 logins was equal to 88%.

Generally, the method considered should be adjusted to the size and quality of the available experimental dataset. The comparative study of two commonly used statistical algorithms: the scaled Manhattan and the Instance-based Tail Density (ITAD) metrics [16], with the state-of-the-art deep learning model TypeNet [17] on small and large datasets, is presented and discussed in the paper [18]. The results serve as a reminder of the general intuition—

deep neural networks produce better results when trained on a sufficiently large dataset, whereas when the learning dataset is undersized simple statistical algorithms perform better.

The usability of keystroke dynamics-based authentication was analyzed in [19], where it serves as a second factor in multi-factor authentication. The scaled Manhattan distance was computed for signup, login (username and full password), and account recovery, which required an additional predefined text. The authors propose employing an OTP factor (one-time password) until a sufficient number of samples have been collected and after the enrollment process is completed whenever keystroke-based authentication fails.

References

1. NIST Special Publication 800-63B-Digital Identity Guidelines. Available online: https://pages.nist.gov/800-63-3/sp800-63b.html#memsecretver (accessed on 21 April 2023).
2. Obaidat, M.S.; Sadoun, B. Keystroke dynamics based authentication. Biom. Pers. Identif. Networked Soc. 1996, 479, 213–229.
3. Maxion, R.A.; Commuri, V. This Is Your Behavioral Keystroke Biometric on Rubbish Data; Carnegie Mellon University: Pittsburgh, PA, USA, 2020.
4. Solano, J.; Rivera, E.; Castelblanco, A.; Tengana, L.; Lopez, C.; Ochoa, M. A Siamese Neural Network for Behavioral Biometrics Authentication. In Proceedings of the ICLR 2021 Conference, Virtual Event, Austria, 3–7 May 2021.
5. Siahaan, C.R.P. Spoofing keystroke dynamics authentication through synthetic typing pattern extracted from screen-recorded video. Big Data 2022, 9, 111.
6. Banerjee, S.; Woodard, D. Biometric Authentication and Identification Using Keystroke Dynamics: A Survey. J. Pattern Recognit. Res. 2012, 7, 116–139.
7. Lowe, W.B.; Harter, N. Studies in the physiology and psychology of the telegraphic language. Psychol. Rev. 1897, 4, 27–53.
8. Coppenrath, L.F. Biopassword Technology Overview. 2001. Available online: http://www.lfca.net/Reference%20Documents/Biometric%20Technology%20Overview.pdf (accessed on 7 March 2022).
9. Gaines, R.S.; Lisowski, W.; Press, S.J.; Shapiro, N. Authentication by Keystroke Timing: Some Preliminary Results; RAND Corporation: Santa Monica, CA, USA, 1980.
10. Umphress, D.; Williams, G. Identity verification through keyboard characteristics. Int. J. Man-Mach. Stud. 1985, 23, 263–273.

11. Joyce, R.; Gupta, G. Identity Authentication Based on Keystroke Latencies. Commun. ACM 1990, 33, 168–176.
12. Gunetti, D.; Picardi, C.; Ruffo, G. Dealing with Different Languages and Old Profiles in Keystroke Analysis of Free Text. In Proceedings of the 9th Congress of the Italian Association for Artificial Intelligence, Milan, Italy, 21–23 September 2005; Volume 3673, pp. 347–358.
13. Killourhy, K.S.; Maxion, R.A. Comparing anomaly-detection algorithms for keystroke dynamics. In Proceedings of the 2009 IEEE/IFIP International Conference on Dependable Systems Networks, Lisbon, Portugal, 29 June–2 July 2009; pp. 125–134.
14. Teh, P.S.; Teoh, A.; Yue, S. A Survey of Keystroke Dynamics Biometrics. Sci. World J. 2013, 2013, 408280.
15. Giot, R.; Rocha, A. Siamese Networks for Static Keystroke Dynamics Authentication. In Proceedings of the 2019 IEEE International Workshop on Information Forensics and Security (WIFS), Delft, The Netherlands, 9–12 December 2019; pp. 1–6.
16. Ayotte, B.; Banavar, M.; Hou, D.; Schuckers, S. Fast Free-Text Authentication via Instance-Based Keystroke Dynamics. IEEE Trans. Biom. Behav. Identity Sci. 2020, 2, 377–387.
17. Acien, A.; Morales, A.; Monaco, V.; Vera-Rodriguez, R.; Fierrez, J. TypeNet: Deep Learning Keystroke Biometrics. IEEE Trans. Biom. Behav. Identity Sci. 2021, 4, 57–70.
18. Wahab, A.; Hou, D. When Simple Statistical Algorithms Outperform Deep Learning: A Case of Keystroke Dynamics. In Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods ICPRAM, Lisbon, Portugal, 22–24 February 2023; pp. 363–370.
19. Wahab, A.; Hou, D.; Schuckers, S. A User Study of Keystroke Dynamics as Second Factor in Web MFA. In Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy, Charlotte, NC, USA, 24–26 April 2023; pp. 61–72.