Keystroke Dynamic Authentication Using Combined MHR (Mean of Horner's Rules) and Standard Deviation

Didih Rizki Chandranegara*[1], Fauzi Dwi Setiawan Sumadi[2]
[1,2]Universitas Muhammadiyah Malang
didihrizki@umm.ac.id[*1], fauzisumadi@umm.ac.id[2]

Abstract

Keystroke Dynamic Authentication used a behavior to authenticate the user and one of biometric authentication. The behavior used a typing speed a character on the keyboard and every user had a unique behavior in typing. To improve classification between user and attacker of Keystroke Dynamic Authentication in this research, we proposed a combination of MHR (Mean of Horner's Rules) and standard deviation. The results of this research showed that our proposed method gave a high accuracy (93.872%) than the previous method (75.388% and 75.156%). This research gave an opportunity to implemented in real login system because our method gave the best results with False Acceptance Rate (FAR) is 0.113. The user can be used as a simple password and ignore a worrying about an account hacking in the system.

Keywords: *Keystroke Dynamic Authentication, Mean of Horner's Rules, Biometric Authentication*

1. Introduction

Many people carry out daily activities through the Internet, such as money transactions, file transfers, gathering information, and chatting [1]. Each activity carried out requires a unique identifier (ID) and password to verify an account that is used for authentication on the system. The passwords as account verification are the most widely used method. Generally, someone uses a password with the date of birth, the name of the parent or the name of boyfriend/girlfriend. However, along with the development of technology today, there are various ways to hack passwords used in the system. For example, the attacks used brute force methods and birthday attack algorithms [2]. This attacking resulted in losses such as leaking of personal information and loss a lot of money in a bank account.

Keystroke Dynamic Authentication (KDA) is the proper solution for handling these problems. KDA is one of the biometric authentication techniques. Biometric authentication is an authentication technique that utilizes what is inside a person or something unique in a person, such as a face, fingerprints, habits, and retina [3][4]. KDA is an authentication that utilizes a person's habit in typing a text with a keyboard. So, everyone will feel safe for using a password with they have and they usually use as a password.

This research aims to improve the accuracy of previous researches, i.e. Gupta & Gopal [5] and Yang & Fang [6]. Researches [5] [6] used the average for the KDA classification. The use of average according to Ho and Kang [1] research is not suitable for data streams such as KDA, so it is proposed to use the mean of Horner's Rule (MHR) that is suitable for data streams. Based on [1], this research proposed to use MHR for the classification of KDA and combined it with a method that was used by Gupta & Gopal [5] and Yang & Fang [6]. An explanation of the method used will be presented in the Research Method and continued with the Results and Discussion of the results of the research that has been carried out.

1.1 Keystroke Dynamic Authentication

Keystroke Dynamic Authentication (KDA) is an authentication that utilizes how to type text by someone using a keyboard [7]. KDA was created to improve the traditional security of authentication such as passwords [8]. The password is commonly used by someone, generally contains what they have such as using birth date or parents name. If only use the password method, it will easy for the attacker to hack a user accounts, and if use a KDA, it becomes very difficult to hack, because it uses typing habits that not everyone has the same typing habits. KDA can combine with the passwords to the information system, so a system security is increasing.

Chandranegara, D., & Sumadi, F. (2018). Keystroke Dynamic Authentication Using Combined MHR (Mean of Horner's Rules) and Standard Deviation. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 4(1). doi:http://dx.doi.org/10.22219/kinetik.v4i1.719
Receive November 01, 2018; Revise November 08, 2018; Accepted November 09, 2018

KDA is one part of biometric authentication because it basically utilizes what is inside a person, i.e. habits. From the needs of the device, KDA only uses that are on the computer, i.e. the keyboard. While other biometric authentication uses adding devices. Additional devices to biometric authentication are needed for this authentication. Face, retina, and fingerprints are biometric authentication that utilizes additional devices that are not available on the computer and the cost of their application is expensive. While KDA only uses the keyboard and all types of laptops or computers must have it. Thus, KDA is a cheap biometric authentication [9], because it only uses existing devices. The advantage of KDA is that someone will not realize that the system used KDA [10].

In KDA, there is a hold time and latency time [11]. Hold time is the duration to press a character until it releases the character again (Key Down to Key Up). Latency time is the duration of releasing a character until pressing the next character again (Key Up to Key Down). Besides hold time and latency time, there is also a flight time [12]. Flight time is the time taken from Key Down to Key Down. Illustration of hold time, latency time, and flight time have been presented in Figure 1.

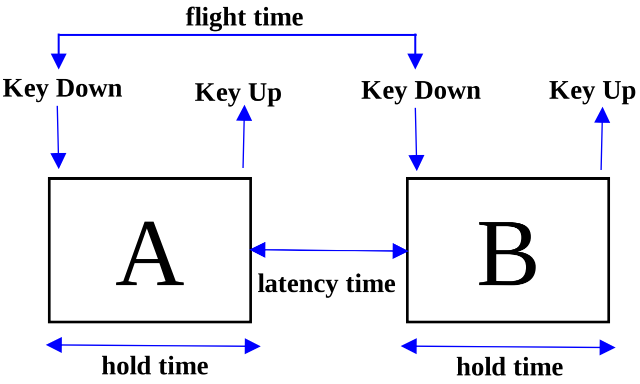

Figure 1. Illustration of Hold Time, Latency Time, and Flight Time

2. Research Method

This research uses a method from Yang and Fang [6] which was improved research from Joyce and Gupta classification methods [5]. The proposed method is replacing the average use with MHR (mean of Horner's Rule). MHR is perfect for data streams such as Keystroke Dynamic Authentication (KDA) [1]. Based on this statement [1], this research was improved research using the proposed method i.e. MHR and classification methods by [6]. The Equation 1 used in the classification of KDA by Joyce and Gupta [5] is:

$$\emptyset = \frac{1}{n}\sum_{i=1}^{i=n} \|M - S_i\| + (1{,}5 * \sigma) \tag{1}$$

In addition, the improved research of the Joyce and Gupta methods [5] conducted by Yang and Fang [6] has been presented in this Equation 2.

$$\emptyset = \frac{1}{n}\sum_{i=1}^{i=n} \|M - S_i\| + (3 * \sigma) \tag{2}$$

Then, the proposed method in this research is using this Equation 3.

$$\emptyset = \frac{1}{n}\sum_{i=1}^{i=n} \|MHR - S_i\| + (3 * \sigma) \tag{3}$$

To determine the MHR value has been presented in this Equation 4 [1].

$$MHR = \frac{\left(\frac{\left(\frac{\left(\frac{x_1 + x_2}{2}\right) + x_3}{2}\right) + x_4}{\dots}\right) + x_n}{2} \tag{4}$$

To determine the user and attacker, this research used description based on previous research [5][6]. Classification process (user and attacker) from previous research:

1. If |M-T| <= ∅, then the logged in user is considered as the real user.
2. If |M-T| > ∅, then the logged in user is considered as an attacker.

The evaluation method used in this research is an accuracy measurement method. Accuracy measurement using True Positive, True Negative, False Positive, and False Negative. Where True Positive is the truth to predict users and True Negative is the truth to predict attacker. Meanwhile, False Positive is an error in predicting the attacker as the legitimate user and False Negative is an error in predicting the user as the attacker. For more details, Table 1 is the confusion matrix tables for determining an accuracy.

Table 1. Confusion Matrix

	User (Prediction)	Attacker (Prediction)
User (Actual)	True Positive (TP)	False Negative (FN)
Attacker (Actual)	False Positive (FP)	True Negative (TN)

The following Equation 5 is used to determine accuracy.

$$Accuration = \left(\frac{TP + TN}{TP + FP + TN + FN}\right) x100\% \tag{5}$$

3. Results and Discussion

3.1 Results

This research used data from Killourhy and Maxion research [13] and dataset can be accessed on this link https://www.cs.cmu.edu/~keystroke/. This data contained 51 users, each user has 400 data Keystroke Dynamic Authentication (KDA) records. This KDA data recording was carried out for 8 days and obtained 30 males and 21 females. The time used in Keystroke Dynamic Authentication data is seconds. Each user in the data-set typed the same text, that is "**.tie5Roanl**" and the use of this text is based on several experiments by [13]. So, each user has different KDA data with another user. In this data-set, there are 31 features consisting of hold time between characters (in dataset symbolized with H), latency time between characters (in dataset symbolized with UD), and flight time between characters (in dataset symbolized with DD). For a more detailed explanation about the dataset, can be accessed on this link https://www.cs.cmu.edu/~keystroke/. The scenario for evaluating the proposed method is:

1. Training data of User take from data 1 to 350.
2. Testing data of User take from data 351 to 400.
3. Each user will be used as an attacker, so the test scenario will be 51 scenarios.
4. Attacker data will be taken from data to 351 to 400 from each user which be an attacker.

From this scenario, testing will be using the program (using php programming language) and the results will be compared with the two previous methods [5][6]. The results of testing that have been presented in line graphs and can be seen in Figure 2. In addition, we also present a table of the average accuracy of testing based on line graphs. The following Table 2 shows the average accuracy of each method.

Table 2. Results of the Research

Method	Mean of Accuracy (%)
Joyce and Gupta [5]	75.388
Yang and Fang [6]	75.156
Proposed Method	93.872

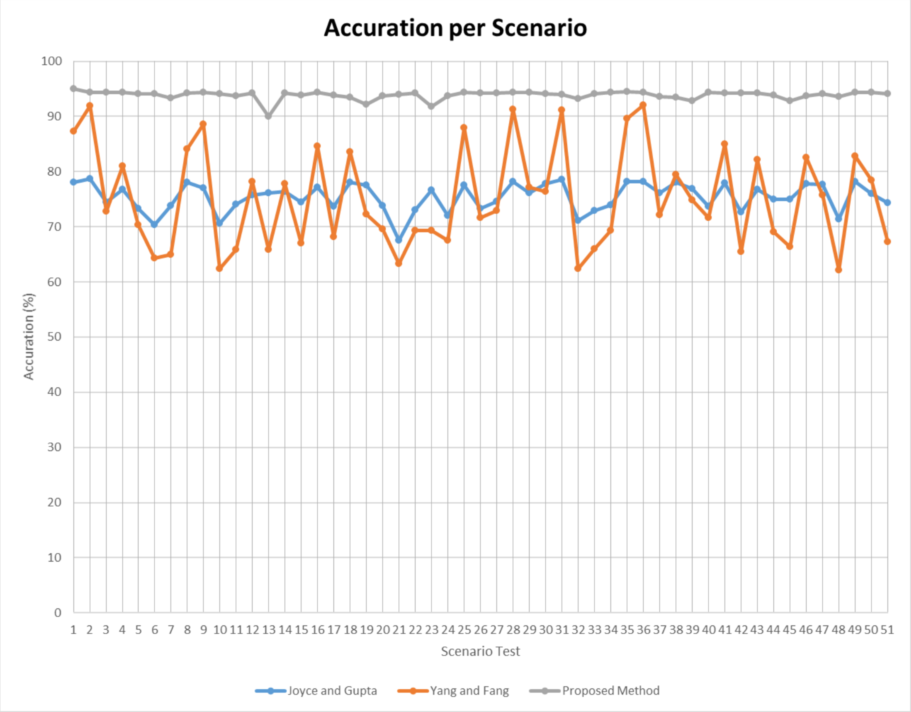

Figure 2. Results of Testing in Line Graphs

3.2 Discussion

Based on the results of the research, the proposed method has an average accuracy level above 90%. While the method used by Joyce and Gupta [5] and Yang and Fang [6] still shows an average accuracy above 70%. Researches by Joyce and Gupta [5] and Yang and Fang [6] still uses the average as a method of determining classification between the user and the attacker. While the proposed method has used MHR (Mean of Horner's Rule) as a determinant of the classification of users and attackers. It is making the proposed method get high accuracy than the previous method. The use of MHR is based on the results of Ho and Kang research [1] which shows that the average use of KDA data is not fully suitable to be used as a determinant of classification. And it is preferable to use MHR because this method is suitable for data streams such as KDA. One of the characteristic data streams is changing quickly over time [1]. It has led researchers to propose MHR as a determinant of the classification of users and attackers on Keystroke Dynamic Authentication (KDA). In KDA, the data will change every update a new password and it’s will change all of KDA data.

Based on the results of the research, if Keystroke Dynamic Authentication (KDA) method of [5] or [6] is applied to the login system, then the probability of the user being rejected will be very high or the probability of the attacker being considered as the legitimate user is very high. To prove this statement, we tested False Alarm Rate (FAR) and False Rejected Rate (FRR) on the proposed method and the previous method [5][6]. FAR is a rate of misclassification to receive an attacker as a legitimate user [12][14][15]. Whereas, FRR is a misclassified rate to reject the user as a legitimate user [12][13][14]. To get the values from FAR and FRR, this research utilizes the Confusion Matrix table to be more easily understood. The following formula is used by this research:

$$FAR = \left(\frac{FP}{TN + FP}\right) \tag{6}$$

$$FRR = \left(\frac{FN}{TP + FN}\right) \tag{7}$$

Table 3 and Table 4 are the average FAR and FRR from the results of the research using Equation 6 - 7 and from all scenarios.

Table 3. False Acceptance Rate (FAR)

Method	Mean of FAR
Joyce and Gupta [5]	0.432
Yang and Fang [6]	0.141
Proposed Method	0.113

Table 4. False Rejected Rate (FRR)

Method	Mean of FRR
Joyce and Gupta [5]	0
Yang and Fang [6]	0.004
Proposed Method	0.009

Based on Table 3, the proposed method has the lowest rate in receiving an attacker as a legitimate user. But on Table 4, it has the highest rate in rejecting a legitimate user to log in. It isn't a problem, because the most important thing in a login system is the ability to reject an attacker to login in the system (lowest FAR). Because the general purpose of the login system using or without KDA is to avoid attackers logging into the system. Besides that, KDA in login system can avoid the loss for many people.

4. Conclusion

Based on the research, it can be concluded that the proposed method used can improve accuracy from the previous method. This research shows that the results of research using MHR are suitable for data streams such as KDA. From the results of FAR testing, shows that the proposed method has the lowest rate (best result) and FRR shows that the proposed method has the highest rate (worst result). The worst results of FRR from the proposed method will not be a problem, because the most important thing in a login system is to reject the attacker to enter the system. However, it is better for the next research to focus on how to reduce the level of FRR in the proposed method. In the next research, a feature selection method can be added to reduce features that are not-important for KDA and can reduce the calculation time when logging into the system.

Notations

feature	: hold time or latency time or flight time
Ø	: threshold per feature
n	: total of training data
M	: average per feature
T	: test data per feature
S_i	: training data *i* per feature
σ	: standard deviation per feature
MHR	: Mean of Horner's Rules per feature
x_n	: training data *n* used for MHR per feature

References

[1] J. Ho and D. Kang*, "One-Class Naïve Bayes with Duration Feature Ranking for Accurate User Authentication Using Keystroke Dynamics",* Applied Intelligence, Vol. 48, No. 6, Pp. 1547-1564, 2018.

[2] C. Lin, J. Liu, K. Lee, *"On Neural Networks for Biometric Authentication Based on Keystroke Dynamics",* Sensors and Materials, Vol. 30, No.3, Pp. 385-396, 2018.

[3] F. Monrose and A. Rubin, *"Keystroke Dynamics as a Biometric for Authentication",* Future Generation Computer Systems, Vol. 16, No. 4, Pp. 351-359, 2000.

[4] M. Ali, J. Monaco, C. Tappert, and M. Qiu, *"Keystroke Biometric Systems for User Authentication",* Journal of Signal Processing Systems, Pp. 1-16, 2016.

[5] R. Joyce and G. Gupta, *"Identity Authentication Based on Keystroke Latencies",* Communications of the ACM, Vol. 33, No. 2, Pp. 168-176, 1990.

[6] W. Yang and F. Fang, *"Application of a Dynamic Identity Authentication Model Based on an Improved Keystroke Rhythm Algorithm",* International Journal of Communications, Network and System Sciences, Vol. 2, No. 8, Pp. 714-719, 2009.

[7] E. Maiorana, P. Campisi, N. Gonzáles-Carballo, and A. Neri, *“Keystroke Dynamics Authentication for Mobile Phones”,* in ACM Symposium on Applied Computing, Pp. 21-26, 2011.

[8] A. Morales, M. Falanga, J. Fierrez, C. Sansone, and J. Ortega-Garcia, *“Keystroke Dynamics Recognition based on Personal Data: A Comparative Experimental Evaluation Implementing Reproducible Research Keystroke Dynamics Recognition based on Personal Data: A Comparative Experimental Evaluation Implementing Reproducible Research”,* in International Conference on Biometrics Theory, Applications and Systems (BTAS), Pp. 1-6, 2015.

[9] D. Stefan and D. Yao, *“Keystroke-Dynamics Authentication Against Synthetic Forgeries”,* in 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2010), Pp. 1-8, 2010.

[10] Ignacio de Mendizabal-V´azquez, Daniel de Santos-Sierra, J. Guerra-Casanova, and C. S´anchez- ´Avila, *“Supervised Classification Methods Applied to Keystroke Dynamics Through Mobile Devices”,* in International Carnahan Conference on Security Technology (ICCST), Pp. 1-6, 2014.

[11] I. Sluganovi´c, A. Karlovi´c, P. Bosilj, M. Šare, and S. Horvat, *“User Authentication Based on Keystroke Dynamics Analysis”,* in MIPRO, 2012 Proceedings of the 35th International Convention, Pp. 2136-2141, 2012.

[12] F. M. Al-Athari and A. Khalaf Husein, *“Selection of the Best threshold in Biometric Authentication by Exhaustive Statistical Pre-Testing”,* International Journal of Computer and Information Technology, Vol.03, No. 04, Pp. 787-791, 2014.

[13] K. S. Killourhy and R. A. Maxion*, “Comparing anomaly-detection algorithms for keystroke dynamics”,* in Proceedings of the International Conference on Dependable Systems and Networks, Pp. 125-134, 2009.

[14] H. Crawford, *“Keystroke Dynamics: Characteristics and Opportunities”,* in Eighth Annual International Conference on Privacy, Security and Trust, Pp. 205-212, 2010.

[15] N. D’Lima and J. Mittal, *“Password Authentication using Keystroke Biometrics”,* in International Conference on Communication, Information & Computing Technology (ICCICT), Pp. 1-6, 2015.