TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 18, No. 2, April 2020, pp. 770~775
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i2.14815

Combined scaled manhattan distance and mean of horner's rules for keystroke dynamic authentication

Didih Rizki Chandranegara, Hardianto Wibowo, Agus Eko Minarno
Informatic Engineering, Universitas Muhammadiyah Malang, Indonesia

Article Info

Article history:

Received Aug 30, 2019
Revised Dec 26, 2019
Accepted Feb 10, 2020

Keywords:

Authentication
Biometric authentication
Keystroke dynamic authentication
Mean of horner's rules
Scaled manhattan distance

ABSTRACT

Account security was determined by how well the security techniques applied by the system were used. There had been many security methods that guaranteed the security of their accounts, one of which was Keystroke Dynamic Authentication. Keystroke Dynamic Authentication was an authentication technique that utilized the typing habits of a person as a security measurement tool for the user account. From several research, the average use in the Keystroke Dynamic Authentication classification is not suitable, because a user's typing speed will change over time, maybe faster or slower depending on certain conditions. So, in this research, we proposed a combination of the Scaled Manhattan Distance method and the Mean of Horner's Rules as a classification method between the user and attacker against the Keystroke Dynamic Authentication. The reason for using Mean of Horner's Rules can adapt to changes in values over time and based on the results can improve the accuracy of the previous method.

Corresponding Author:

Didih Rizki Chandranegara,
Informatic Engineering,
Universitas Muhammadiyah Malang, Indonesia.
Email: didihrizki@umm.ac.id

1. INTRODUCTION

Generally, to access a system service, users need an account that contains a username and password [1, 2]. The main key to securing an account is a password. At present, passwords are one of the popular authentication methods [3-5]. Usually, the contents of the password used by the user contain a variety of information they have (or what they know), such as full name, date of birth to the name of his parents [6, 7]. Passwords are a simple authentication method that is very easy to implement. That is why there are still many systems in cyberspace utilizing this conventional method. Because of its ease of implementation, many ways can be done to guess passwords from system users such as dictionary attack and brute force attacks [8, 9]. However, there is a technique that can be done so that the account is not easily broken into by adding some special characters (example: "<% $ @!") [3]. But, it is not easy to remember for users, because users must remember the characters they use every time they log in to the system [1] and they cannot easy log into the system [9]. And if the user has not used his account for a long time, there is an indication that the user will not be able to log into the system because he/she has forgotten the password used. So, it makes users frustrations because cannot log into the system [10].

Keystroke Dynamic Authentication (KDA) is one of the right solutions in several previous problems. KDA is an authentication technique that utilizes the habit of typing someone as a login parameter from the user of a system [6, 11, 12]. The purpose of KDA is to increase the security of using passwords that have been widely used and handle various account security issues that are often broken into by irresponsible

users (hackers or attackers) [3, 6]. KDA is one of the Biometric Authentication techniques. Biometric Authentication utilizes something unique from users such as the face, fingerprints, and habits (in this case KDA) [3, 6, 13]. And every person face, fingerprints, and habits can not be imitated by others (one of the habits is typing characters using the keyboard or KDA). This also shows that the application of KDA to a system is very safe [14]. Then, the main reason for using KDA in this research is that it does not require expensive costs (low costs) and does not need any additional devices [14-16] (only uses the keyboard). This differentiates KDA with another Biometric Authentication which using adding devices (such as face or fingerprints) [17]. Another advantage of KDA is that the characters used in the password do not have to utilize special characters, but can use the alphabet and numeric characters [6, 15]. Because utilizes the KDA method, users who enter into the system will not realize that the system they are using has used the KDA method for their account security.

There are KDA researches that utilize the Scaled Manhattan method [3, 18]. They utilize the average in the research conducted. The use of averages has a weakness for data streams such as KDA ie the value does not change with time [15]. That is, a user's typing speed will change over time (maybe faster or slower depending on certain conditions). The use of averages is not suitable for this problem, so we propose the use of Mean of Horner's Rules (MHR) which can adapt to changes in values over time. Also, by using MHR on KDA, it can improve accuracy in the classification between attackers and users rather than using averages [6, 15]. So in this research, we will do a combination of the Scaled Manhattan method and MHR to improve accuracy in the classification between attackers and users. And, for more details on the methods used, the final results and discussion of this research can be seen in the next chapter.

2. RESEARCH METHOD

This research uses Scaled Manhattan Distance [3, 19] combined with Mean of Horner's Rules (MHR) [15]. The purpose of this combination is to improve the accuracy of the classification between attackers and users. This has been proven from the results of research from Chandranegara and Sumadi [6] that utilize a combination of MHR and the accuracy of the methods developed is improved compared to the previous method. Where the classification method used without MHR produces an accuracy of approximately 75% and when combined with MHR it becomes approximately 93% (increasing by 18%). While the Dynamic Keystroke data used is derived from the results of Killourhy and Maxion [19]. Following is the formula of the Scaled Manhattan Distance method [3, 19]:

$$\emptyset_n = \sum_1^p \left| f_{p,n} - \overline{g_n} \right| / a_n \tag{1}$$

where p is total of training data and n is a feature of the data. Whereas f_(p,n) is the training data of the *n*th feature with p=1,...p. (g_n)¯ is the average of training data per feature and a_n is the absolute deviation of training data per feature. To get an absolute deviation you can use a formula like the following [3, 19]:

$$a_n = \frac{1}{q} \sum_1^q \left| f_{q,n} - \overline{g_n} \right| \tag{2}$$

where q is total of training data and n is a feature of the data. $f_{q,n}$ is the training data of the *n*th feature with *q=1,...q.* Furthermore, to find the MHR value, the following formula can be used [6, 15]:

$$MHR = \frac{\left(\frac{\left(\frac{\left(\frac{x_1 + x_2}{2} \right) + x_3}{2} \right) + x_4}{\dots} \right) + x_n}{2} \tag{3}$$

where X_n is the *n*th data from the training data.

This research proposes a combination of Mean of Horner's Rules (MHR) which can be seen as follows :

$$\emptyset_n = \sum_{m=1}^m \left| f_{m,n} - MHR_n \right| / (a_n) \tag{4}$$

where this combination is done by replacing the average value with MHR. The purpose of using this MHR is to improve the accuracy of the previous method. This is reinforced from the results of Chandranegara and Sumadi's research [6] which states that accuracy increases by replacing the average using MHR. For classification between attackers and users we use classifications like the following [6]:

- If $|MHR_n - T_n| \leq \emptyset_n$, then the user is considered as an actual user.
- If $|MHR_n - T_n| > \emptyset_n$, then the user is considered as an attacker.

Where *T* is the testing data and *n* is a feature of the testing data.

As an evaluation of KDA method which aims to find out how good the proposed KDA method is in accepting users or rejecting attackers, in this research we use FAR (False Acceptance Rate) and FRR (False Rejected Rate) values [15, 20, 21]. FAR is a possible system/method for accepting an attacker as a user [15, 20, 21]. Whereas FRR is the possibility of a system/method to reject users and detect them as attackers [15, 20, 21]. How to get the FAR and FRR values can be seen in formulas (5) and (6), provided that the smaller the value of the FAR or FRR, the better the results of the KDA classification applied [6, 22].

$$FAR = \frac{number\ of\ acceptance\ attacker}{total\ number\ of\ attacker} \quad (5)$$

$$FRR = \frac{number\ of\ rejected\ user}{total\ number\ of\ user} \quad (6)$$

In addition to FAR and FRR, we also evaluate using accuracy with the following formula [6]:

$$Accuracy = \left(\frac{TP+TN}{TP+FP+TN+FN}\right) x100\% \quad (7)$$

where TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). TP is the success to accept users as actual users and TF is the success to detect attackers. Whereas FP is a misclassification for accepting an attacker and detecting it as a actual user and FN is a misclassification for refusing an actual user and detecting it as an attacker. To get the accuracy value as explained before, we use several scenarios like the following:

a. Dynamic Keystroke Data is divided into 2 types i.e. training data and testing data.
b. Data Training for every user is the first 350 data from a dataset. For illustration training, User "A" training data uses data from 1 to 350, from a total of 400 KDA data.
c. Data Testing for every user is the last 50 data from a dataset. For illustration testing, User "A" uses KDA data from 350 to 400, from a total of 400 KDA data as testing data.
d. Furthermore, each user in this data will be used as an attacker for every other user. Thus, as many as 51 attack scenarios will be formed (where the total users of the data used are 51 people). And the attacker data used is the last 50 data of dataset that is used as an attacker.

Testing methods used in this research use a program (using php programming) that is made in accordance with the proposed method and previous methods and adapted to predetermined scenarios.

3. RESULTS AND ANALYSIS

3.1. Dataset

This research uses Keystroke Dynamic data from Killourhy and Maxion [19]. In this data, there are 51 users (30 male and 21 female) and each user has 400 Dynamic Keystroke data. This data was obtained by them within 8 days, where every day obtained Dynamic Keystroke data as much as 50 data perusers. The time used in this data is seconds. The character used in Keystroke Dynamic data recording is "tie5Roanl". The use of this character has also been based on several attempts and the result is that this character has a high level of password security. In this data, each user types characters and records them. There are several important feature elements contained in this data, i.e. [6, 7, 20, 23, 24, 25] (illustration of these important features and contained from the data used can be seen in Figure 1):

a. Hold time (H) is the time needed to press a character (Key-Down to Key-Up).
b. Up-Down (UD) is the time between releasing (Key-Up) characters to pressing the next (Key-Down) character or commonly referred to as Latency Time.
c. Down-Down (DD) is the time taken when pressing the first character (Key-Down) to press the second character (Key-Down) or commonly referred to as Flight Time.

Total features in this data are 31 features. Where each feature consists of:

a. Holdtime character ".", Character "t" to the last character that is "l" and pressing the "return" button is also included. So that the total is 11 features.
b. Up-Down (UD) characters "." And "t" to UD between the last character with the "return" button. So that the total is 10 features.
c. Down-Down (DD)/Flight Time between the characters "." Until the last character, "l" and pressing the "return" button are also entered. So the total is 11.

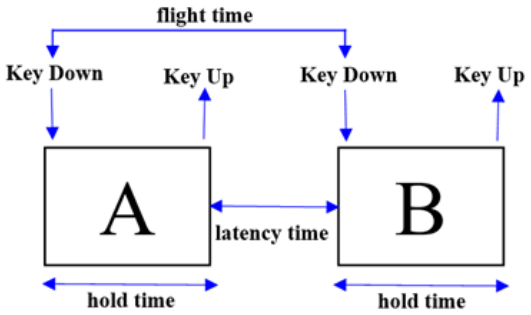

Figure 1. Illustration of KDA Features [6]

3.2. Result and analysis

Based on the test results, the proposed method produces a less good accuracy of 50.113%. While the accuracy of the previous method is 50.335%. However, the FAR value of the proposed method has decreased from the previous method (the proposed method has a FAR value of 0.976 and the previous method has a FAR value of 0.98). The increase does not occur at accuracy but in the FAR value. Because it has not shown better accuracy, we have modified the proposed method by adding coefficient 5 and can be seen in formula (7).

$$\emptyset_n = \sum_{m=1}^{m} \left| f_{m,n} - MHR_n \right| / (5 * a_n) \quad (7)$$

After the modification, the accuracy is quite high. The reason for using the coefficient number 5 is based on several experiments using other coefficients from 1 to 7 (The results of the coefficient experiment can be seen in Table 1). Based on the test results, it appears that coefficient 5 has FAR and FRR values of 0.356 and 0.305 see Table 1. The FAR and FRR values of the coefficient 5 show almost the same value and can be said to be balanced. Whereas in other coefficients, the FAR value is low but the FRR value is high and vice versa, the FRR value is low but the FAR value is high.

Table 1. Result of using coeffesien in proposed method

Coeffesien	FAR	FRR
1	0.976	0.021
2	0.863	0.082
3	0.673	0.172
4	0.490	0.241
5	0.356	0.305
6	0.251	0.367
7	0.179	0.425

There are reasons why we don't use other coefficients that have the lowest FAR or FRR values, i.e.:

a. If the FAR value is high, then the possibility of the system accepting the attacker as an actual user is higher.
b. If the FRR value is high, then the possibility of the system rejecting actual users or assuming actual users as attackers are higher.

These two reasons are our main benchmarks for using coefficient 5. Also, this reason is based on the results of previous studies [6, 15]. The results of the test in the form of accuracy using the proposed method given coefficients and the previous method have been presented in Table 2. And the results of this test are the average accuracy obtained from 51 preplanned scenarios.

Table 2. Results of Research

Method	Average of Accuracy (%)
Scaled Manhattan Distance	50.335
Combined Scaled Manhattan Distance with MHR (Coeffesien=5)	66.963

4. CONCLUSION

Based on the results of the research conducted, it appears that the accuracy of the proposed method has not increased compared to the previous method. This is because the value of the Scaled Manhattan Distance Modification produced is less suitable for accepting users and rejecting attackers. So we try to add coefficient to increase the value. And the results show that its accuracy can be increased even if not significantly. And the best coefficient used in this proposed method is number 5. Because based on the results of tests conducted previously, shows that coefficient 5 gives the smallest FAR and FRR values compared to other coefficients. However, although accuracy does not increase if it does not add coefficient, this proposed method can reduce the FAR (false acceptance rate). Means, the proposed method without coefficient has a good result on FAR but not on the accuracy value.

In next research, it is expected to be able to add feature selection so that the computational classification is reduced and can also select features that are considered important in Keystroke Dynamic Authentication. Also, we can do some modifications to other methods that apply averages as their classification. And based on our research, the accuracy value cannot be used as a benchmark that the method is good or not, but we can use other parameters besides accuracy as in the case of KDA namely the FAR and FRR values. Then based on the results of this research, this proposed method can be applied to real or desktop-based login systems. And users will not be aware if the login method has been applied Keystroke Dynamic Authentication security.

ACKNOWLEDGEMENTS

This research is supported by Laboratorium Informatika Universitas Muhammadiyah Malang. The authors wish to thank Universitas Muhammadiyah Malang for providing the funding.

REFERENCES

[1] Zahid S., Sean B., Bojan C., "Normalizing variations in feature vector structure in keystroke dynamics authentication systems," *Software Quality Journal*, vol. 24, pp. 137-157, 2016.
[2] Romain G., Mohamad E., Christophe R., "Greyc keystroke: a benchmark for keystroke dynamics biometric systems," 3rd *International Conference on Biometrics: Theory, Applications, and Systems*, vol 6, Washington, 2016.
[3] Aythami M., Mario F., Julian F., Carlo S., Javier O., "Keystroke dynamics recognition based on personal data: A comparative experimental evaluation implementing reproducible research," *International Conference on Biometrics Theory, Applications and Systems (BTAS),* 2015.
[4] Lin C. H., Liu J, C., Lee K. Y., "On Neural Networks for Biometric Authentication Based on Keystroke Dynamics," *Sensors and Materials*, vol. 30, no. 3, pp. 385-396, 2018.
[5] Antal M., Szabó L. Z., "An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices," 20th *International Conference on Control Systems and Computer Science*, 2015.
[6] Chandranegara DR, Sumadi FDS., "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*, vol. 4, no. 1, pp. 13-18, 2018.
[7] Al-Athari F. M., Hussain A. K., "Selection of the Best threshold in Biometric Authentication by Exhaustive Statistical Pre-Testing," *International Journal of Computer and Information Technology*, vol. 3, no. 4, 2014. (ISSN: 2279–0764)
[8] Kobojek P., Saeed K., "Application of recurrent neural networks for user verification based on keystroke dynamics," *Journal of telecommunications and information technology*, vol. 3, pp. 80-90, 2016.
[9] Alsultan A., Warwick K., Wei H., "Improving the performance of free-text keystroke dynamics authentication by fusion," *Applied Soft Computing*, vol. 70, pp. 1024-1033, 2018.
[10] Alsultan A., Warwick K., Wei H., "Non-conventional keystroke dynamics for user authentication*," Pattern Recognition Letters*, vol. 89, pp. 53-59, 2017.
[11] Pinto P., Patrão B., Santos H., "Free typed text using keystroke dynamics for continuous authentication," *IFIP International Conference on Communications and Multimedia Security*, Springer, 2014.
[12] Morales A., Fierrez J., Tolosana R., Ortega-Garcia J., Galbally J., Gomez-Barrero M., et al., "Keystroke biometrics ongoing competition," *IEEE Access*, vol. 4, pp. 7736-7746, 2016.

[13] Rybnik M., Panasiuk P., Saeed K., Rogowski M., editors, "Advances in the keystroke dynamics: the practical impact of database quality," *IFIP International Conference on Computer Information Systems and Industrial Management*, 2012.
[14] Chandrasekar V., Kumar S. S., Maheswari T., "Authentication based on keystroke dynamics using stochastic diffusion algorithm*," Stochastic Analysis and Applications*, vol. 34, no. 1, pp. 155-164, 2016.
[15] Ho J., Kang D. K., "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.
[16] Ivannikova E., David G., Hämäläinen T., "Anomaly detection approach to keystroke dynamics-based user authentication," *2017 IEEE Symposium on Computers and Communications (ISCC)*, 2017.
[17] Chang T. Y., Tsai C. J., Tsai W. J., Peng C. C., Wu H. S., "A changeable personal identification number-based keystroke dynamics authentication system on smart phones," *Security and Communication Networks*, vol. 9, no.15, pp. 2674-2685, 2016.
[18] Araújo L. C., Sucupira L. H., Lizarraga M. G., Ling L. L., Yabu-Uti J. B. T., "User authentication through typing biometrics features," *IEEE transactions on signal processing*, vol. 53, no. 2, pp. 851-855, 2005.
[19] Killourhy K. S., Maxion R. A., "Comparing anomaly-detection algorithms for keystroke dynamics," *2009 IEEE/IFIP International Conference on Dependable Systems & Networks*, 2009.
[20] Teh P. S., Teoh A. B. J., Yue S., "A survey of keystroke dynamics biometrics," *The Scientific World Journal*, 2013.
[21] Wu P. Y., Fang C. C., Chang J. M., Kung S. Y., "Cost-effective kernel ridge regression implementation for keystroke-based active authentication system," *IEEE transactions on cybernetics*, vol. 47, no. 11, pp. 3916-3927, 2016
[22] Jagadamba G., Sharmila S., Gouda T., "A secured authentication system using an effective keystroke dynamic," *Emerging research in electronics, computer science and technology*, Springer, pp. 453-460, 2014.
[23] Al-Jarrah M. M., "An anomaly detector for keystroke dynamics based on medians vector proximity," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 3, no. 6, pp. 988-993, 2012.
[24] Alghamdi SJ, Elrefaei LA, "Dynamic user verification using touch keystroke based on medians vector proximity," *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*, 2015.
[25] Wankhede SB, Verma S., "Keystroke dynamics authentication system using neural network," *International Journal of Innovative Research and Development*, vol. 3, no. 1, pp. 157-164, 2014.