Adaptive Behaviometric for Information Security and Authentication System using Dynamic Keystroke

by Dewi Yanti Liliana

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 10, No. 1, 2012

Adaptive Behaviometric for Information Security and Authentication System using Dynamic Keystroke

Dewi Yanti Liliana
Department of Computer Science
University of Brawijaya
Malang, Indonesia
dewi.liliana@ub.ac.id; dewi.liliana@gmail.com

Dwina Satrinia
Department of Computer Science
University of Brawijaya
Malang, Indonesia
dwina.satrinia@gmail.com

***Abstract*—The increasing number of information systems requires a reliable authentication technique for information security. Password only is not enough to protect user account because it is still vulnerable to any intrusion. Therefore an authentication system using dynamic keystrokes can be the simplest and the best choice. Dynamic Keystroke Authentication System (DKAS) becomes an effective solution which can be easily implemented to gain a high security information system with the aid of a computer keyboard. DKAS verify users based on their typing rythm. Two main stages of DKAS is the enrollment stage to register user into the system, and the authentication stage to check the authenticity of user. Moreover, we use a local threshold to make it becomes adaptive behaviometric for each user. From the experiment conducted, the accuracy rate in distinguishing genuine and impostor user is 91.72%. This shows that the adaptive method of DKAS has a promising result.**

Keywords- authentication system, behaviometric, dynamic keystroke, local threshold

I. INTRODUCTION

The increasing use of information systems in any fields causes a high-demand on a reliable authentication system for information security. Authentication based on biometrics is widely used because of its robustness. Biometrics is a method to recognize human based on intrinsic features or characteristics human has [1]. Physiological biometrics uses unique physical characteristics of individual such as fingerprint, face, palm print, iris, or DNA to identify user and has proven to be a powerful method for authentication systems [1, 2, 3]. Nevertheless, these systems need additional devices (e.g. camera, fingerprint reader, microphone, etc.) to capture human features. Meanwhile, behavioral traits of human or so-called behaviometric which is related to human behavior [4, 5], such as typing rhythm or typing pattern can be implemented on authentication systems without any additional devices. This research implemented behaviometric for authentication system using dynamic keystroke which only needs a computer keyboard to capture the distinct features on typing.

In 2005, Hocquet et.al, conducted a research on authentication system using the combination of password and dynamic keystroke which incorporated three methods; statistical measurement, measure of disorder, and direction similarity measure [5]. The combination method was simple, needed only small size training data, and used global threshold for classifying genuine and impostor users. Global threshold is a constant threshold for all users. The problem was to determine this constant value based on prior knowledge of data. In this research we propose a local threshold setting which can be adaptively adjusted for each different user. Local threshold is adopted from the average score of each user which is obtained during the enrollment phase.

II. DYNAMIC KEYSTROKE AUTHENTICATION SYSTEM

Keystroke means key press. While dynamic keystroke is a biometric which concern about how a user interacts with a keyboard, typing rhythm of a person associated with the habit of typing the password, words, or text [6]. It requires only a keyboard as an input device. Dynamic keystroke also can be implemented for remote access. In addition, biometric based on dynamic keystroke can be used with or without user consciousness.

Password is commonly used on an authentication system for its simplicity, but is less secure because vulnerable to some kinds of attack such as key loggers, spyware, and can be hacked using simple brute force techniques. To enhance the system security and cost efficiency, the password-based authentication system can be combined with dynamic keystroke authentication system (DKAS).

There are two stages on DKAS to distinguish between genuine and impostor user namely, the enrollment stage and the authentication stage (see fig. 1).

At the enrollment stage user sign up their login details such as user name and password which is retyped for several times. The system takes the user dynamic keystrokes ten times for each enrollment, extracts the features, and trains the system to create a reference template of user's typing pattern. The reference template is stored in a database. At the authentication stage, the user enters the login details to be matched with user's reference template which is already stored in the database. This phase consists of collecting user dynamic keystrokes, feature extraction, and feature matching with reference template in the database. The verification process yields two kinds of action: accepted or rejected user access. The first action occurs when the user is the genuine one, while the other action occurs for the impostor user.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

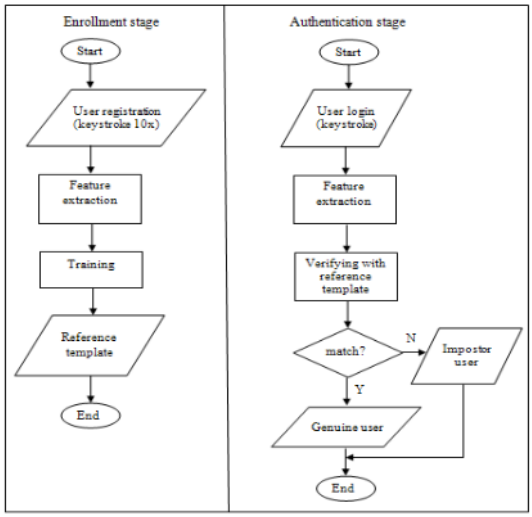


Figure 1. Flowchart of Dynamic Keystroke Authentication System

Four dynamic keystrokes used as features for the authentication system can be seen on illustration of fig. 2.

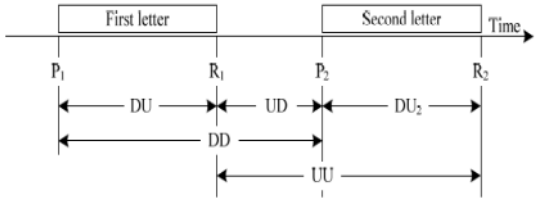


Figure 2. Features of Dynamic Keystroke

Those four features are explained bellow:

1. PP (Press-Press) or DD (down-down) or digraph1: the time between one key press and the next key press (P2-P1).
2. PR (Press-Release) or DU (down-up) or duration: the length of key press (R1-P1).
3. RP (release-press) or UD (Up-down) or latency: the time between key release and the next key press (P2-R1).
4. RR (release-release) or UU (up-up) or digraph2: the time between key release and the next key release (R2-R1).

III. METHODOLOGY

The initial step in this paper is started with the formation of reference templates. Moreover, three methods namely, statistical scoring, measure of disorder, and direction similarity measure will be performed. The last step is the adaptive local threshold setting.

A. The Formation of Reference Templates

In order to verify a user based on dynamic keystrokes, the system needs to create a model or reference template for each user. Reference template is a combination of user keystrokes acquired during the enrollment process which is converted into a more solid form, but still can represent a user keystroke patterns [7]. This research utilized a statistical mean and standard deviation for the reference template formation which can be obtained using equation 1 and 2, respectively.

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} t_x^i \quad (1)$$

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_x^i - \mu_x)^2} \quad (2)$$

where $i=1,2,...,n$ is the number of training samples, $x=1,...,m$ is the number of features used, t_x^i denotes the feature x on the sample i, μ_x and σ_x denote mean and standard deviation of feature x, respectively.

B. Statistical scoring

In the verification process feature matching is performed. It compares the feature of the user test data with the reference template that has been formed on the enrollment stage. Statistical scoring is employed for feature matching. This method will verify the user based on statistical data such as mean and standard deviation. The equation for calculating statistical score is written in Eq.3:

$$Score_{stat} = \frac{1}{n}\sum_{i=1}^{n} e^{-\frac{|t_i-\mu_i|}{\sigma_i}} \quad (3)$$

where $t_i=1,...,n$ is the i-th test feature, e is a constant with value of 2.71828, μ_i and σ_i denote mean and standard deviation of reference template vector, respectively.

C. Measure of Disorder

Measure of disorder method is used to compare two ways of typing on the keyboard by studying the similarity between sequences of time features generated as reference templates with sequences of time features which is being tested [8].

To compute the distance between the user keystroke input with the reference template then several steps must be carried out as follows:

1. Rate or rank individual features of each user keystroke input and the comparison reference template. Ordering is done from the smallest to the largest feature value.
2. Calculate the magnitude of differences in rank order or ranking of any existing features on the template with user ratings on keystroke input
3. Calculate the score of disorder using equation 4.

$$score_{disorder} = 1 - \frac{\sum_{i=1}^{n}|R_i^t - R_i^u|}{Max_{disord\ er}} \quad (4)$$

where R_i^t is the i-th feature rank obtained from rank vector, R_i^u is the i-th feature rank obtained from the user input, and N denotes the number of element or existing features which hold two condition as follows: $Max_{disor\ der} = \frac{N^2}{2}$ if N is even; and

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$Max_{disorder} = \frac{N^2-1}{2}$ if N is odd.

D. Direction Similarity Measure

Direction similarity measure (DSM) is a simple approach that is discriminatively compares user's typing patterns. The idea of this method is to determine the consistency of the user typing habit. This idea is adopted from the rhythm of the music [8]. In music where the rhythm of a melody is determined by the duration of a tone (the tone is full, half, quarter, etc.), the keystroke is represented by the dynamic rhythm of ups and downs or how quick a keystroke is pressed.

In the calculation of DSM, there is a ΔD symbol which is used as a sign of change in the direction of two successive keystrokes. As an example, ΔD is positive if there is any time reduction between two keystrokes (faster), and ΔD is negative if there is any additional time between two keystrokes (slower). Figure 3 shows the ΔD signing.

DU1	DU2	DU3	DU4
245	297	326	268
ΔD :	-1	-1	+1

Figure 3. An example of ΔD signing

DSM score can be calculated using the equation 5:

$$Score_{DSM} = \frac{m}{n-1} \quad (5)$$

where m is the number of ΔD which has the same sign, and n is the total features. To compare the user keystroke template with the user keystroke input, what must be considered is the change in sign of ΔD. If the sign of ΔD from the user reference template equal to the value of ΔD of user keystroke input, then the value of m increases. The final value of m is divided by the number of features minus 1.

E. The incorporation of methods

In this paper the three methods (statistical scoring, measure of disorder, and direction similarity measure) are incorporated by using scoring level which will be done using weighted sum rule operator. The final merged score can be calculated with equation 6:

$$score_{final} = \Sigma(w_i * score^i) \quad (6)$$

where Σw_i=1, score^1 = statistical score; score^2 = measure of disorder score; skor^3 = DSM score.

If the $score_{final}$ of the test user is greater than the user threshold value, then the user will be recognized as a genuine user. Otherwise, it will be recognized as an impostor.

F. Local Threshold

The threshold for the verification system is similarity value between the test inputs with the model. If the results of feature matching score threshold, then the user is recognized as an impostor, and if the results of feature matching score ≥ threshold, then the user is recognized as an actual or genuine user.

There are two kinds of threshold, global and local threshold. The global threshold value is set equal to all users, and the local threshold value is set specifically to each user. The problem is to determine the global threshold value required prior knowledge of the data. Therefore, the determination of local threshold value can reduce the problem. Moreover, local threshold can be adaptively adjusted for each different user. There are some ways to estimate local threshold value can be chosen, using the actual user data, impostor data, or a combination of both. The equation used to determine the local threshold value is on Eq. 7:

$$\theta = \mu_{user} - \alpha.\sigma_{user} \quad (7)$$

where θ denotes local threshold, $\mu_{user}, \sigma_{user}$ denotes mean and standard deviation score from user enrollment, respectively, and α denotes a constant factor obtained from the experiment.

The determination of threshold values from user registration data is easy to implement but is less effective because sometimes when the user on registration gets disorders such as drowsiness, talk to or in any uncomfortable situations that are bothering in dynamic keystroke patterns representation. If the threshold was estimated on a situation like this, it will result in decreased accuracy in recognizing user's system. To overcome this problem, we used a method to estimate the weighted scores of local threshold value.

Weighted score is a method to estimate the threshold that gives the weights on the scores based on distance from the user's score to the average score [9]. Scores that were located far from the average are considered as outliers of the user which might be due to a disturbance when users type a password in the registration process. Weighting factor w_i is the parameter of the sigmoid function. w_i values can be calculated by the equation 8:

$$w_i = \frac{1}{1+e^{-C.d_i}} \quad (8)$$

Where C is a constant empirically gained from the experiment with the best value = -3. d_i denotes the distance of score$_i$ to the average score ($d_i = |score_i - \mu_{score}|$). Thus, we got the final score S_T by using equation 9:

$$S_T = \frac{\sum_{i=1}^{N} w_i.Score_i}{\sum_{i=1}^{N} w_i} \quad (9)$$

The constant C determines the shape of the sigmoid function used to set the weights. $score_i$ and μ_{score} of the training set obtained by a leave-one-out approach. Standard deviation is calculated from $score_i$ against weighted score S_T. The S_T value will replace the μ value of user, and the standard deviation of weighted score will replace the σ user in determining the threshold value. Here are steps on leave-one-out to get $score_i$ value:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

1. Take a feature vector of n feature vectors used as input during registration for the test.
2. Create a comparison matrix of n-1 remaining feature vectors, then create a reference template of the comparison matrix
3. Compare the test input in step 1 with a reference template that is formed in step 2, using the method used in the verification process to get $score_i$.
4. Repeat steps 1-3 with all possible combinations of the features found on other user registration data so as to produce n numbers of $score_i$.
5. Calculate μ_{score} which is an average score of the comparison.

IV. EXPERIMENTS AND RESULTS

Tests carried out using two groups of data that is a typing sample based on user passwords. The first group is users with passwords that usually have been typed by them e.g. their name, etc. The second group is users who use unusual typed words as the password or words chosen at random. Each group consists of the actual and impostor user.

System performance is measured using two error rate: False Rejection Rate (FRR), describes the percentage of a biometric system fails to recognize the actual user and False Acceptance Rate (FAR), describes the percentage of the biometric system identifies incorrect impostor as the actual user. To measure the accuracy of the system, we also measure the Equal Error Rate (EER) obtained when FAR value is equal to FRR (in other words, the intersection of FRR and FAR line). EER is used to compare the performance of different biometric systems [5].

The experiment conducted three kinds of testing: weight value testing that produced the lowest EER value; testing the accuracy of a system that used a local threshold; and testing a system using a global threshold. All tests were using two different groups of data as well as the overall data.

Based on tests done on 826 typed samples, the resulting value of the lowest EER is 8.22%, obtained when the score of statistical weight is 0.7, and the weight score of measure of disorder (MOD) & DSM are 0.15 respectively (see Fig. 4).

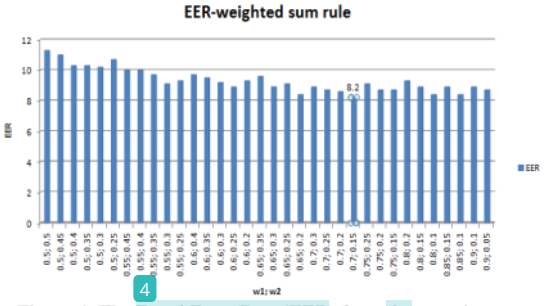


Figure 4. The Equal Error Rate (EER) from the experiment.

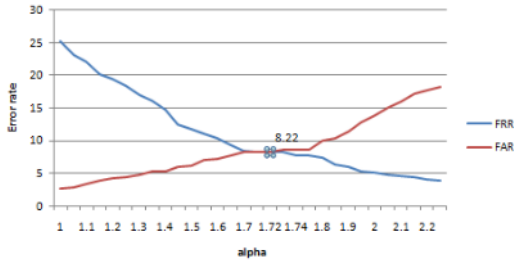
The accuracy rate of the authentication system with local and global threshold setting is shown in Table I.

TABLE I. THE ERR COMPARISON OF LOCAL AND GLOBAL THRESHOLD

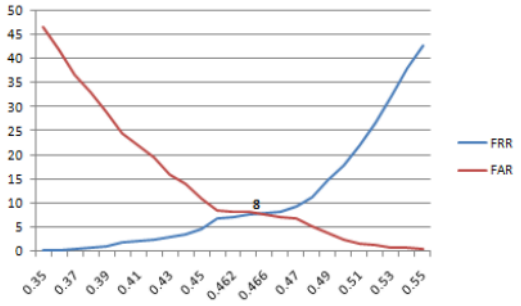
Data	EER (%)	
	Local	Global
all data	8.22	8
Group 1	4.49	4
Group 2	12	10

From the test result (see table 1), it can be seen that the EER test in group 1 (table 1 row 4) is significantly lower than group 2 (table 1 row 5). This shows that the accuracy rate of dynamic keystroke authentication system depends on the choice of words as passwords. The more accustomed the user with the word, the more the ability of system to recognize users.

From the experiment of comparing global and local thresholds, we got the result which is shown as graphs of error rate in fig. 5. The EER for local threshold is 8.22% with the accuracy rate 91.72%, obtained when the value of α is 1.71. While the EER for global threshold is 8% with the accuracy rate 92%, using the global threshold value = 0.466. When compared with a global threshold, the accuracy rate of a system that uses a local threshold can be said is equally better in verifying the user. The advantages of setting a local threshold is the threshold value for each user can be adaptively estimated using the user data only from the registration process, even without prior knowledge of the data.



(a)



(b)

Figure. 5. Graphs of error rate (a) Local Threshold (b) Global Threshold

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. CONCLUSION

Dynamic keystroke authentication system is able to verify the user using statistical method, measure of disorder, and direction similarity measure that recognized the user based on the adaptive local threshold. The use of the word or phrase as a password influences the accuracy rate of the system. The accuracy of the system using the local threshold is 91.72%, obtained when the value of α is 1.71.

REFERENCES

[1] N.K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems", IBM systems Journal, vol. 40, pp. 614-634, 2001.

[2] S. Tulyakov, F. Farooq, and V. Govindaraju, "Symmetric Hash Functions for Fingerprint Minutiae", Proc. Int'l Workshop Pattern Recognition for Crime Prevention, Security, and Surveillance, pp. 30-38, 2005.

[3] M.A. Dabbah, W.L. Woo, and S.S. Dlay, "Secure Authentication for Face Recognition", presented at Computational Intelligence in Image and Signal Processing, CIISP 2007, IEEE Symposium, 2007.

[4] http://biosecure.it-sudparis.eu/public_html/biosecure1/public_docs_deli/BioSecure_Deliverable_D10-2-3_b3.pdf

[5] Hocquet, Sylvain, J. Ramel and H. Cardot, "Fusion of Methods for Keystroke Dynamic Authentication", Fourth IEEE workshop on Automatic Identification Advance Technology, 2005.

[6] Hocquet, Sylvain, Jean-Yves Ramel & Hubert Cardot, "User Classification for Keystroke Dynamics Authentication", International Conference on Biometric, Springer-Verlag Berlin Heidelberg. Page 531-539, 2007.

[7] P.S. Teh, B.J.T. Andrew, T. Connie, and S.O. Thian, "Keystroke dynamics in password authentication enhancement", Expert Systems with Application,Vol. 37, Page 8618-8627, 2010.

[8] F. Bergadano, D. Gunetti, and C. Picardi, "User Authentication through Keystroke Dynamics", ACM Transactions on Information and System Security (TISSEC), Page 367-397, New York: ACM New York, 2002.

AUTHORS PROFILE

Dewi Yanti Liliana obtained Bachelor of Informatics from Sepuluh Nopember Institute of Technology Surabaya, Indonesia, in 2004, and Master of Computer Science from University of Indonesia, Depok, Indonesia, in 2009. She is currently working as a Lecturer for the Department of Computer Science, Faculty of Mathematics and Natural Sciences, University of Brawijaya Malang, East java, Indonesia. Her research interests include pattern recognition, biometrics system, computational algorithm, computer vision and image processing.

Dwina Satrinia is a graduate student at the Department of Computer Science, Faculty of Mathematics and Natural Sciences, University of Brawijaya Malang, East java, Indonesia. Her research interests include pattern recognition and biometrics system.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Adaptive Behaviometric for Information Security and Authentication System using Dynamic Keystroke

ORIGINALITY REPORT

14%	10%	11%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Multimedia University Student Paper	3%
2	Submitted to Universitas Brawijaya Student Paper	2%
3	www.ijeat.org Internet Source	2%
4	"Advances in Biometrics", Springer Science and Business Media LLC, 2007 Publication	1%
5	Pin Shen Teh, Shigang Yue, Andrew B.J. Teoh. "Improving keystroke dynamics authentication system via multiple feature fusion scheme", Proceedings Title: 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 Publication	1%
6	H. Cardot. "Fusion of Methods for Keystroke Dynamic Authentication", Fourth IEEE	1%

Workshop on Automatic Identification Advanced Technologies (AutoID 05), 2005
Publication

7	centaur.reading.ac.uk Internet Source	1 %
8	Pin Shen Teh, Andrew Beng Jin Teoh, Connie Tee, Thian Song Ong. "A multiple layer fusion approach on keystroke dynamics", Pattern Analysis and Applications, 2009 Publication	<1 %
9	Z. S. Shariatmadar, K. Faez. "An Efficient Method for Finger-Knuckle-Print Recognition by Using the Information Fusion at Different Levels", 2011 International Conference on Hand-Based Biometrics, 2011 Publication	<1 %
10	www.archive.org Internet Source	<1 %
11	Pin Shen Teh, Andrew Beng Jin Teoh, Thian Song Ong, Han Foon Neo. "Statistical Fusion Approach on Keystroke Dynamics", 2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, 2007 Publication	<1 %
12	www16.us.archive.org Internet Source	<1 %

13	researchportal.port.ac.uk Internet Source	<1%
14	"On keystrokes as Continuous User Biometric Authentication", International Journal of Engineering and Advanced Technology, 2019 Publication	<1%
15	Augustin-Catalin Iapa, Vladimir-Ioan Cretu. "Modified Distance Metric That Generates Better Performance For The Authentication Algorithm Based On Free-Text Keystroke Dynamics", 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2021 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		