Selection of the Best threshold in Biometric Authentication by Exhaustive Statistical Pre-Testing

Faris M. Al-Athari
Professor and Head, Department if Mathematics
Faculty of Science and Information Technology
Zarqa University, Jordan
Email: f-athari {at} zu.edu.jo

Abdulameer Khalaf Hussain
Department of Computer Science
Jerash University – Faculty of IT
Jerash, Jordan

Abstract— **This paper presents a new approach to identify and improve the best key typing speed of the user when he/she enters the password in order to be used for authentication purpose. Keystroke authentication is considered an advanced authentication method in that it depends on the biometric behavior.**
This idea is based on the several times of pretesting for each new user to type the password before the actual enrollment of the character information of the password. For example, the first pretest may enter the user 10 times and then selecting the best speed. The process continues to subject the user for more other pretest with more trials numbers. After each trial, the user is subjected to type his/her password once. Learned from these attempts to accustom the user to be familiar with the password and then we can get the best speed of typing which is used to distinguish the authenticated user from others. This method depends on different statistical parameters to ensure the best authenticated user. The pretest of each new user allows us to get the best typing speed and so this method is different from other keystroke biometric methods.

Keywords- Authentication, Biometric Keystroke, Pre-test, Typing speed, Statistical biometric measurements.

I. INTRODUCTION

The importance of authentication systems is that they allow different entities to be recognized before using resources. The traditional authentication methods use the classical couple of username and password. This type of schemes, which is based only on one factor, suffers from various security holes [1]. To ensure a strong authentication we must use multiple authentication factors to improve security. In this case, individuals are authenticated with the help of at least two authentication methods using one or several different factors among: (1) something we know; (2) something we have; (3) something we are.

Biometric systems can play an important role to provide the strong authentication scheme by providing the factor what we are when used with one of the two other factors. We can provide strong authentication in the password authentication scheme (what we know) by combining it with keystroke dynamics [2], which is a behavioral biometric modality monitoring the way individual's type on the keyboard (what we are). There are several types of key stroke dynamics systems and are generally based on very long texts [3], passwords [4] or shared secrets [5] although several studies used a shared secret without referring to this term.

The biometric sample can be captured by two methods either statically (i.e., at login phase) or continuously (i.e., during the computer session). If a scheme uses a shared secret it means that all users use the same password. The system always acts as an authentication system, because only a certain group of people is aware of this secret (what we know) while all the members of the group type it differently (what we are).

During the verification phase, the system checks if the password is the required one, if it is not, the user is rejected, otherwise, the system checks if the keystroke dynamics match. The objective of biometric systems is to verify the identity of an entity which can access to a resource. In the case of physical access, this resource can be a building or a room, whereas in the case of logical access, this resource can be an application on a computer.

There are three main families of biometric modalities:
1: Biological: recognition based on the analysis of biological data linked to an individual (e.g., DNA, EEG analysis,).
2: Behavioral: based on the analysis of an individual behavior while performing a specific task (e.g., keystroke dynamics, signature dynamics, gait,).
3: Morphological: based on the recognition of different physical patterns, which are, in general, permanent and unique (e.g., fingerprint, face recognition,).

Most biometric authentication systems are generally composed of two main modules: (a) the enrollment module which consists in creating a template (or reference) for the user with the help of one or several biometric captures (or samples), and (b) the verification module which consists in verifying if the provided sample belongs to the claimed user by comparing it with its template. After verification phase, the system takes a decision to decide to accept or to reject the user depending on the result of the comparison. In addition we can

This study is funded by the Deanship of research and Graduate studies in Zarqa University/ Jordan, Grant No. 1/1/2/202.

use an optional (c) adaptive module which updates the template of a user after a successful authentication.

Many works have already been done on the evaluation of biometric systems [6, 7, 8].

The evaluation of biometric systems may be performed within three different aspects:
1: Performance: the objective is to measure various statistical criteria on the performance of the system (Capacity [9], Equal Error Rate (EER), Failure To Enroll (FTE), Failure To Acquire (FTA), computation time, Receiver Operating Characteristic (ROC) curves, False Acceptance Rate (FAR), False Rejection Rate (FRR) etc [6]).

2: Acceptability and user satisfaction: this gives some information on the individuals' perception, opinions and acceptance with regard to the system [6, 10].

3: Security: this quantifies how well a biometric system (algorithms and devices) can resist several types of logical and physical attacks such as Denial of Service (DoS) attack or spoofing or mimicking attacks [11].

The most metrics used for performance evaluation are:
1: FAR False Acceptance Rate which represents the ratio of impostors accepted by the system;
2: FRR False Rejection Rate which represents the ratio of genuine users rejected by the system;
3: EER Equal Error Rate which is the error rate of the system when it is configured in order to obtain a FAR value equal to the FRR one.

The three aspects (performance security, acceptability and user satisfaction) should be taken into account simultaneously when comparing different biometric systems: we cannot say that a system is good if it provides very low error rates (i.e., very good performance) but has a very low user acceptance (i.e., a high probability to be refused by users) [12, 13].

The main objective of any keystroke dynamics system is to enhance more security for password-based authentication systems which suffer of many drawbacks [14]: (i) passwords can be shared between users, (ii) passwords can be stolen (written on a piece of paper, from the database where it is stored, through network sniffing, ...), (iii) passwords can be guessed (social engineering [15]). Keystroke dynamics introduces an additional parameter to the password authentication process which is something that qualifies the user or his behavior (i.e., the way of typing passwords). Using this additional parameter strengthens the password authentication.

II. RELATED WORKS

One of researches introduces a k-nearest neighbor approach in order to make a classification of users' keystroke dynamics profiles. To provide an authentication it is necessary to check an input against the profiles within the cluster which has significantly reduced the verification load. **[16]**

In [17] an investigation of user authentication using keystroke biometric. The authors proposed an effective new distance metric to deal with keystroke dynamic data such as scale variations, feature interactions and redundancies. This new distance metric are evaluated on the CMU keystroke dynamics benchmark dataset and are shown to be superior to algorithms using traditional distance metrics.

A method of using inputs which are down and up times and the key ASCII
Codes captured while the user is typing a string are proposed. The authors analyzed four features (key code, two keystroke latencies, and key duration) and seven experiments were performed combining these features. The best results were achieved utilizing all features, obtaining a false rejection rate of 1.45% and a false acceptance rate of 1.89%. [18].

In [19] a study was performed develop and evaluate techniques to authenticate valid users, using the keystroke dynamics of a user's PIN number entry on a numerical keypad, with force sensing resistors. Added with two conventional parameter lists of elements, i.e. digraph latency times and key hold times, keying force was chosen as a third element. The study conducted two experiments. The first experiment was to evaluate whether the three types of elements derived from keystrokes have a significant effect for subjects and to examine how trials and session effects generated the variation of the three elements. The second experiment was to demonstrate the system performance by calculating the False Rejection Rate (FRR) and the False Acceptance Rate (FAR) of the system. In the second experiment, a total of 20 keystrokes were recorded from each subject one week after the memorizing session, in order to evaluate the FRR of the system.

In [20] an authentication method was proposed using behavioral biometrics of human or so-called behaviometrics to be the best choice of such system. The system used dynamic authentication system which is an effective solution to gain a high security information system. Verification is done by comparing the input feature vectors with template references. Feature matching employed is the combination of statistical method, measure of disorder and direction similarity measure. From the experiment conducted, the accuracy rate in distinguishing genuine and impostor users is 91.72%.

III. PROPOSED SYSTEM

In this method, and in order to achieve a strong authentication scheme, each user for the first time must be subjected to exhaustive training before the actual registration. The system consists of two main phases, the training enrollment phase and the validation phase. In enrolment phase, the user passes different tests with a selected number of password repetitions. The password must be very strong. If the user types an erroneous password, two other trials are permitted to that user otherwise he/she is rejected. The authentication scheme here depends on biometric characteristics of the user; in this paper we use the typing speed of the password by the keyboard.

In the enrollment phase , the user must enter his/her password n times and in each time the speed of typing between successive characters are registered .This is considered the first trial to train the user to be accustomed more with the password. Information of typing speed between successive characters of the password is maintained in a 2-D matrix which contains n-1 elements in each row and this row represents ith trial of n-1 successive timing speed. Each column of the matrix represents the typing speed of successive character. For example, if the user is subjected to enter password of m characters 5 times, the rows represent the 1^{st} typing speed ,2^{nd} typing speed and the final row is the 5^{th} typing speed. The number of columns is m-1 and each column represents one of each typing speed but between each successive characters.

The user then is subjected to another training session with more number of typing the password. The user continues training depending on the system strategy.

In each training session, the system measures the thresholds for different calculations. One of these calculations is to get the average and standard deviation for each row of each training session which represents the average and standard deviation typing speed of the whole password and also the average and standard deviation of each column to check these values of typing speed between successive characters. The important calculation is to get the range of typing speed in each column.

In the validation phase, the user must enter the password once after each training session but not directly after the last typing speed. The user must wait for a suitable time and then enter his/her password. The system must check all calculations in the registration phase of that session and compare it to see if the final typing speed is within the range obtained in the registration phase.

Let T= $\{T_1,T_2, \ldots..,T_n\}$ be the number of training sessions.
Let N be a set of ordered pairs which represents the number of password typing in each training session so $N=\{T_1,N_1),(T_2,N_2),\ldots.(T_n,N_m)\}$

Let thLow , thHigh be the minimum speed and maximum speed for each column of each training session .

Then we notice that for each training session the range R which is thHigh <=R>=thLow can be reduced to be more and more nearest to the actual typing speed of the authorized user ,after completing all training session , when the actual enrolment is done for future authentication of that user.

Figure 1 illustrates the structure of enrollment matrix

(T1,mn)	m_1 1st TS	2nd TS	.....	n-1 TS
	.	.	.	.
	.	.	.	.
	m_m .	.	.	.

Figure 1 . Enrollment Matrix Structure

If we suppose R_1, R_2, …….R_n be the all ranges obtained from all n sessions then the calculations of these ranges can be accumulated and analyzed to get the best thresholds to estimate the nearest typing speed.

One of the most statistical measurements is to determine the mode either for fixed speeds or for a range of speeds in each column.

The calculations used for this matrix are :

1: Dwell Time (DT) which refers to the amount of time between pressing and releasing a single key. In other words, how long a key was held pressing down. It is also worth noticing that several terms for DT appeared in the literature such as duration time [21, 22] and hold time [23]. DT can be calculated by:

$$D T n = R n - P_n \quad ,$$

where R and P indicate the time stamp of release and press of a character, respectively, while n indicates the position of the intended DT.

2: Flight Time (FT) which refers to the amount of time between pressing and releasing two successive keys. It may also be termed as latency time [24, 25], interkey time [26] or interval time [27, 28]. It always involves key event (press or release) from two keys, which could be similar or different characters. The formula to calculate each form is listed as follows:

$$F T t y p e 1 , n = P n + 1 - R n ,$$

$$F T t y p e 2 , n = R n + 1 - R n ,$$

$$F\,T\,type\,3\,,\,n = P\,n+1 - P\,n\,,$$

$$F\,T\,type\,4\,,\,n = R\,n+1 - P\,n\,,$$

where R and P indicate the time stamp of release and press of a character, respectively, while n indicates the position of the intended F T.

Figure 2 illustrates these calculations [29].

Figure 2: Dwell Time and Flight Time

All the typing speeds must be checked against different imposters to prove the strength of this system. If the typing speed of any imposter is within the authenticated speed ranges, the user must subjected to private information for each authenticated user to solve this problem. If, in addition to typing speed matching, the user provides the correct private information, the system accepts that user as an authenticated user; otherwise the system rejects that user.

IV. RESULTS

In this paper we perform a number of pre-tests beginning with 10 trials for the password and ending with 100 trials for the same password. Table 1 summarizes the results obtained from these trials.

TABLE 1 Results of all Pre-test

Number of trials	Average of typing speed	STD	Percentage of Mode of Ranges for each successive letters in each trial						
10	84.9	4.3	50	40	70	30	60	50	41
20	83.2	4.1	55	43	72	35	63	57	44
30	82.3	3.9	57	45	75	39	67	60	47
40	80.1	3.4	61	49	79	43	69	65	52
50	79.8	3.1	66	52	81	46	73	69	55
60	78.1	2.8	69	55	84	49	76	71	58
70	77.8	2.4	73	58	86	51	78	74	63
80	77.1	2.1	75	61	88	55	82	77	66
90	68.3	1.9	79	65	91	62	84	82	68
100	61.1	1.2	83	69	94	66	87	85	72

From this table we see that in the first test in which the user types the password 10 times the average of typing speed is 84.9 and this speed is reduced when the numbers of typing of the password increases until we get the best typing speed when the user types the password 100 times. We also see standard deviations between each successive characters of the password is significantly reduced as the number of typing speed is increased which means that the user , as typing trials increase , will be accustom with the password speed typing to get the best training for the best typing speed. The average typing speed was 84.9 for first 10 trials and it becomes 61.1 in the last 100 trials. Figure 3 illustrates a graph for describing the significant reduction of typing speed.

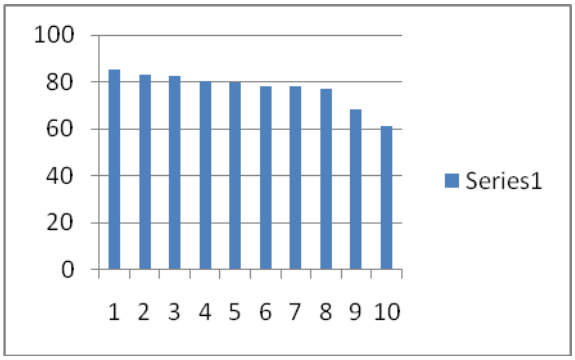

Figure 3 . Reduction of typing speeds

From table 1 we can get the best low threshold and high threshold which are computed as following:
thLow= High average speed of first 50 trials=84.9-79.8=4.8
thHigh = High average speed of last 50 trials= 78.1-61.1=17.0
So the best typing speed must be within two these ranges i.e., 4.8<=R<=17.0 relative to each average typing speed.

In each trial session the system is subjected to a different imposter as illustrated in table 2 (this table represents the trial of one imposter for each legal trial):

TABLE 2 Imposters Trials

Number of trials	Average Speed of the authenticated user	Imposter trials	States
10	84.9	101.9	Within the range
20	83.2	100.2	Within the range
30	82.3	87.1	Within the range
40	80.1	79.1	Within the range
50	79.8	97.9	Out of The Range
60	78.1	97.4	Out of The Range
70	77.8	98.1	Out of The Range
80	77.1	102.9	Out of The Range
90	68.3	98.6	Out of The Range
100	61.1	105.6	Out of The Range

This procedure is implanted for about 50 imposters and 50 legal users for each trial and we get the an FAR with 0.23% and FRR with 0.09% and these values can be improved as we increase the number of trials more than 100 trials of typing the password.

To enhance the identity of the authenticated users, each user must maintain private information. This procedure will solve the problem of matching the range of the average speed of the authenticated users with the range of any imposter. One such private information, which is in the form of question, is" Who is a memorable person from your childhood?" This procedure

specially is used when the average typing speed is within the range of high threshold and low threshold.

V . CONCLUSION AND ANALYSIS

The training procedure used in this paper leads to encouraging results. As the user is subjected to more number we get the best typing speed of that user. The average typing speed is significantly reduced from the first trials with a little number of typing the password to the trials which gave more typing of the same password. The system proposed gets the best average typing speed by comparing the average speed of any user after these numbers of trials with the high and low thresholds of the average typing speed . The result of testing average speed also reduces the number of users which are out the average speed range as the number of trials increase which means that the system gets the best average typing speed. The system also results in very good results of decreasing the number of imposters who are accepted as valid users and the number of legitimate users who are rejected as invalid users through the calculations of FAR and FRR. Finally and in order to get a strong authentication scheme , each user has private information in the case of matching the average speed of the invalid users with the average typing speed of the valid user .

REFERENCES

[1] A. Conklin, G. Dietrich, and D. Walz, "Password-based authentication: A system perspective", in: Proceedings of the 37th Hawaii International Conference on System Sciences, Hawaii, p. 10, 2004.

[2] A. Peacock, X. Ke, and M. Wilkerson, "Typing patterns: A key to user identification", IEEE Security & Privacy , PP. 40–47 , 2004.

[3] D. Gunetti, and C. Picardi, "Keystroke analysis of free text", ACM Transactions on Information and System Security (TISSEC) 8 (3) ,PP. 312–347 , 2005.

[4] S. Hocquet, J.-Y. Ramel, and H. Cardot, "User classification for keystroke dynamics authentication", in: The Sixth International Conference on Biometrics (ICB2007), PP. 531–539, 2007.

[5] M. Obaidat, and B. Sadoun, "Verification of computer users using keystroke dynamics, Systems, Man and Cybernetics", Part B, IEEE Transactions PP. 261–269, 1997.

[6] M. Theofanos, B. Stanton, and C. A. Wolfson, "Usability & Biometrics: Ensuring Successful Biometric Systems", National Institute of Standards and Technology (NIST), 2008.

[7] ISO, "Biometric performance testing and reporting", Tech. rep., ISO/IEC 19795-1:2006(E) ,2006.

[8] A. Mansfield, and J. Wayman, "Best practices in testing and reporting performance of biometric devices", NPL Report CMSC 14 (02) ,2002.

[9] J. Bhatnagar, and A. Kumar, "On estimating performance indices for biometric identification", Pattern Recognition , PP.1803 – 1815,2009.

[10] M. El-Abed, R. Giot, B. Hemery, and C. Rosenberger, "A study of users' acceptance and satisfaction of biometric systems", in: 44th IEEE International Carnahan Conference on Security Technology (ICCST'10), San Jose, California, USA, PP. 1–10, 2010,.

[11] ISO, Information technology - security techniques - security evaluation of biometrics, Tech. rep., ISO/IEC 19792-1:2008(E), 2008.

[12] F. Cherifi, B. Hemery, R. Giot, M. Pasquet, and C. Rosenberger, "Behavioral Biometrics for Human Identification: Intelligent Applications", IGI Global, 2009, Ch. Performance Evaluation Of Behavioral Biometric Systems, pp. 57–74, 2009.

[13] G. Romain , E. Mohamad , H. Baptiste , and R. Christophe," Unconstrained Keystroke Dynamics Authentication with Shared Secret" , Computers & Security, PP. 427-445, 2011.

[14] M. Sasse, S. Brostoff,and D. Weirich, "Transforming the 'weakest link' – human/computer interaction approach to usable and effective security", BT Technology Journal 19 , PP. 122–131. ,2001.

[15] I. Winkler, and B. Dealy, "Information Security Technology?... Don't Rely on It A Case Study in Social Engineering", in: Proceedings of the Fifth USENIX UNIX Security Symposium, p. 6, 1995.

[16] D. Poonam, and K.P. Chaudhri ," Typing Pattern Recognition Using Keystroke Dynamics", Communication in Computer and Information Science ,Volume 296, pp 275-280 , 2013.

[17] Z. Yu , Y.Deng, and A.Jain , "Keystroke dynamics for user authentication", Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference , PP. 117 – 123 , June 2012.

[18] C. F. Livia , H. R. Luiz , L. Lee , G. Miguel , B. T.Joao , User Authentication Through Typing Biometrics Features , IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 53, NO. 2, FEBRUARY 2005.

[19] K. Kotani , and K.Horii , "Evaluation on a keystroke authentication system by keying force incorporated with temporal characteristics of keystroke dynamics ", Behaviour & Information Technology , volume 24 , issue 4 , PP. 289-302, 2005

[20] L.Dewi, and S .Dwina , "Adaptive Behaviometrics using Dynamic Keystroke for Authentication System" , International Conference on Future Information Technology IPCSIT vol.13 (2011) © (2011) IACSIT Press, Singapore, 2011.

[21] F. Bergadano, D. Gunetti, and C. Picardi, "User authentication through keystroke dynamics," ACM Transactions on Information and System Security, vol. 5, no. 4, pp. 367–397, 2002.

[22] H. Saevanee and P. Bhatarakosol, "User authentication using combination of behavioral biometrics over the touchpad acting like touch screen of mobile device," in Proceedings of the International Conference on Computer and Electrical Engineering (ICCEE '08), pp. 82–86, December 2008.

[23] N. Pavaday and K. M. S. Soyjaudah, "Enhancing performance of Bayes classifier for the hardened password mechanism," in Proceedings of the IEEE Africon 2007 Conference, pp. 1–7, September 2007.

[24] A. M. Ahmad and N. N. Abdullah, "User authentication via neural network," in Proceedings of the 9th International Conference on Artificial Intelligence: Methodology, Systems, and Applications, pp. 310–320, London, UK, 2000.

[25] H. Davoudi and E. Kabir, "A new distance measure for free text keystroke authentication," in Proceedings of the 14th International CSI Computer Conference (CSICC '09), pp. 570–575, October 2009.

[26] C. Zhang and Y. Sun, "AR model for keystroker verification," in Proceedings of the 2000 IEEE International Conference on Systems, Man and Cybernetics, vol. 4, pp. 2887–2890, October 2000.

[27] T. Shimshon, R. Moskovitch, L. Rokach, and Y. Elovici, "Continuous verification using keystroke dynamics," in Proceedings of the International Conference on Computational Intelligence and Security (CIS '10), pp. 411–415, December 2010.

[28] S.-S. Hwang, H.-J. Lee, and S. Cho, "Improving authentication accuracy using artificial rhythms and cues for keystroke dynamics-based authentication," Expert Systems with Applications, vol. 36, no. 7, pp. 10649–10656, 2009.

[29] O. Jon, " Authentication Solutions Through Keystroke Dynamics", Enterprise Strategy Group March , 2006.