Advanced Authentication Scheme Using a Predefined Keystroke Structure

Abdulameer K. Hussain and Mohammad M. Alnabhan

Computer Science Department, Jerash University, Jerash, 26150, Jordan

ABSTRACT

This paper presents an advanced keystroke authentication model improving users' validation strength. The proposed system is based on defining a keystroke structure for each authorized user, to be used in the user login attempts. This structure is composed based on two components; the user's typing time deviation thresholds; and a unique user secret code which is distributed between password's characters based on time distances. The strength of the proposed method depends primarily on the amount of information distributed among typing time, and on reducing the deviation of these times. During the preliminary evaluation, it was confirmed that the proposed system has achieved an improved authentication level, and the system model was highly accepted between participating users.

KEYWORDS

Authentication, Keystroke, Dynamics, Predefined structure, Time Distance.

1. INTRODUCTION

Authentication is the process of determining whether a user is allowed to accesses a particular system or resource. The major objective of authentication system is allowing entities to be recognized before using resources. Several authentication methods are available starting from alphanumeric passwords until the use of biometrics and smart cards. However, the use of these technologies has raised several concerns such as the acceptability and lack of flexibility, and lack of robustness against imposters. In addition, traditional method such as the couple of username and passwords are required to be effective for authentication, easy and quickly executable, which can be considered as conflicting and difficult for humans. However, to guarantee strong authentication it is required to integrate multiple authentication methods. For example, it is possible to provide strong authentication in the password authentication scheme by combining it with keystroke dynamics [1].

Keystroke is a behavioural biometric modality monitoring the way individuals' type on the keyboard [2]. The basic idea of keystroke dynamics is based on the assumption that people type in uniquely different characteristic manners. Hence, this method depends on identifying users certain habitual typing rhythm patterns [3]. Different names of keystroke dynamics are used: keyboard dynamics, keystroke analysis, typing biometrics and typing rhythms [8]. One of the advantages of keystroke dynamics is that it is inexpensive because it can be used without any additional hardware. In addition, the user acceptance of a keystroke dynamics biometric system is considered very high [16, 10].

DOI:10.5121/ijcsit.2014.6212

Several research works had addressed the usage of keystroke dynamics in improving authentication systems performance. However, still keystroke drawbacks such as users typing time deviation needs to be considered. This work considers developing keystroke dynamics to reach an effective and reliable authentication solution; this was achieved by defining a new keystroke structure for each authorized user and by solving the deviations in user password typing time.

2. RELATED WORK

In [12] keystroke dynamics was applied to measure users typing process using statistical methods, in which users were divided into different groups in order to speed up the required dynamic computation. Additionally, [4] describes preliminary experimental results describing using keystroke timing as a basic of authentication system in which a textual material and a statistical model was developed and used within an experimental study. Most keystroke dynamics studies had been evaluated using datasets where users typed the same fixed string [7], [6], while very few of them used different strings for each user [11].

Authors in [9] presented a filtering scheme and adaptation mechanism to improve the stability and effectiveness of keystroke authentication. In which, the typing characteristics of users are measured by n-dimensional vectors and an ellipsoidal hypothesis space, which is evolved using a genetic algorithm. In [13] a novel keystroke dynamics authentication system was presented. This model utilizes two sets of derived values to constructs a template for identifying the user based on his typing style. The first set of derived values is computed based on the collected measurements, and the second set is computed based on the first set of derived values. [14] shows the possibility of using neural networks especially in static keystroke dynamics verification, in which researchers created a template for each user by using approximately 30 user samples and 45 impostors samples; where the samples represents the timing information that are extracted from the typing of the name of the user. In the same concern, the behaviour of user typing was used along with password based security to achieve enhanced security. This was achieved by analyzing the basic user behaviours/activities and finally training users by neural network and classifying them as legal or intruder [15].

In addition, individual's typing behaviour was considered in [17], in which a new function was presented to train users through keystroke dynamics and a set of validation rules were applied to validate system users. Furthermore, keystroke dynamics were utilized in [5] to be associated with PIN codes used in ATM machines, in which a compromising algorithm was implemented, and used to withdraw the security threat, that might happen when the imposter get hold of both user-ID (user card) and password.

Accordingly, several research works has addressed the usage of keystroke dynamics in improving authentication systems performance. However, still keystroke drawbacks such as users typing time deviation needs to be considered. This work considers developing keystroke dynamics to reach an effective and reliable authentication solution; this was achieved by defining a new keystroke structure for each authorized user and by solving the deviations in user password' typing time

3. PROPOSED SYSTEM MODEL

This system depends upon constructing a predefined keystroke structure for each user to ensure improved authentication strength. The system considered a strong users password especially for sensitive applications. Figure 1 below describes the proposed system steps. The first phase is

described as the enrolment phase, in which users are trained several times to enter password characters before the actual registration phase, in order to measure the typing timing periods and the deviations thresholds accurately. In addition, the time deviation of typing speed over trails is also considered enrolment phase.

Table 1 illustrates the procedure of measuring the timing periods between each successive characters of the password. This table represents a matrix maintaining the time periods between each successive characters of the password. Suppose, the password consists of n characters, then T11 to T1m represents the typing time between the first character and the second characters for m trials, T21 to T2m represents the timing periods between the second characters and the third characters, and T1n to Tnm represents the typing periods between the character before the last character of the password and the last character. In addition, two timing thresholds for each column in table 1 must be identified for each successive character. The upper range threshold denoted as (th1) and the lower range threshold denoted as (th2), in which the user's typing time must lie between these thresholds.

Table 1: Registration Matrix

Trial No	**Typing Time between character 1 and character 2**	**Typing Time between character 2 and character 3**	**…**	**Typing Time between character n-1 and character n**
1	T11	T21	…	Tn1
2	T12	T22		Tn2
m	T1m	T2m		Tnm

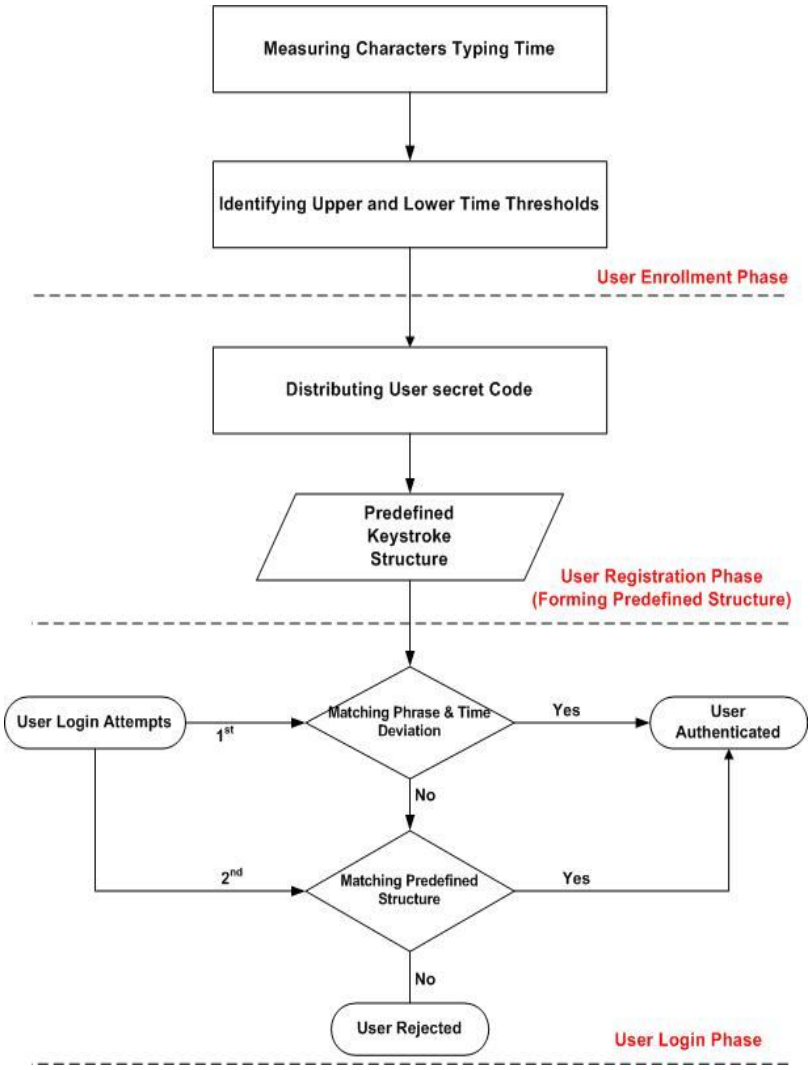

Figure1: Proposed System Model

The second phase is responsible for forming the predefined keystroke structure, which consists of two parts; the password characters and users unique secret code distributed between these characters based on the typing time deviation thresholds measured in phase 1. Suppose, the secret information is S, this can be divided into different parts (S0, S1, … Sn-2) for password of length n characters, figure 2 below illustrates the predefined structure including both parts:

>=th11<=th12|| S0	>=th21<=th22 || S1	…	>=thn-1<=thn	Sn-2

Figure 2: The predefined structure of keystroke Dynamics and Secret Information

Where th11 and th12 represents the upper and lower range of thresholds for the first and second characters of the password, th21 and th22 represents the upper and lower range of thresholds for the second and the third password characters. thn-11 and thn represents the upper and lower range

of thresholds for the last two password characters before. The last step in the system model, is described as the login phase, where users entre their passwords. The system calculates the typing time as in the enrolment phase, and then checks the upper and lower thresholds for each successive character. If the new entry lies within these thresholds, the user is considered authenticated and will be successfully logged into the system. If the same authenticated user makes some distances from the thresholds, then the system rejects that user. In this case, the system asks the user to retry logging to the system by typing his keystroke structure, which represents the password characters accompanied with segments of the unique code S. If the entered structure matches the specific predefined structure being formed for this user, then the user is considered entirely authenticated.

4. RESULTS AND ANALYSIS

In order to evaluate the proposed system, 10 measurement attempts were conducted allowing each user to enter password characters, in order to measure typing time deviations (in millisecond) between two successive characters. Table 2 below summarizes the typing time deviations for one single user considering 10 trials. The user utilises a strong password consisting of set of special characters.

Table2: Time distances between successive characters of the use's password (during 10 trials)

1st char & 2nd char.	2nd & 3nd char.	3rd char. & 4th char.	4th char. & 5th char.	5th char. & 6th char.	6th char. & 7th char.	7th & 8th char.	8th char. & 9th char.	n cha. & 2nd char.	Average time of each row
H 344	L 218	H 359	187	343	172	L 203	280	156	251
343	H 312	249	H 250	343	203	218	297	141	H 261
280	265	171	188	358	H 219	H 234	H 312	110	237
234	250	187	203	358	188	218	312	H 156	234
250	249	L 156	250	327	156	219	296	141	227
234	234	172	L 187	H 359	218	219	296	L 109	225
281	234	187	203	297	171	219	280	140	223
234	250	156	187	343	187	203	296	109	L 218
234	250	249	219	L 296	187	203	281	125	227
L 234	265	203	265	312	L 141	218	L 234	109	220

Note: H stands for high threshold and L stands for low threshold for each column.

For each user, the time between successive characters are calculated and stored in a specific profile. The average time distance in each trial is registered measuring the upper range (denoted as H in Table 2) and the lower range (denoted as L in Table 2); which corresponds to the thresholds th11 and th22 mentioned in figure 2.

After measuring the time deviation thresholds; the predefined keystroke structure for the experimental user was formed. The length of the user's secret code is selected depending on the user's password and on the length of each splitting segments of this private information. During the experimental scenario, the length of the user's password was 9 characters, so the suitable length of the user secrete code or private information must be at least twice of the password's length (i.e., 18 characters). For example, the phrase "secureapplications" is selected as the user's private information, then the predefined structure after the distribution of this phrase considering typing time distances as described in table 2, will appear as the following:

```
>=234 <=344se>= 109<=156cu>=218 <=312re>= 156<=359ap>=
187<=250pl>=296 <=359ic >=141 <=219at >=203 <=234io >=218 <=312ns
```

Figure 3: First sample of the predefined keystroke structure

Using the average of typing time deviations thresholds as described in the last column of table 2; the shape of the predefined keystroke structure will appear as the following:

```
>= 218<=261se>= 218<=261cu>= 218<=261re>= 218<=261ap>=
218<=261pl>= 218<=261ic >= 218<=261at >= 218<=261io >=
218<=261ns
```

Figure 4: Second sample of the predefined structure

The same measurement steps were repeated for 10 participating users, in which the model acceptance among participants and the authentication successful rate was almost 80% during the experimental trials. In addition, the presented authentication model has solved the problem of large deviations in keystroke dynamics. As show in figures 2 and 3, the predefined structures have shown a strong authentication solution, in which user is considered authenticated after providing the correct password characters within the right ranges of typing time deviations, or after entering the password characters accompanied with segments of the unique code representing his keystroke structure, which should match the predefined structure being formed for the user during the registration phase as described in figure 1.

5. CONCLUSIONS

The proposed authentication model solves the problem of large deviations in keystroke dynamics and provides improved keystroke authentication level. This was achieved by defining new keystroke structure for each system user. The structure consists of two parts; the password characters and segments of user secrete code distributed among password characters based on users typing time thresholds. The proposed authentication model consists of several phases; starting with the enrolment phase which defines the time distance thresholds. Following, the predefined keystroke structure is formed using the time thresholds and user's secret code. Afterwards using the keystroke structure, users can be authenticated and logged in the system. A set of preliminary measurement trials were conducted evaluating the proposed model phases and determining the system performance and successful rates. It was confirmed that the system has achieved a strong authentication level and the system model was highly accepted between participating users. However, in future measurements it is worth increasing the number of participating users and using different password and secret codes samples. This will provide an increased validity to system evaluation process.

REFERENCES

[1] Kang, P., Hwang, S.-s. Cho, S., "Continual retraining of keystroke dynamics based authenticator", in: S.-W. Lee, S. Li (Eds.), Proceedings of ICB 2007, of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Vol. 4642, pp. 1203–1211, 2007. 04/010970000/seminars/Ilonen.pdf (accessed January 2013)

[2] Revett, K., "A bioinformatics based approach to user authentication via keystroke dynamics", International Journal of Control, Automation and Systems, vol.7, no.1, pp.7–15, 2009.

[3] Monrose, F., Rubin, A., "Authentication via Keystroke Dynamics", ACM Conference on Computer and Communications Security, pp.48-56, 1997.
[4] Gaines, R., Lisowski, W., Press, S., Shapiro, N., "Authentication by keystroke timing some preliminary results", Rand Report R-2526-NSF, Rand Corporation, 1980.
[5] Giot, R., El-Abed, M., and Rosenberger. C., "Greyc keystroke: a benchmark for keystroke dynamics biometric systems". Proceeding of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009), pp.1–6, 2009.
[6] Gunetti, D., Picardi, C., Keystroke analysis of free text, ACM Transactions on Information and System Security (TISSEC) 8 (3) (2005) 312–347.
[7] Hocquet, S., Ramel, J.-Y., Cardot, H., "User classification for keystroke dynamics authentication", in: The Sixth International Conference on Biometrics (ICB2007), pp. 531–539, 2007.
[8] Ilonen, J., "Keystroke dynamics", Lappeenranta University of Technology, Finland, 2003, [Online: http://www.it.lut .fi/kurssit/03-
[9] Jae, L. Sung-Soon, C., and Byung, M., "An evolutionary keystroke authentication based on ellipsoidal hypothesis space", Proceedings of the 9th annual conference on Genetic and evolutionary computation, pp.2090-2097, 2007
[10] Kacholia, V., Pandit, S., "Biometric Authentication using Random Distributions (BioART)", 2003, [online: http://shashankpandit.com/papers/bioart/paper.pdf, (accessed January 2013)
[11] Balagani S., Phoha V., Ray A., and Phoha. S., "On the discriminability of keystroke feature vectors used in fixed text keystroke authentication", Pattern Recognition Letters, vol.32.no.7, pp.1070 – 1080, 2011.
[12] Manpreet, K., and Rajinder, V., "Security System Based on User Authentication Using Keystroke Dynamics", International Journal of Advanced Research in Computer and Communication Engineering. vol.2, no.5, 2013.
[13] Mechthild, R .,.Kellas, D., and Yvonne, J., "Keystroke dynamics authentication techniques" , patent , Publication number US8332932 B2 , Dec 11, 2012
[14] Cho, S., Han, H., Han, C., and Kim. H.-I., "Web-based keystroke dynamics identity verification using neural network", Journal of organizational computing and electronic commerce, vol.10, no.4, pp.295–307, 2000.
[15] Preet, S., "Enhanced Password Based Security System Based on User Behavior using Neural Networks", International Journal Information Engineering and Electronic Business, vol.2, pp.29-35, 2012.
[16] Bleha, S. Slivinsky, C. Hussien, B. "Computer-access security systems using keystroke dynamics", IEEE Transactions On Pattern Analysis And Machine Intelligence vol.12 pp.1216–1222.
[17] Sally, A., and Izzeldin, O., "An Application of the Keystroke Dynamics Biometric for Securing PINs and Passwords", World of Computer Science and Information Technology Journal (WCSIT), vol.1, no.9, pp.398-404, 2011.

Authors

Mohammad Alnabhan finished his bachelor degree in computer science, from Mu'tah University, in 2004. He received his master degree in computer science from Anglia Ruskin University (ARU), in 2006. Alnabhan earned his PhD degree from Brunel University in 2009; his research field was on mobile computing. Where, he developed an innovative Location Based Services (LBS) model focused towards disabled pedestrians. After completing his PhD, Mohammad Alnabhan was appointed as an Assistant Professor in the computer science department at Jerash University, where, he was involved in teaching a great variety of computer science courses in both undergraduate and postgraduate levels. During his academic career, alnabhan has confirmed outstanding research ability, where he has published more than twenty research articles in highly reputed Journals and international conferences. In which, his research interest includes mobile computing, context adaptive computing, QoS measurements, m-learning, and Location Based Services (LBS).