Password Recovery Mechanism Based on Keystroke Dynamics

Soumen Roy, Utpal Roy and D.D. Sinha

Abstract Automated password recovery process includes needing the user to response a "secret question" defined as part of the user registration process. The second mechanism in use is having the user offer a "hints" during record-keeping that helps the user remember his password. Here, system may be compromise through the use of brute-force attacks, inherent system weakness or easily guessed secret questions and answers. The third mechanism is One Time Password (OTP), where personal phone or alternative user ID is needed. The fourth mechanism is our proposed, password recovery mechanism based on behavioral biometric characteristics, keystroke dynamics data. Here, users are not only identified by their secret questions answers or hints, but their typing style is also accounted for. It improves the security rank and can be used to identify the real user. In this paper, we have also suggested some future plans that also can be effectively implemented by this technique.

Keywords Keystroke dynamics · Password recovery mechanism · Behavioral biometrics · Computer security · Manhattan distance · Euclidean distance · Mahanobolis distance · Z score

S. Roy (✉) · D.D. Sinha
Department of Computer Science and Engineering, University of Calcutta, 92 APC Road, Calcutta 700009, India
e-mail: soumen.roy_2007@yahoo.co.in

D.D. Sinha
e-mail: devadatta.sinha@gmail.com

U. Roy
Department of Computer and System Sciences, Visva-Bharati University, Santiniketan 731235, India
e-mail: roy.utpal@gmail.com

J.K. Mandal et al. (eds.), *Information Systems Design and Intelligent Applications*,
Advances in Intelligent Systems and Computing 339,
DOI 10.1007/978-81-322-2250-7_24

1 Introduction

Social networking sites, online form fill-up, E-mail ID providers are considered to have user-login mechanism (Knowledge based user authentication technique), where password recovery mechanism is one part. Among all password recovery mechanisms, "secret questions answers" and "password hints" are very popular. But if an attacker collects our personal information and check answer one after another against single question, then attacker may change the password or can get the access of the system. In order to defeat their technique in practice a higher level of security and performance together with low cost version is demanded to an accepted level of error, to be designed.

Now-a-days, Knowledge-based user authentication technique is not limited to password or PIN. Our behavioral biometric properties such as keystroke dynamics: a technology to segregate and distinguish people based on their typing rhythms, is also accounted for. Our proposed system uses personal information of clue for a password or secret question answer at the time of registration, compressing of characters as well takes into account the typing style of depressed characters entered. Here, typing style is a key issue to identify the authentic user, which is stored in the database. It will help recover the password, as strong as biometric properties which are meshed up with personal information that we are habituated to press it daily. It ensures the consistence typing style of a user.

We have collected typing style of three common passwords six times each from 15 different individuals having different age and education level using JAVA language at a fixed keyboard which laboratory made sample password database has been used to train the system. Here system records all the key press and release timing and calculates the duration of depressed characters, latency time between various down and up key sequence latencies for each sample, then found out the actual timing template by applying some statistical methods. Then some features mining mechanism like Euclidean distance, Manhattan distance are used to calculate the score. Minimum score is to be finding out and system decides the corresponding user is valid or not. Thus we can reduce the probability of brute-force or shoulder surfing attacks. The rhythm of the password as it is entered is used to improve the security level and validates the authenticity of the user rather than only Secret question answer or hints. Here, system calculates duration of depressed characters and pause duration between each subsequent characters entered. This timing parameters promise parameters like biometric characteristics that may facilitate non-intrusive, cost-effective and continuous monitoring.

Keystroke Dynamics is not new technique. Bryan and Harter first formally investigated it in 1897 as part of a study on skill gaining in telegraph operators. Spillane suggested in an IBM technical bulletin in 1975 that typing rhythms might be used for identifying the user at a keyboard. Forsen et al. [1] in 1977 conducted first round tests of whether keystroke dynamics could be used to differentiate typists. Gaines et al. [2] produced an extensive report of their investigation with seven typists into keystroke dynamics in 1980. After then, S. Bleha submitted his

Ph.D. thesis on Recognition system based on keystroke dynamics in 1988 [3]. Joyce and Gupta [4] proposed an identity authentication based on keystroke latencies in 1990. Monrose and Rubin [5] proposed keystroke dynamics as a biometric for authentication in 2000. Keystroke dynamics research has been going on for the more than thirty three years. Many methods have been proposed during that time. But all the keystroke dynamics features are not considered in any proposed system. Here, we have considered five effective factors.

Automated password recovery process includes requiring the user to answer a "secret question" defined as part of the user registration process. The second mechanism in use is having the user offer a "password hints" during record-keeping that help the user remember his password. Here, system may be compromised in an attack through the use of brute force, inherent system weakness or easily guessed secret questions and answers. The third mechanism is One Time Password (OTP), where cell phone or alternative user ID is needed.

Keystroke dynamics technique is used in our proposed system where typing style (key pressing and releasing time) will be calculated which is unique [6]. Our system takes all the key pressing and releasing timing and then calculates five timing factors; key hold time and four key latencies of some sequence of key press and release timing and stores it into the database for future purpose. In our system, we have to press any sequence of characters (good to choose some common words like name, address, E-ID, …) as claimed string, then the system compares with the stored data and decides whether user is authenticated or not and only then the system offers new password or PIN.

2 Password Recovery Mechanism

Password recovery mechanism is essential technique in knowledge base user authentication technique, which mechanism provides the facilities to choose a new password once again after verification of the valid user. It is one factor technique where one factor such as "password hints" or "secrete answer" or user's phone number or email address. But our system takes two factors one is hints or secret answer and keystroke dynamics data.

People are still unimaginable and lazy to choose strong password. Generally, we, as people pick up some words for password from relatively small dictionary which makes easy to break the password for attackers. If we choose stronger password for different access control, it is hard to remember different strong passwords for different systems. Keystroke Dynamics is a technique, which can solve this problem, where we have nothing to remember. Here typing style identifies the user [7–10].

Conventional website authentication method uses the following password recovery method.

2.1 Secret Question Answer Method

A System, which asks some selective secret questions to the user as a reference data obtained at the time of registration for comparing purpose in future.

2.2 Password Hints Method

It is a clue for password, artificially it is known to the user.

2.3 One Time Password (OTP)

Automatically one random password will be generated and it is sent to the user's alternative E-mail ID or Cell phone by the system. This password is valid for one time.

3 Keystroke Dynamics

Keystroke dynamics is a behavioral biometrics which is the technique of examining the way a user types on a keyboard and identify him/her based on his/her regular typing pattern [11]. It is the study of whether people can be well-known by their typing pattern, much like handwriting is used to identify the author of a written text.

Our typing style can be easily calculated by simple program which can calculate key pressing and releasing time of each key and then generates key-hold time and keys-latency times of all down up events [7] (Tables 1 and 2).

Table 1 Key press and release time of fixed-text "kolkata123"

Entered key	Key press time	Key release time
k	0	109
o	172	281
l	375	484
k	609	733
a	749	889
t	1,326	1,451
a	1,482	1,623
1	1,950	2,059
2	2,169	2,278
3	2,387	2,496

Table 2 Sample key hold and key latency times with standard deviation for the recorded fixed-text "kolkata123" for a user

Key	Key hold time	Key hold (Sd)	Down down key latency	Down down key latency (Sd)	Up up key latency	Up up key latency (Sd)	Up down key latency	Up down key latency (Sd)	Down up key latency	Down up key latency (Sd)
k	78	16.7988	187	9.8082	213	29.0034	291	16.1431	109	16.9823
o	104	12.7436	205	6.1482	213	19.0526	317	8.0125	101	13.0461
l	111	6.5727	267	36.4609	291	30.5188	403	36.1026	156	31.4038
k	135	21.1849	140	19.7737	116	27.3861	252	20.9189	5	8.0125
a	111	23.1733	231	28.8166	231	30.3546	343	26.1763	119	33.8674
t	111	15.3493	127	18.4282	119	12.6016	231	15.7353	15	13.8852
a	104	7.7460	377	18.3085	351	21.6241	455	23.0998	273	21.4756
1	78	17.5271	174	11.9248	208	16.2049	286	8.0125	96	15.4337
2	112	6.3875	205	11.8828	197	16.2111	309	12.0499	93	9.8184
3	104	12.7436	–	–	–	–	–	–	–	–

4 Proposed System

4.1 Raw Data Collection

Raw data is a set of personal information and there typing style. Here all the event time will be calculated and that will be stored in the database.

List of the following information are captured when keystroke event happen.

Event	Two events are PRESS and RELEASE of key.
Key code	It is ASCII code of the entered key.
Timestamp	System calculates the time of all the key press and release event occurs. It is usually represented in millisecond (ms).

4.2 Data Extraction

After obtaining user's raw keystroke timing information we extracted data in different ways, taking into account the occurrence in time of specific events, number of specific event occurrence in a period of time, simultaneous occurrence of events and others. For the purpose of keystroke dynamics most used measures are user ID, password and their dwell time, flight time, overlapping of specific keys, method of error correcting, cursor navigation-specific keys, key pressure, sequenced combined keys timing (di-graph, tri-graph), typing speed, finger movement style on keyboard etc. In free text authentication fixed password may not be needed but huge data sets are required to identify the valid user.

Our system calculates all key press and release time and generates 5 timing factors (Key Hold time, Down down, up up, up down and down up key latencies) by equations defined in [7].

4.3 Features Selection

Some basic features are Di-graph which represents the time information of two consecutive entered keys, Tri-graph which represents the time information of three consecutive entered keys, Dwell time which refers to the key hold time, Flight time which refers to the duration time between pressing and releasing two consecutive entered keys, Error rate which is the occurrence of errors and how to resolve is also a good factor, Shift and Control key: Event of pressing shift or control key and there arrangement of pressing the event also a feature. But in our system we have taken five timing factors and one fixed common words for identifying the valid user.

4.4 Decision Making

The capture sample is sent to the conclusion module as soon as its score is considered. This module is answerable for comparing the match score with an identity recognition threshold, determining if the challenge score is valid or not.

Claimant's keystroke data is presented to the system and equated to the reference stored template via some distance based algorithms. A final decision will be finished based upon the outcome of classification or matching algorithm to govern if a user is genuine or otherwise.

4.5 Profile Updating

Characteristics of human may change over time. So update mechanism is needed to update template after acceptable verification or identification. Sometimes, score of different algorithms differs. It would be better if we combined all scores in a single equation like mean value calculation with given weights of all scores.

4.5.1 Overall System Model

Access control system can be accessed by two ways one is normal procedure another is secret question answer or hints. Our system will takes some common strings and rhythm then compares with the stored keystroke data which is generated from normal access. Flow of the control is defined in Fig. 1.

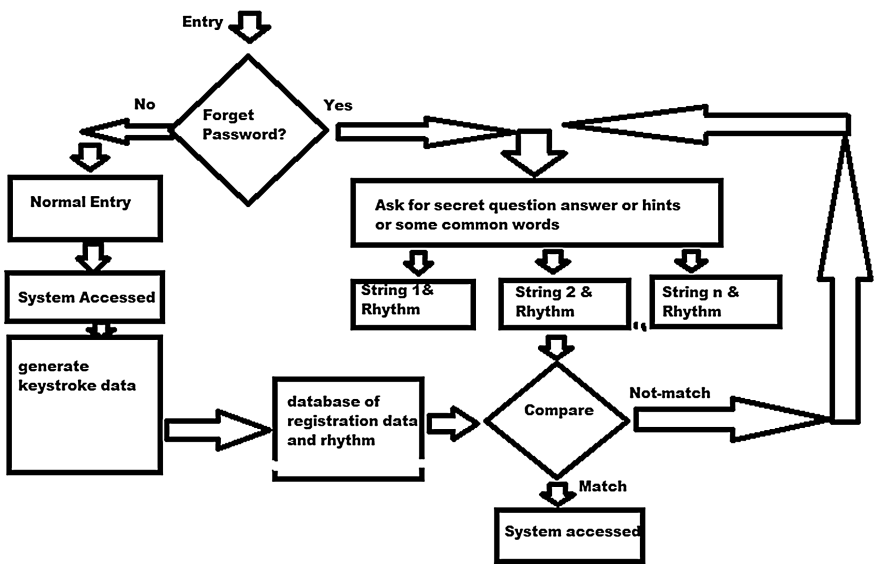

Fig. 1 Total control flow in our system

5 Keystroke Dynamics Algorithms

Many classification and distance based or score calculation methods have been applied in keystroke dynamics training over the last three decades [7, 12]. Following are the distance based algorithms were used to evaluate the system.

5.1 Manhattan Distance

Manhattan distance is one distance based method which calculates the score and minimum score will be treated as perfect match and corresponding user will be treated as a genuine user. Manhattan distance is formulated below:

$$M = \sum_{i=1}^{n} (|x_i - y_i|) \tag{1}$$

where $x = (x_1, x_2, x_3, \ldots, x_n)$ represents stored vector and $y = (y_1, y_2, y_3, \ldots, y_n)$ represents the claim vector of the exercise sample.

5.2 Manhattan with Standard Deviation Distance

The standard deviation of each feature is calculated as in Eq. 2. Here α_i represents standard deviation.

$$M_s = \sum_{i=1}^{n} (|x_i - y_i|)/\alpha_i \tag{2}$$

5.3 Euclidean Distance

The score is calculated as the squared Euclidean distance between the stored vector and claim vector as in Eq. 3.

$$E = \sqrt[2]{\sum_{i}^{n} (|x_i - y_i|)^2} \tag{3}$$

5.4 Mahanabolis Distance

The standard deviation of each feature is calculated, where Mahanabolis distance is presented in Eq. 4.

$$Eh = \sqrt[2]{\sum_{i}^{n}((|x_i - y_i|)/\alpha_i)^2} \tag{4}$$

5.5 Z Score Values

The z score is calculated in Eq. 5:

$$Z = \sum_{i=1}^{n}(|X_i| - \mu(|X_i|))/\alpha_i \tag{5}$$

where $\mu(x_i)$ are mean value and α is standard deviation.

6 Evaluation an Analysis

There are some performance measurement parameters that are used to evaluate performance of different biometric system [12].

6.1 False Acceptance Rate (FAR)

FAR defined by the following equation:

$$\text{FAR} = \frac{\text{Number of falsely accepted illegitimate users}}{\text{Total number of imposters}}\% \tag{6}$$

6.2 False Rejection Rate (FRR)

FRR defined by the following equation

$$\text{FRR} = \frac{\text{Number offalsely denied legitimate users}}{\text{Total number of genuine users}}\% \tag{7}$$

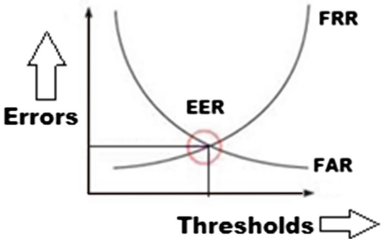

Fig. 2 Equal error rate or cross error rate

6.3 Equal Error Rate (EER)

ERR is the rate at which both ERR and FAR are equal. See the Fig. 2.

In our simulation program in JAVA, we have recoded each key press and release time for six sample of passwords (size of password <=10) and calculated key hold time and latency time between various down and up key sequence latencies, which are shown in Fig. 3. After then we have calculated score by different algorithms with calculated mean, which are very much similar. Calculated core is defined in the Table 3.

Euclidean distances of the string "123456" of 15 users are very similar. Where "kolkata123" and "password" strings are strong to identify the user which is shown in Fig. 2.

As per the experimental result, we got 0.133 % EER for the string "kolkata123" where 0.4 and 0.53 % EER for the string "password" and string "123456" separately. Size of the string and common used words are best choice to choose secrete question answer or hints where we are habituate to press this type of words daily. We have collected data from the users those are belonging to Kolkata, India, so they

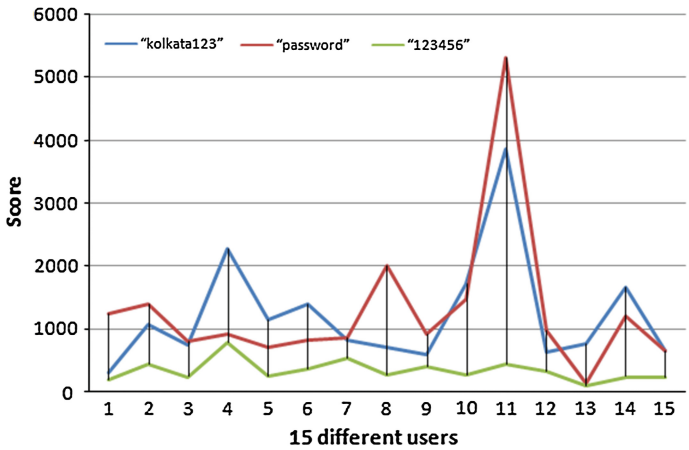

Fig. 3 Euclidean distances of different set of data for the password "kolkata123", "password" and "123456"

Table 3 Score calculation by different algorithms with the fixed-text "kolkata123" for 6 same sample of a user

Samples	Manhattan distance	Manhattan with Sd	Euclidean distance	Mahanabolis distance	Z score
1	758.0	39.5596	161.3877	7.4959	32.8560
2	567.0	28.9450	129.3947	5.7496	24.8224
3	412.0	25.3889	79.6618	4.8356	21.6220
4	525.0	31.8757	98.4937	5.9892	18.8801
5	637.0	36.5833	114.4596	6.3400	28.6692
6	685.0	35.4427	139.2875	6.4269	31.5871

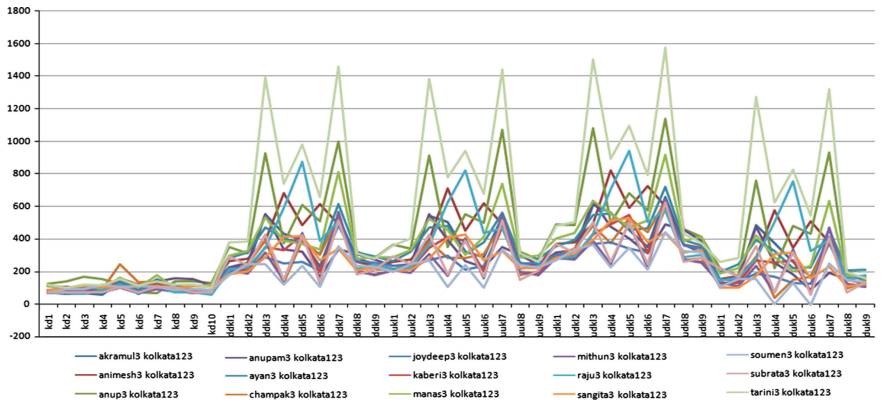

Fig. 4 15 same sample of passwords "kolkata123" for 15 different users

are habituated to press string kolkata. We got the better result for the string "kolkata123" because it is a longer size string than other tested strings.

In our experiment, we have collected rhythm of 270 predefined fixed-texts from 15 users from Kolkata and seen that the user's typing styles are dissimilar as per following line chart generated by the experimental database. Good suggestion is c*hoose the password what we press daily such as user ID*, *name*, *place* etc. Otherwise three factors will affect the system, finger movement time, key searching time and different keyboard (Fig. 4).

7 Comparison with Existing Systems

Generally we forget the secret question answer or hints for not frequently using the system. So it would be difficult to access through secret question answer or hints method. OTP would be best probable solution but in OTP, alternative e-mail is required or cell phone. It takes few times. But our system does not need any

alternative e-mail ID or cell phone and no need to remember secret question answer or hints.

Brute-force attacks or shoulder surfing attacks may happen in secret question answer or hints method. But there is no chance in OTP. In our system no one can copy our typing style even if he/she watch our typing style.

8 Discussion

Our simulation program, which we have written in JAVA, is working very well and gave promising result. Where all five timing factors are considered to key issue which is unique and cannot be copied or stolen by other users.

Typing style is a behavioural biometric characteristic which may change over time. Person's typing style may vary subsequently during a day or between two days. System needs updating mechanism where each record will be appended by the latest sample to the stored reference data. Size of the stored data may be increased.

Further, improvingly, keystroke dynamics depends upon mental state of the users. In that case keystroke dynamics may change to an extent. Such change yet to be improved so that system can cope with the situation.

In this technique, using different keyboards may affect the way. But if we consider bi-graph (two subsequent keys) or tri-graph (three subsequent keys) or syllable duration, this problem can be resolved or artificial keystroke by general setting [13] is the best solution or keystroke sound which is explained in [14].

If we solve this type of problems, this technique may give best possible solution, which is very easy to implement with the exiting knowledge based authentication technique.

To recognize keystroke dynamics, no extra security apparatus is needed. So this technique is cost effective and can be easily implemented with small alterations.

9 Conclusion

Among three password recovery mechanisms there are, secret question answer, password hints and OTP, OTP is the best but we may not have any alternative account or cell phone. In this situation keystroke dynamics is the best possible alternative solution where biometric characteristics are considered to distinguish people. It can be implemented with existing password recovery mechanism with small alterations, which enhances the security level and provides the better performance of the system. This technique can be also effectively implemented in distance-based examination, cyber criminal investigation, identifying back door account etc. But this technique, as of now, suffers from accuracy level and performance. In order to realize this technique in practice a higher level of security and

performance together with low cost version is demanded with an error to an accepted level. Hence, it is highly essential to identify the controlling parameters and optimize the accuracy, performance as well as cost with new algorithms, this is our future job.

Acknowledgments Authors acknowledge Mr. Champak Chakraborty, Department of Physics, Bagnan College and Mr. Sourjya Roy, Department of English, Bagnan College for reading the manuscript carefully.

References

1. Forsen, G., Nelson, M., Raymond Staron, J.: Personal attributes authentication techniques. Technical report RADC-TR-77-333, Rome Air Development Center, Oct 1977
2. Gaines, R., Lisowski, W., Press, S., Shapiro, N.: Authentication by keystroke timing: some preliminary results. Rand Rep. R-2560-NSF, Rand Corporation, California (1980)
3. Bleha, S., Slivinsky, C., Hussien, B.: Computer-access security systems using keystroke dynamics. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 1217–1222 (1990)
4. Joyce, R., Gupta, G.: Identity authorization based on keystroke latencies. Commun. ACM **33** (2), 168–176 (1990)
5. Monrose, F., Rubin, A.D.: Keystroke dynamics as a biometric for authentication. Future Gener. Comput. Syst. **16**(4), 351–359 (2000)
6. Killourhy, K.S.: A scientific understanding of keystroke dynamics. Ph.D. Thesis, Computer Science Department, Carnegie Mellon University, Pittsburgh (2012)
7. Roy, S., Roy, U., Sinha, D.D.: Enhanced knowledge-based user authentication technique via keystroke dynamics. Int. J. Eng. Sci. Invention (IJESI) **3**(9), 41–48 (2014)
8. Roy, S., Roy, U., Sinha, D.D.: Rhythmic password-based cryptosystem. In: 2nd International Conference on Computing and System, pp. 303–307. University of Burdwan, West Bengal, India (2013)
9. Roy, S., Roy, U., Sinha, D.D.: Modified knowledge-based user authentication technique. In: 7th International Conference on Mathematical Science for Advancement of Science and Technology, MSAST, vol. 2, p. 236. IMBIC, Kolkata, India (2013)
10. Roy, S., Roy, U., Sinha, D.D.: Combined user authentication technique. In: International Conference on Recent Trends in Science and Technology (ICRTST), pp. 106–113. College of Engineering and Management, Kolaghat, West Bengal, India (2013)
11. Pin, S.T., Andrew, B.J.T., Shigang, Y.: A survey of keystroke dynamics biometrics. Sci. World J. **2013** (2013). Article ID: 408280
12. Shimaa, I.H., Mazen. M.S., Hala, H.: User authentication with adaptive keystroke dynamics. IJCSI **10**(4) (2013)
13. Janakiraman, R., Sim, T.: Keystroke dynamics in a general setting. In: Advances in Biometrics (ICB 2007). Lecture Notes in Computer Science, Seoul, Korea, vol. 4642, pp. 584–593. Springer, Berlin
14. Roth, J., Liu, X., Ross, A., Metaxas, D.: Biometric authentication via keystroke sound. Int. Conf. Biometrics (ICB) **1**(8), 4–7 (2013)