IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

Authenticating Mobile Phone User using Keystroke Dynamics

Baljit Singh Saini[1*], Navdeep Kaur[2], Kamaljit Singh Bhatia[3]

[1]Dept. of CSE, SGGSWU, Fatehgarh Sahib, Punjab, India
CSE, Lovely Professional University, Phagwara, Punjab, India
[2]Dept. of CSE, SGGSWU, Fatehgarh Sahib, Punjab, India
[3]Dept. of ECE, IKGPTU, Batala Campus, Punjab, India

[*]*Corresponding Author: baljitsaini28@gmail.com, Tel.: 9464727063*

Available online at: www.ijcseonline.org

Accepted: 06/Dec/2018, Published: 31/Dec/2018

***Abstract*—** Since few decades, the simple password authentication has either replaced or compounded with biometrics (such as Facial Recognition, Fingerprint Scan etc.) to provide better security. Keystroke Dynamics is behavioral biometrics that can perform continuous authentication to detect intruders. In this paper, we investigate whether user specific password gives better performance than artificially rhythmed password. Also, impact of sensory data on overall performance of the system is examined. Finally, Genetic Algorithm is used to optimize the features. The features used to analyze the user data were hold time, flight time and X, Y and Z axis reading from accelerometer sensor. Results showed that user data gives better performance than artificially rhythmed passwords. Best accuracy of around 90% was achieved by using user specified passwords and optimizing the results with genetic algorithm.

***Keywords*—**Keystroke dynamics, Typing behaviour, Mobile, Authentication, Biometrics

I. INTRODUCTION

Artificial Intelligence (AI) has changed the way how modern machines work. It is currently being used in all the major fields of work (such as Medicine, Business analytics, Digital Security, etc.). AI facilitates a machine to learn on its own and update by itself. These cognitive abilities of the machine are now considered as key to many of the real-world problems. To prevent data breaches, the digital resources (data, computing, network, etc.) are protected from the intruders using two processes: authentication and authorization. Authentication is the process of recognizing and verifying the identity of claimed user[1]. Authorization is the process of granting access to a resource. Authentication is broadly classified into three types:

1. Knowledge based (e.g. password, 4-digit pin)
2. Ownership based (e.g. access card, token)
3. Biometric based
    a) Physiological (e.g. finger print, Iris)
    b) Behavioral (e.g. hand writing, typing rhythms, voice)

Amongst the three, biometric-based authentication is an excellent way to verify a user’s identity. Unlike, passwords, tokens or smart cards, biometrics couldn’t be lost, stolen, or shoulder surfed [2]. Amid the Physiological and Behavioral, Physiological is static, it could be exploited using the existing technology. Thus, Behavioral biometrics has a better potential to be used for user authentication. A typical biometric system follows the methodology as shown in Figure 1.

It has two phases – enrollment and authentication. During the enrollment phase, the user data is collected, and after feature extraction, a classification model is applied to build up the user profile. While during the authentication phase the same classification model is used to the features extracted from the data acquired during the authentication phase. A fresh profile is built and matched with the existing one. If both profiles get matched the user is recognized as a legitimate user otherwise is rejected as being an imposter.

The paper is organized as follows. Section I gives a brief introduction to biometrics. In section II the basics of keystroke dynamics is discussed followed by a review of literature in section III. Section IV introduces the problem statement. In section V the experimentation conducted is discussed along with the results obtained. Lastly, in section VI the research is concluded by giving future directions.

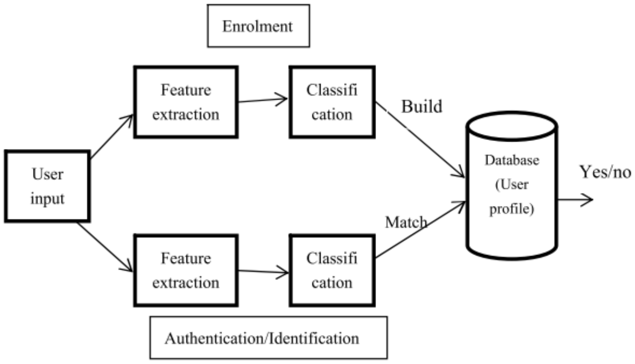

Figure 1: Biometric System

II. KEYSTROKE DYNAMICS

Keystroke Dynamics is behavioral biometrics that is based on user's typing patterns. Every user has a unique typing pattern on computer keyboard like that of user's signature [3]. The origin of keystroke dynamics dates to World War II where the sender of the telegraph is identified by a methodology known as the 'Fist of the Sender'[4]. It uses the rhythm, syncopation, and pace of the telegraph keys. Keystroke dynamics has the following advantages over the other authentication systems

- It does not need any extra hardware. (i.e., cost effective)
- It is highly non-intrusive.
- The user does not need to put any extra effort.

The above two or more methods are combined to achieve a strong authentication. Amongst the three, biometric-based authentication is an excellent way to verify a user's identity. Unlike, passwords, tokens or smart cards, biometrics couldn't be lost, stolen, or shoulder surfed. Amid the Physiological and Behavioral, Physiological is static; thereby it could be exploited using the existing technology. Thus, Behavioral biometrics has a better potential to be used for user authentication. This type of authentication is achieved by using machine learning algorithms such as neural networks. This paper provides brief insight of a typing rhythm based behavioral biometrics, Keystroke Dynamics.

Authentication using keystroke dynamics can be static of continuous. In static authentication, the user is authenticated only once during login while in a continuous method the user is monitored during the entire session.

The most common features used for building a user profile are derived from press and release time of adjacent keys as shown in Figure 2.

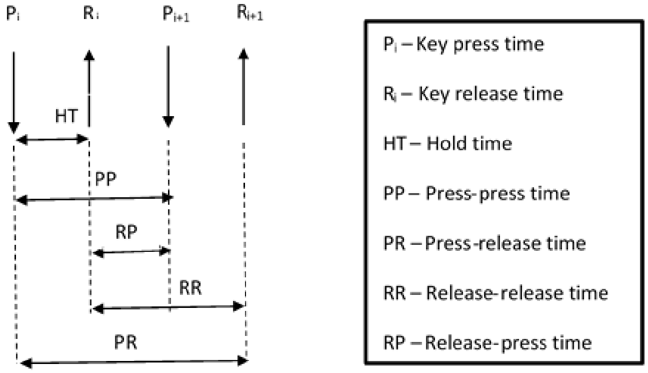

Figure 2: Keystroke Features

Apart from these, the pressure of key press, trigraph and typing speed are also used for building a user profile. In case of mobile phone, the angular axis readings are also considered.

The performance of such a system is judged based on three error rates: false acceptance rate (FAR), false rejection rate (FRR) and an equal error rate (EER). The FAR is the measure of successful imposter attempts. The FRR is the measure of unsuccessful genuine attempts. The EER is the value at which FAR becomes equal to FRR.

III. LITERATURE REVIEW

Authors in [2] present the working of various algorithms used in keystroke dynamics. The authors built a C++ toolkit by using xview library for analyzing the data. This toolkit facilitates a quick way to create rough properties on the data set by dividing the users into distinct groups. They have explained the formulae used in the following algorithms: Euclidean distance measure, Weighted probability measure, Non-weighted probability, Bayesian-like classifier. They achieved best identification rate 87.18% (approximately) using the weighted probabilistic classifier on a dataset of 63 users. The identification rated obtained by using Euclidean distance and the non-weighted scoring approach is 83.22% and 85.63% respectively. They favor the using structured text (transcription) instead of arbitrary text (i.e., free composition).

In [5] the authors have investigated the ability of neural networks to successfully authenticate users, according to their interactions with a mobile phone's keypad. Using PIN code, they got 18.1% FAR, 12.5% FRR and 15% ERR. They suggested that Keystroke dynamics can become a part of a hybrid authentication system.

Sim & Janakiraman in [6] found that word specific digraphs are good for free text keystroke dynamics. They have also proved that typing behavior of the user is dependent on context (E.g. the user's typing of the word “IN” as a whole word is different from “IN” in “BEGIN”). Of the top ten most frequently used words, they obtained complete classification with the word “IN”. Compared to non-word specific n- graphs, word specific n-graphs perform better. The authors defined a finger set to identify trained users and non-trained users.

In [7] detectors(algorithms) were compared on a dataset that was collected from 51 users with 400 entries of the passphrase “.tie5Roanl” in 8 sessions(50 entries per session). They used Windows XP operating system with an external reference clock that is accurate up to ±200 microseconds. They achieved best performance (least EER) with Manhattan distance.

In [8] to overcome the problems of shorter PIN length (e.g. three characters long), they have used artificial rhythms and tempo cues, by using these they obtained better performance with reduction of error (EER) from 13% to 4%. They suggested applying the analysis to more diverse group of users. They mentioned that in future one amongst the FAR and FRR would be important and the issue could be addressed by selecting proper threshold. They have also shown following accepted hypothesis: (a) Artificial Rhythms with cues are useful for users using both the hands for typing. (b) The average error rate (EER) involving Artificial Rhythms with cues is lower than Natural Rhythm without cues.

The research by Giot et. al. [9] focuses on benchmarking datasets to help researchers directly work on their proposed algorithm instead of spending time for data collection. They developed a software “GREYC-Keystroke” for data collection. It can be download from(“GREYC-KeyStroke Dataset | E-Payment &amp; Biometrics Research Unit,” n.d.). The software can be used to share the collected data with the world. They observed that there were huge number of failures during static data acquisitions. The reasons were already cited earlier, in the report. Their database is available for free and can be used to compare the efficiency of the methods used for identification in keystroke dynamics.

In [10] the BeiHang database is studied. It is like GREYC Keystroke database, which is open for public. But to use the existing database, a request should be sent to the corresponding author. It is at (“BeiHang Dataset,” n.d.). They used Gaussian, Neural network classifier and variants of SVM algorithm to compare their performance against two variants of databases. They obtained best results with SVM expansion reduction method with an ERR of 11.8327 and 26.6014 for database A and B respectively.

Bhatt & Santhanam [1], explains the 2013's trends of keystroke dynamics. They explained the components and working of the biometric system. This paper cites the research done by various researchers. This paper mentions that keystroke dynamics can be used for finding if a person's influence of alcohol or not. They mention that the appropriate length of password needs to be determined, to authenticate the user easily. Also, the template stored to authenticate the user needs to be able to adapt themselves with a change in behavioral nature of the user.

IV. PROBLEM

To authenticate a user successfully, the keystroke data should be consistent and discriminable. Artificial cues (such as Pauses and musical rhythms) had proven to be successful in achieving high discriminability. With the hypothesis that user type their own password more consistently and uniquely than any regular text, we would like to investigate (research) if the user's password or artificially rhythmed text gives better authentication results.

V. EXPERIMENTATION AND RESULTS

The sample data is collected from 9 individuals using an android application (that only accepts input without typographical errors) in an unconstrained environment (Each candidate installed the app in his/her mobile). The android application collects the accelerometer data along with the timestamp of each key press and release. The artificially rhythmed string is based on corresponding user specific password. Data Entries: Each candidate typed ten user-specific passwords (including their own password) and artificially rhythmed string. is typed 20 times, amounting to 400 entries per user.

From the keystroke timestamps data, the hold time and flight time features are extracted. Hold time is computed by calculating the difference between the time stamps of press and release of the key. The flight time is calculated as the difference between release timestamp of a key and immediate press timestamp of the next key. From the sensor data, the mean, median and standard deviation of acceleration along x, y and z axis for each password entry is extracted. The extracted features are used to train Bagged Trees classifier to categorize users according to their typing rhythm.

Figure 3 shows the methodology adopted to carry out the research. The hold time and flight time are clubbed together and are referred to as keystroke features in the rest of the paper while the features extracted from the axis reading are referred to as accelerometer features.

Table 1: Previous Research Work done in the field of Keystroke Dynamics

Author	F	Algorithm	E	Text Type	N#	S#	FAR	FRR
Robinson et al[11]	1,2	Minimum intra-class distance, non-linear, Inductive learning	D	U	20	-	9	10
Haider et al. [12]	1	Fuzzy logic, Neural Network & statistical techniques	S	C	-	-	6	2
Gutirrez et al.[13]	1	Naive Bayesian, Statistical Decision Trees, Instance based	S	C	-	-	1.6	14.3
Kacholia & Pandit[14]	1	Random dist. function	S	-	20	440	3.4	2.9
Arau´jo et al.[15]	1,2	Fuzzy Logic	S	C	10	200	3.4	2.9
Eltahir et al. [16]	1,3	Autoregressive	S	C	22	22	5	-
Sang et al.[17]	1	Support Vector Machine	S	-	10	-	0.02	0.1
Joshi & Phoha [18]	1	Competition between Self Organizing Maps for Authentication	S	C	43	873	0.88	3.55

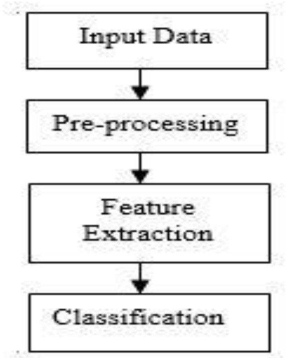

Figure 3: Research Methodology

Table 2 shows the accuracy of the system. The accuracy is much better if accelerometer data is clubbed with keystroke features. The accuracy is better as the user is more comfortable while typing the text of his/her choice. Also, the accuracy is better with user-specific input as compared to artificial rhythm. With a combination of features, accuracy of 88.22% is achieved which is 28% higher than just using keystroke features. This shows that the angle at which the device is held plays an essential role in determining the user typing pattern.

Table 2: Accuracy with accelerometer data

	Keystroke features	Accelerometer + keystroke features
User Specific	60.66%	88.22%
Artificial rhythm	57.66%	75%

Further analysis was carried by optimizing using a genetic algorithm. Optimizing reduced the final set of features from 27 to 21 and the accuracy of the system was increased to 90.22%. Table 3 shows the results after optimization.

	Keystroke features	Accelerometer + keystroke features

User Specific	70%	90.22%
Artificial rhythm	60.23%	80.41%

Figure 2 and Figure 3 shows the ROC curve for user specific. It can be seen that the results of the ROC curve are much better after optimization as against the results without optimization.

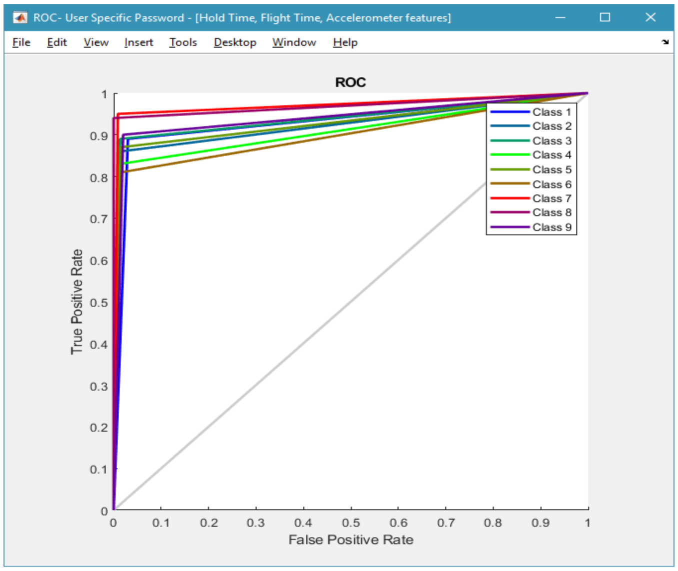

Figure 4: ROC without optimization

VI. CONCLUSION AND FUTURE SCOPE

This report attempts to present an overview of research on keystroke dynamics over the past few decades along with the new research results answering the questions: Does the text typed by the user has an impact on keystroke dynamics system? Can we embed the sensor-based data to extract relevant features, to improve the performance of the keystroke dynamics-based user authentication system?

From the results, the Bagged Tree classifier has shown better performance with user-specific password dataset than an artificially rhythmed (random) password. Thus, the text typed by the user does have impact on keystroke dynamic system's performance.

Features extracted from accelerometer data improved the performance of the keystroke dynamic system. Thus, the mobile-based authentication would be better with sensory data combined with keystroke data.

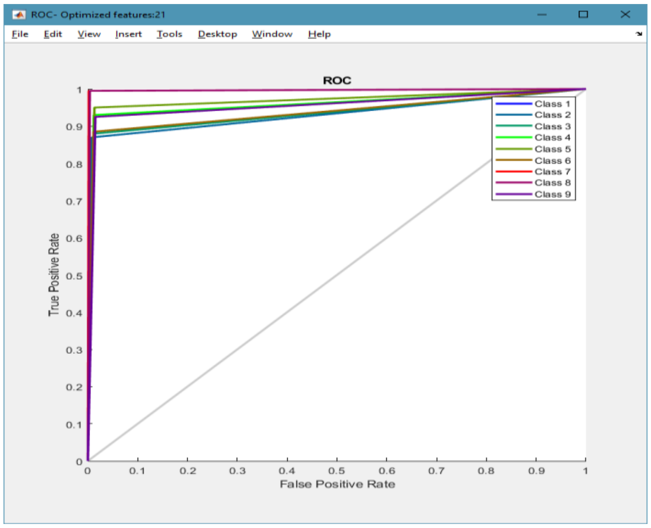

Figure 5: ROC after optimization

REFERENCES

[1] S. Bhatt and T. Santhanam, "Keystroke dynamics for biometric authentication - A survey," 2013 Int. Conf. Pattern Recognition, Informatics Mob. Eng., pp. 17–23, 2013.

[2] F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication," Futur. Gener. Comput. Syst., vol. 16, no. 4, pp. 351–359, 2000.

[3] I. Bhardwaj, N. D. Londhe, and S. K. Kopparapu, "Study of imposter attacks on novel fingerprint dynamics based verification system," IEEE Access, vol. 5, pp. 595–606, 2017.

[4] R. Solanki and P. Shukla, "Estimation of the User's Emotional State by Keystroke Dynamics," Int. J. Comput. Appl., vol. 94, no. 13, pp. 21–23, 2014.

[5] N. L. Clarke, S. M. Furnell, B. M. Lines and P. L. Reynolds, "Keystroke dynamics on a mobile handset: a feasibility study," Information Management & Computer Security, vol. 11, no. 4, pp. 161–166, 2003.

[6] T. Sim and R. Janakiraman, "Are digraphs good for free-text keystroke dynamics?," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.

[7] K. S. Killourhy and R. Maxion, "Comparing Anomaly Detectors for Keystroke Dynamics," IEEE Computer Society Press, Los Alamitos, California, pp. 125–134, 2009.

[8] S. Hwang, S. Cho, and S. Park, "Keystroke dynamics-based authentication for mobile devices. Computers and Security, vol. 28, no. 2, pp. 85–93, 2003.

[9] R. Giot, M. El-Abed and C. Rosenberger, "GREYC keystroke: A benchmark for keystroke dynamics biometric systems. IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems, BTAS, 2009.

[10] Y. Li, B. Zhang, Y. Cao, S. Zhao, Y. Gao and J. Liu, "Study on the BeiHang Keystroke Dynamics Database," In International Joint Conference on Biometrics, pp. 1-5, 2011.

[11] J. A. Robinson, V. M. Liang, J. A. M. Chambers, and C. L. MacKenzie, "Computer user verification using login string keystroke dynamics," IEEE Trans. Syst. Man, Cybern. Part ASystems Humans., vol. 28, no. 2, pp. 236–241, 1998.

[12] S. Haider, A. Abbas, and A. K. Zaidi, "A multi-technique approach for user identification through keystroke dynamics," IEEE Int. Conf. Syst. man Cybern., vol. 2, pp. 1336–1341, 2000.

[13] F. Gutiérrez, “Biometrics and data mining: Comparison of data mining-based keystroke dynamics methods for identity verification,” MICAI 2002 Adv. …, pp. 460–469, 2002.

[14] V. Kacholia and S. Pandit, “Biometric Authentication using Random Distributions (BioART),” Proc. 15th Can. IT Secur. Symp. Canada, vol. 5, pp. 1–8, 2003.

[15] L. C. Freire Araujo, M. Gustavo Lizarraga, L. H. Rabelo Sucupira, J. B. Tadanobu Yabu-uti, and L. L. Lee, “Typing Biometrics User Authentication based on Fuzzy Logic,” Lat. Am. Trans. IEEE (Revista IEEE Am. Lat., vol. 2, no. 1, pp. 69–74, 2004.

[16] W. E. Eltahir, M. J. E. Salami, A. F. Ismail, and W. K. Lai, “Dynamic keystroke analysis using AR model,” Industrial Technology, 2004. IEEE ICIT ’04. 2004 IEEE International Conference on, vol. 3. p. 1555–1560 Vol. 3, 2004.

[17] Y. Sang, H. Shen, and P. Fan, “Novel Impostors Detection in Keystroke Dynamics by Support Vector Machine,” Lect. Notes Comput. Sci., vol. 3320, pp. 666–669, 2004.

[18] S. S. Joshi and V. V. Phoha, “Investigating hidden Markov models capabilities in anomaly detection,” in Proceedings of the 43rd annual southeast regional conference on - ACM-SE 43, vol. 1, 2005.

Authors Profile

Mr. B. S. Saini did his Bachelor of Technology from Guru Gobind Singh Indraprastha University, New Delhi and completed his Masters in Technology from Guru Nanak Dev University, Amritsar. He is currently pursuing his Ph,D in the field of Keystroke Dynamics from Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab. He is currently working as Assistant Professor in Lovely Professional University, Phagwara, Punjab. He has a total teaching experience of 9 years and research experience of 4 years.

Dr. N. Kaur is serving as Professor in the Deaprtment of Computer Science, Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab. She is doctorate from IIT Roorkee in the field of distributed database. Her research interests are information security, mobile computing, cloud computing and software engineering.She has published over 100 papers in various journals and proceedings.

Dr. K. S. Bhatia is serving as Assistant Professor in the Deaprtment of Electronics and Communication Engineering, Inder Kumar Gujaral Punjab Technical University, Batala Campus. He is a researcher and a prolific author. There are about more than 80 research papers and four Books of International level into his credit. He is doctorate in the field of Optical-OFDM and wireless communication from a reputed university of INDIA. He has guided 29 research students at M. Tech level and 04 at Ph. D. level are in process. He has about 13 years of experience in teaching and research.