ISSN: 0971-1260
Vol-22-Issue-16-August-2019

A Brief Review on Keystroke-Based User Identification System

Baljit Singh Saini
Associate Professor, Lovely Professional University, Phagwara, India

Navdeep Kaur
Professor, CSE, Sri Guru Granth Sahib World University, Fatehgarh Sahib, India

Cherry Khosla
Assistant Professor, CSE, Lovely Professional University

ABSTRACT

With the advent of artificial intelligence, the technology has transformed into smart and sophisticated technology. Since few decades, the simple password authentication has been either replaced or compounded with biometrics (such as Facial Recognition, Fingerprint Scan etc.) to provide better security. Now, the onetime authentication systems are shifting towards a smart (i.e., continuous) authentication system. Keystroke Dynamics is behavioral biometrics that can perform continuous authentication to detect intruders. In this paper, the origin and various works carried out in Keystroke Dynamics is discussed.

Keywords: Keystroke Dynamics, Authentication, Biometrics, Identification, Typing Pattern, Artificial Intelligence, Password, Typing Rhythm

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. Data Security is considered one of the major problems of digital world. For example, On 3rd October 2017, Yahoo revealed that the August 2013, data breach of Yahoo has affected over 3 billion user accounts.To prevent these data breeches, the digital resources (data, computing, network etc.) are protected from the intruders by means of 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)

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

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 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 in nature, 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.

KEYSTROKE DYNAMICS

Keystroke Dynamics is behavioral biometrics that is based on user's typing patterns. Every user has unique typing pattern on computer keyboard like that of user's signature [3]. It uses the rhythm, syncopation and pace of the telegraph keys. The researchers' effort has made enhanced commercial software possible. E.g. Biopasswor, This uses a technique called password hardening where the user's typing rhythm is checked along with the password to verify the user.

Keystroke dynamics has the following advantages over the other authentication systems:

1. It doesn't need an extra hardware. (i.e., cost effective)
2. It is highly non-intrusive.
3. The user doesn't need to put any extra effort.
4. Continuous authentication can be done throughout the session.
5. It can be used in an intrusion detection system

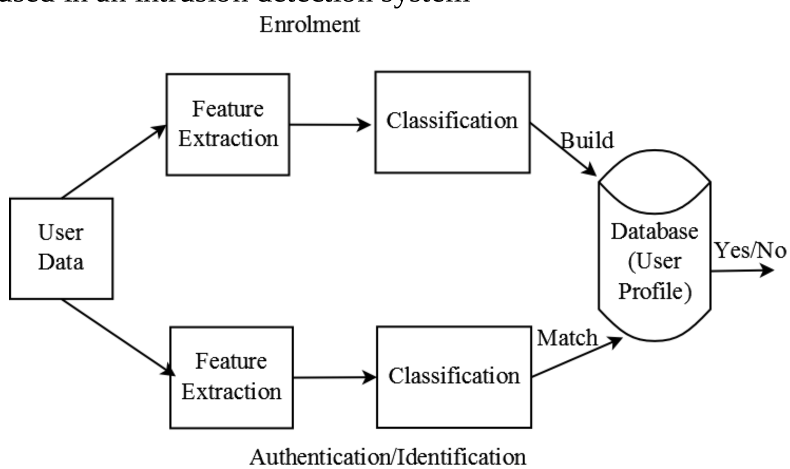

Figure 1: Biometric System

A typical keystroke dynamics system consists of two phases with four modules (see Figure 1):

1. Enrollment Phase: In this phase, the user's data is recorded to extract features that are unique to the user and then store them in the database.
2. Authentication Phase: Here, the user's input is validated with the saved data in the database.

User Input

The input text length, clock precision and training data influence the performance of keystroke dynamics system [4]. The user input can be either static or dynamic. Static input consists a fixed

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

structured text (transcription) that should be entered by user. Whereas in dynamic input the user can either have a fixed text or enter random text (free composition) in every input, i.e. the user may enter different text in both the phases. The former type of input is used for entry level authentication purpose i.e. at the login time only [1]. The dynamic can be used for both, entry level authentication as well as continuous authentication (free composition). According to an earlier study [5], there is no significant difference between the two methods, transcription and free composition. The raw data of the input consists of the key and its press and release timestamps. If mistakes are not allowed during acquisition of user password [6], a huge number of failures occur.

These erroneous captures are due to several reasons:

1. Long length of the password (17 characters).
2. Difference in the user keyboard i.e. the user is not accustomed with the keyboard used for getting the input.
3. User's state of mind i.e. forgetting password, disturbed by the environment etc.

When the keystroke latencies were paired with key pressure, the error rate has decreased from 12.2 to 6.9 in [7]. Thereby pressure can also be taken during the user input for better authentication. To measure pressure, special keyboards which are generally expensive than the standard keyboard are required. In general, desktop computers, laptops and smartphones are not pressure sensitive. Therefore, majority of the researchers limit their work to standard keyboard only. Notably, the environment plays a vital role in influencing an individual's typing behavior. Depending on the computer's operating system, resolution of data captured, keyboard dimensions, device, posture, lighting condition, the user's typing rhythm may be affected. For example, In MS Windows, the key press and key release keyboard events could not be distinguished if their latency is below 15.625ms [5]. An environment where all the parameters that influence an individual's typing rhythm are relatively neutral to all the users is known as controlled environment. Otherwise it is known as uncontrolled environment. Uncontrolled environment provides a realistic scenario to test and deploy the system, the first experiment was performed by[8]. They mentioned that the separate data recorded at different time is more reliable. Controlled environment can be used for gaining knowledge by comparing various approaches and algorithms used in making the system. To facilitate peer review, publishing the collected dataset and specifying the environment conditions (such as OS and applications used, computer's configurations, its work load etc.) would yield comparable results.

Use of keystroke dynamics for smart phones has considerable benefits as compared to its use in hard keyboard-based systems [9]. Comparatively, less number of fingers are used with soft keyboard of smartphones. The position has no considerable impact on the user's typing rhythm as the smartphone would be handled in the same manner in different postures. Due to the high availability of movement sensors (such as gyroscope, air pressure, accelerometer etc.) that had been proven effective for authentication [10], sensor enhanced keystroke dynamics came into existence. In this, the sensor data must be collected along with traditional keystroke rhythms.

Feature Extraction

Ever since the debut of the term, Keystroke Dynamics, latency has been the most commonly used feature by early researchers [11]. The basic types of latencies are:

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

1. Press to Press (PP) or Digraph latency: It is the time elapsed between two successive key presses.
2. Press to Release (PR): It is the time delay between a key press and its release.
3. Release to Press (RP): It is the latency between release of a key and press of next key.
4. Release to Release (RR): It is time duration between two consecutive key releases.

The first feasibility study to use the above features was conducted by [12] at Rand Corporation.

The following latencies are considered when more than two keys are used for feature extraction:

1. Tri-Graph latency: three keys are used to derive the metrics.
2. N-Graph latency: n-keys are used to derive the metrics.

Using these features either individually or in combination, sub-features (like typing speed, key pressure etc.) can be derived [13]. It is observed that trigraph latency gives better classification results than digraphs [11, 14] found that the hold times are very important than inter key timings. Figure 2 shows the various latencies.

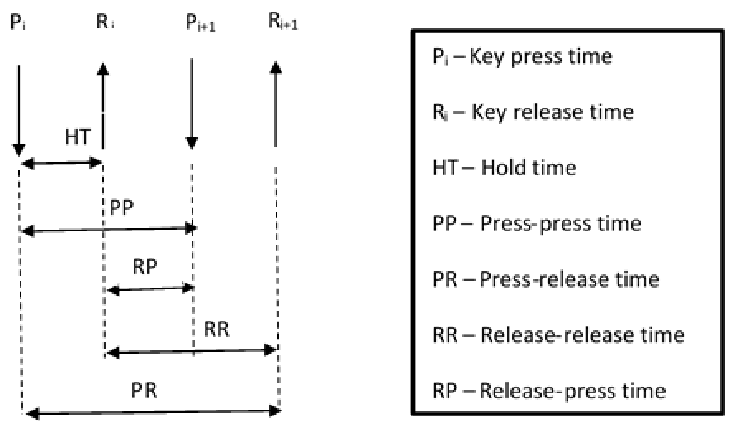

Figure 2: Most common features extracted in keystroke dynamics

The performance of keystroke dynamics system is measured by using the following three basic types of errors:

- False Acceptance Rate (FAR): It is the percentage of erroneously accepted imposters (invalid users) as legitimate user.
- False Rejection Rate (FRR): It is the percentage of legitimate users who are incorrectly categorized as imposters.
- Equal Error Rate (EER) or Cross Error Rate (CER): It is the value at which FAR and FRR are equal. Figure3 shows the relation between FAR, FRR and ERR.

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

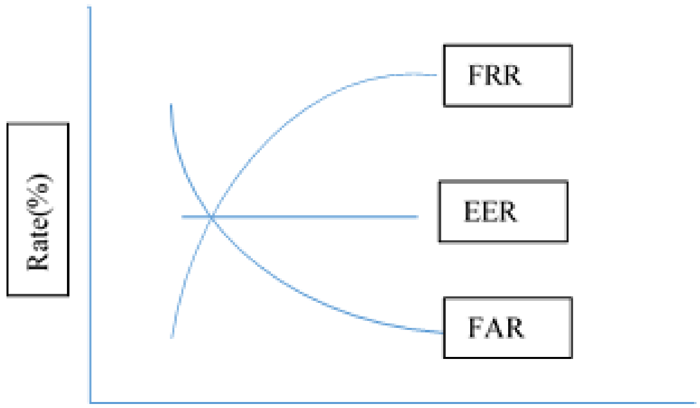

Figure 3: Relationship between FAR, FRR and EER

The lesser these values are, the better is the performance of the system.

TABLE 1
STATISTICAL METHODS WITH FAR AND FRR VALUES

Author	Algorithm	E	Text Type	No. of subjects	Samples	FAR	FRR
[12]	T-test	C	S	7	36	0	4
[15]	Mean and Standard Deviation latency	C	S	17	34	6	12
[16]	keystroke Digraph Latencies	C	S, D	17	72	5	5.5
[17]	Bayes Classifier	C	S	32	171	2.8	8.1
[18]	Absolute distances	C	S	33	975	0.25	16.36
[19]	Distance Measure	U	D	31	-	≈0	≈ 2.0
[20]	Mean	U	S	11	880	0	-

E- Environment, S - Static, D - Dynamic,C - Controlled, U - Uncontrolled.

CLASSIFICATION

Classification aims at finding the best category/group which is closest to a classified pattern. The extracted features of the user in the enrollment phase are used to create template of the user behavior (profile). During the authentication phase, the extracted features are compared with the existing user template to find the best match to determine if the user is legitimate user or imposter. There are four major classification methods:

1) STATISTICAL METHODS
   In the earlier times, these were the most popularly used method by majority of the researchers. To build templates, simple statistical methods like mean and standard deviation were used. For comparison of hypothesis, t-test or distance measures like Euclidean or Manhattan distance were used. Table 1 shows various statistical methods used by various researchers along with their results.
2) ARTIFICIAL NEURAL NETWORKSArtificial Neural Networks (also referred as

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

TABLE 2
NEURAL NETWORK APPROACHES

Author	Algorithm	E	Text Type	No. of subjects	Samples	FAR	FRR
[21]	Perceptron Algorithm	C	S	24	1400	8	9
[22]	Potential Function	C	S	30	6750	2.2	4.7
[23]	Minimum Intra Class Distance Classifier	-	-	137	-	23	24
[24]	Backpropagation	C	S	151	-	1.11	0
[25]	ART-2	U	S	50	-	-	~10
[26]	Backpropagation	-	S	200	-	0	1
[27]	Convolutional neural network	U	D	10	1000	0	6
[28]	Multi Layer Perceptron	C	S	7	91	0.3	1.5

E- Environment, S - Static, D - Dynamic, C - Controlled, U - Uncontrolled

Neural Networks) are adaptive non-linear statistical learning algorithms that are inspired from biological neural networks. This type pf learning can be either supervised or unsupervised. The backpropagation algorithm is most popular method in supervised learning. Table 2 shows the list of research conducted in keystroke dynamics using neural network algorithm.

3) PATTERN RECOGNITION

Pattern recognition aims to classify patterns into different classes. Several pattern recognition techniques had been proposed and used by various researchers. It contains algorithms such as the Fishers linear discriminant (FLD), Bayes classifier and support vector machine (SVM). Table 3 displays various pattern recognition algorithms that are being used by researchers.

4) OTHER TECHNIQUES

The optimizing techniques like Evolutionary algorithms (e.g. genetic algorithm) and hybrid algorithms fall into this category. Ant colony optimization and Extreme machine learning achieved a feature reduction of 46.51% compared to Particle swarm optimization and Genetic Algorithm [35]. To analyze hold time and latency, the concept of artificial rhythm with cues was used by [36] and obtained EER of 4%. In free text(continuous) authentication, unstable (noisy/ gibberish) keystrokes are generated from user activities like playing game etc. It was found that this noisy data increases FRR. Filtering this type of data using spelling checker yielded better FRR instead of using regular expression [37]. With Ant Colony Propagation, [13] achieved classification error of 0.059 and Accuracy of 92.8. Though the hybrid techniques yield good results [39], yet they consume more system resources for computation. If the complexity of the combined algorithm is exponential or factorial in nature then a standalone application for system with limited resources (e.g. mobiles, laptops etc.) wouldn’t be possible.

 ISSN: 0971-1260

TABLE 3
PATTERN RECOGNITION ALGORITHMS AND THEIR RESULTS

Author	Algorithm	E	Text Type	No of users	S	FAR	FRR
[23]	Minimum intra-class distance, non-linear, Inductive learning	D	U	20	-	9	10
[29]	Fuzzy logic, Neural Network & statistical techniques	S	C	-	-	6	2
[30]	Random dist. function	S	-	20	440	3.4	2.9
[31]	Fuzzy Logic	S	C	10	200	3.4	2.9
[32]	Autoregressive	S	C	22	22	5	-
[33]	Competition between Self Organizing Maps for Authentication	S	C	43	873	0.88	3.55
[7]	Hidden Markov Models	S	C	20	600	0	-
[34]	Parallel decision trees	S	C	43	387	0.88	9.62
[38]	Fuzzy, Neural Network	S	C	25	500	1	1.5

E- Environment, S- Samples, S - Static, D - Dynamic, C - Controlled, U – Uncontrolled.

CONCLUSION

This paper attempts to present an overview of research on keystroke dynamics over the past few decades along with the research proposal. Keystroke Dynamics is a behavioral biometric authentication technique that has great potential for becoming the standard authentication model in near future. It offers continuous authentication unlike the traditional login time authentication system. The boundaries for application of keystroke dynamics are not yet well defined. The following are a few applications of keystroke dynamics:

- Detecting the age of a person.
- Detecting the mood of the person.
- Both, One time and continuous Authentication of User.
- Determine if a person is under the influence of alcohol or not.

Though there is significant increase in this field of research since 2002, yet the field has many unexplored parameters. The FAR, FRR and ERR of keystroke dynamics is usually high that it do not meet standards in some standard access control systems, such as the EN-50133-1(Kevin S Killourhy & Maxion, 2009). There are many parameters such as keyboard dimensions, operating system used, light conditions, human mental health/condition, the device being used etc., that influence the keystroke biometric system. The research work considering more parameters might yield better results. This is a promising field to carry out the research work. For e.g. the following areas can be explored:

- The statistical influence of operating system on keystroke biometrics.
- Estimate the percentage of alcohol consumed by a user using keystroke dynamics
- Determine the human personality using key patterns of the user.
- Security standards to govern the use of keystroke dynamics as a web-based service.
- Determine the optimum length of the string to be used for keystroke-based user identification system.

REFERENCES

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

[1] S. Bhatt, and T. Santhanam, “Keystroke dynamics for biometric authentication - A survey,” International Conference on Pattern Recognition, Informatics and Mobile Engineering, pp. 17–23, 2013.
[2] R. Young, "Keystroke Dynamics: Establishing Keyprints to Verify Users in Online Courses." Computers in the Schools, vol. 36(1), pp. 48-68, 2019.
[3]I. Bhardwaj, N. 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] N. F. Sadikan, R.Azizul, and M. Farhan."A survey paper on keystroke dynamics authentication for current applications," AIP Conference Proceedings. vol. 2173 (1), pp. 1-12, 2019
[5] P. H. Pisani and A. C. Lorena. A. C. “A systematic review on keystroke dynamics. Journal of the Brazilian Computer Society,”vol. 19(4), pp. 573–587, 2013.
[6]F.Didem, and T. Omer. "An Exploration of Machine Learning Methods for Biometric Identification Based on Keystroke Dynamics," Artificial Intelligence and Machine Learning Applications in Civil, Mechanical, and Industrial Engineering. pp. 258-269, 2019.
[7] T. Y. Chang, C. J. Tsai, and J. H. Lin, “A graphical-based password keystroke dynamic authentication system for touch screen handheld mobile devices,” Journal of Systems and Software, vol. 85(5), pp. 1157–1165, 2012.
[8] H. Lee, "A parameterized model to select discriminating features on keystroke dynamics authentication on smartphones," Pervasive and Mobile Computing,vol. 54, pp. 45-57, 2019.
[9] B. S. Saini, N. Kaur and K. Bhatia, "Analyzing user typing behaviour in different positions using keystroke dynamics for mobile phones," Journal of Discrete Mathematical Sciences and Cryptography, vol. 22(4),pp. 591-603, 2019.
[10] I Lamiche, G. Bin, Y. Jing, Z. Yu and A. Hadid."A continuous smartphone authentication method based on gait patterns and keystroke dynamics," Journal of Ambient Intelligence and Humanized Computing, vol. 10(11), pp. 4417-4430, 2019.
[11] S. P. Banerjee and D. Woodard.“Biometric Authentication and Identification Using Keystroke Dynamics: A Survey,” Journal of Pattern Recognition Research, vol. 7(1), pp. 116–139, 2012.
[12] R. S. Gaines, S. J. Press, and N. Shapiro.“Authentication by keystroke timing some preliminary results. No. RAND-R-2526-NSF. RAND CORP SANTA MONICA CA, pp. 1–51, 1980.
[13] M. Karnan, M. Akila, and N. Krishnaraj.“Biometric personal authentication using keystroke dynamics: A review,” Applied Soft Computing Journal, vol. 11(2), pp. 1565–1573, 2011.
[14] A. Kołakowska, "Usefulness of keystroke dynamics features in user authentication and emotion recognition." Human-Computer Systems Interaction,pp. 42-52, 2018.
[15] D. Umphress and G. Williams.“Identity verification through keyboard characteristics,“ International Journal of Man-Machine Studies, vol. 23(3), pp. 263–273, 1985.
[16] J. Leggett and G. Williams.“Verifying identity via keystroke characteristics,” International Journal of Man-Machine Studies, vol. 28(1), pp. 67–76, 1988.
[17] S. A. Bleha and M. S. Obaidat.“Computer users Verification Using the Perceptron Algorithm,” IEEE Transactions on Systems, Man and Cybernetics, vol. 23(3), pp. 900–902, 1990.
[18] R. Joyce and G. Gupta. “Identity authentication based on keystroke latencies,” Communications of the ACM, vol. 33(2), pp. 168–176, 1990.

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

[19] D. Gunetti and C. Picardi.“Keystroke analysis of free text,” ACM Transactions on Information and System Security, vol. 8(3), pp. 312–347, 2005.
[20] S. Giroux, R. S. Wachowiak and M. P. Wachowiak. “Keystroke-based authentication by key press intervals as a complementary behavioral biometric. IEEE International Conference on Systems, Man and Cybernetics, pp. 80–85, 2009.
[21] S. Bleha, C.Slivinsky, B.Hussien, A. Shoukry, J. Feder, and T. Fountain.“Computer-access security systems using keystroke dynamics,” Pattern Analysis and Machine Intelligence, IEEE Transactions On, 12(12), 1217–1222, 1990.
[22] M. S. Obaidat andB. Sadoun. “Verification of computer users using keystroke dynamics,” IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 27(2), pp. 261–269, 1997
[23] J. A. Robinson, V. M. Liang, J. A. Chambers, and C. L. MacKenzie.“Computer user verification using login string keystroke dynamics,” IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans, vol. 28(2), pp. 236–241, 1998.
[24] D. T. Lin. “Computer-access authentication with neural network based keystroke identity verification,” International Conference on Neural Networks, vol. 1, pp. 174–178, 1997.
[25] J. Bechtel, G. Serpen, and M. Brown.“Passphrase authentication based on typing style through an ART 2 neural network,” International Journal of Computational Intelligence and Applications, vol. 2(02), pp.131-152, 2002.
[26] C. V. Priya and K. S. Angel. "Keystroke dynamics based authentication framework using back propagation neural network," International Journal of Pure and Applied Mathematics vol. 119(10),pp. 1097-1104, 2018.
[27] C-H, Lin, J-C, Liu, and K-Y Lee. "On Neural Networks for Biometric Authentication Based on Keystroke Dynamics." Sensors and Materials,vol. 30(3),pp. 385-396, 2018.
[28] A. Salem andS. Mohammad. "A novel security scheme for behavioral authentication systems based on keystroke dynamics," Security and Privacy, vol. 2(2), pp. 212-219, 2019.
[29] S. Haider, A. Abbas, and A. K. Zaidi. “A multi-technique approach for user identification through keystroke dynamics,” IEEE International Conference on Systems, Man and Cybernetics, vol. 2, pp. 1336–1341, 2000.
[30] V. Kacholia and S. Pandit.“Biometric Authentication using Random Distributions,” Proceedings of the 15th Canadian IT Security Symposium, pp. 1–8, 2003.
[31] L. C. Freire Araujo, L. M. Gustavo and L. H. Sucupira, J. B. Tadanobu and L. L. Lee.“Typing Biometrics User Authentication based on Fuzzy Logic,“Latin America Transactions, IEEE, vol. 2(1), pp. 69–74, 2004
[32] W. E. Eltahir, M. J. Salami, A. F. Ismail, and W. K. Lai.“Dynamic keystroke analysis using AR model,” IEEE International Conference on Industrial Technology, vol. 3, pp. 1555-1560, 2004
[33] S. S. Joshi and V. V. Phoha. “Investigating hidden Markov models capabilities in anomaly detection,” In Proceedings of the 43rd annual southeast regional conference, vol. 1, pp. 98-102, 2005
[34] Y. Sheng, V. V. Phoha, and S. M. Rovnyak.“A parallel decision tree-based method for user authentication based on keystroke patterns,” IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 35(4), pp. 826–833, 2005
[35]D. Shanmugapriya and G. Padmavathi, “A Survey of Biometric keystroke Dynamics: Approaches, Security and Challenges,” International Journal of Computer Science and Information Security, vol. 5(1), pp. 5-13, 2009

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

[36] Hwang, S. seob, Cho, S., & Park, S. (2009). Keystroke dynamics-based authentication for mobile devices. Computers and Security, 28(1–2), 85–93, 2009

[37] J. Huang, D. Hou and S.Schuckers.“Effects of text filtering on authentication performance of keystroke biometrics,” IEEE International Workshop on Information Forensics and Security, pp. 1-6, 2017.

[38] P. M. Koh and W. Kin. "Keystroke Dynamics Identification System using ABC Algorithm," IEEE International Conference on Automatic Control and Intelligent Systems, pp. 108-113, 2019.

[39] C. Dwivedi, D.Kalra, D. Naidu and S. Agarwal. "Keystroke dynamics based biometric authentication: A hybrid classifier approach," IEEE Symposium Series on Computational Intelligence, pp. 266-273, 2018.