IJCSE International Journal of Computer Sciences and Engineering Open Access

Review Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

Exploration of Keystroke Dynamics Based Authentication on Fixed-Text and on Free-Text

M. Rathi[1*], A. V. Senthil Kumar[2]

[1]Department of Computer Technology, Dr. NGP Arts and Science College, Coimbatore, India
[2]Department of MCA, Hindusthan College of Arts and Science, Coimbatore, India

[*]*Corresponding Author: rathi.vidu@gmail.com, Tel.: +91-0422-2369259*

Available online at: www.ijcseonline.org

Accepted: 06/Jan/2019, Published: 31/Jan/2019

Abstract— Computer security is the protection of computing systems and the data that is stored or accessed. It is very important to ensure that the information remains confidential and only those who should access that information can. Username and password alone are insufficient in complex applications. Hence, a strong authentication method such as Biometric authentication method is required to verify one's identity using the unique biological characteristics of an individual. Existing security approaches can be strengthened by one of the behavioral biometric based technique known as Keystroke Dynamics. The main objective of this paper is to explore particularly on Keystroke Dynamics based Authentication (KDA). This technique can be applied in different domains like intrusion detection, online learning and assessment, e-banking etc., to authenticate the users. This paper presents a review of its applications using fixed-text (static passwords) and free-text (continuous). Comparing these two types of keystroke authentication methods, Free-text kind of authentication process was found to be better as it is not limited with username and password during the log-in session; it is continued until the end of the log-on session.

Keywords— Authentication, Keystroke dynamics, Behavioral Biometrics, Fixed Text, Free text, Biometric authentication

I. Introduction

Usage of the Internet for online shopping, e-learning, social interactions and net-banking have grown rapidly which poses security as a major concern. Networks are sometimes weak, which are most vulnerable and hijacked easily. Efforts are taken to secure the network through Network topologies and security protocols. The recognition process authenticates the validity of the user to determine if that user is a valid user or not. Authentication is the process of determining whether someone or something is, in fact, who or what it is declared to be. As the dependence upon computers and computer networks grows, the need for authentication has increased. Figure 1 shows the different types of user authentication methods. There are three conventional methods used for authenticating a person. These are possession based, knowledge based, and biometrics based. Possession based includes anything that a 'user must have' in their possession to log in, such as a smart-card, keys, a smart phone or passport etc. Knowledge based include that a 'user must know' in order to log in, such as his user name and password or personal identification number (PIN). Biometrics based includes biometric user data to login. Biometric methods can be further classified based on the physiological and behavioral characteristics of a person. Physiological biometrics are those features that describe 'who the user is' depending on their physical attributes such as fingerprints, face, iris, retina scanning, hand geometry etc. Behavioral biometrics are related to the behavior pattern of a person, such as signature, voice, keystroke, mouse movement, etc.

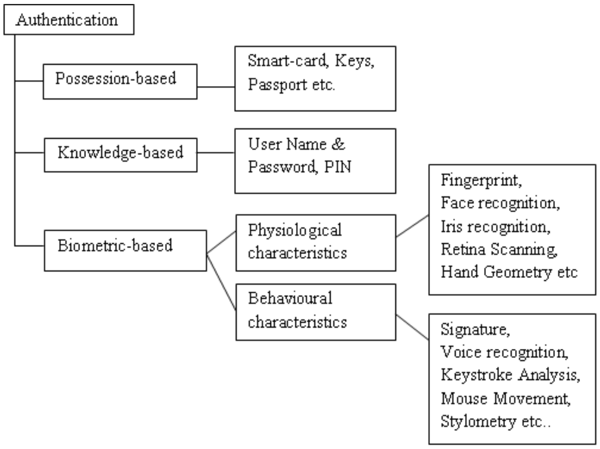

Authentication
Possession-based
Smart-card, Keys, Passport etc.
Knowledge-based
User Name & Password, PIN
Biometric-based
Physiological characteristics
Fingerprint, Face recognition, Iris recognition, Retina Scanning, Hand Geometry etc
Behavioural characteristics
Signature, Voice recognition, Keystroke Analysis, Mouse Movement, Stylometry etc..

Figure 1. Authentication Methods

Although there exist a variety of authentication techniques, Keystroke Dynamics is a powerful behavioral biometric system that involves authenticating a person based on his typing rhythm on the digital devices like smart phones , a desktop/laptop computers or usage of mouse to validate one's identity [2]. Keystroke dynamics is "not what you type, but how you type" [3]. In this approach, no external hardware is required to authenticate the user. Static keystroke authentication is performed on typing samples produced using predetermined text for all individuals under observation whereas the user is authenticated during an entire logon session in the continuous (free-text) keystroke authentication. The advantages of using keystroke biometric systems include low cost of implementation, unique feature for each individual, strengthens the security of the applications, continuous authentication system etc.

Authentication through keystroke dynamics is carried out in two main phases. First being the enrolment phase, and the second being verification phase [4]. Figure 2 describes the actions carried out during the KDA process. The first phase is the enrolment phase which acquires data from the user, such as username, password, user's typing behavior etc. The system extracts the keystroke timing features and a template is created for each user's typing behavior. This template is stored in the database as user's profile.

In the second phase, called the verification phase, whenever the user accesses the system, the system collects the user's keystroke timing features and compares with the template in the database using the classification model. If the data collected during the enrolment phase and the verification phase matches, the user is granted access to the system otherwise not. The components of biometric authentication include a reader or a gadget to read, software to verify the samples received, and a database to securely store the biometric information.

In the second phase, called the verification phase, whenever the user accesses the system, the system collects the user's keystroke timing features and compares with the template in the database using the classification model. If the data collected during the enrolment phase and the verification phase matches, the user is granted access to the system otherwise not. The components of biometric authentication include a reader or a gadget to read, software to verify the samples received, and a database to securely store the biometric information.

Figure 2 shows the different stages that are in KDA process. Data is collected from the individuals in different environments like in a specific lab environment (controlled) or from any place (uncontrolled). After this stage, the features like dwell time, flight time are extracted from the collected data (Figure 3). Dwell time is the time interval between a key press and key release. Flight time is the time interval between a key press of one key and the key press of another. These features are used by the classification model to verify if the keystroke pattern stored in the database and the keystroke extracted during the verification phase are the same or not. If both are same, then the user is a valid user otherwise he is an impostor.

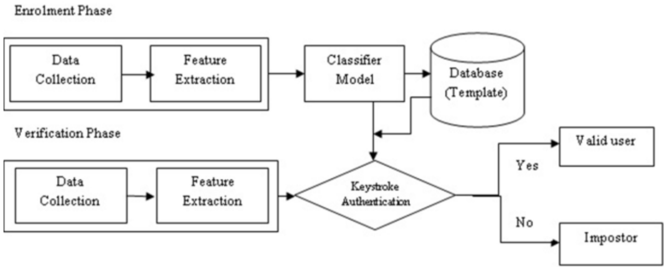


Figure 2. Keystroke Dynamics Authentication (KDA) process

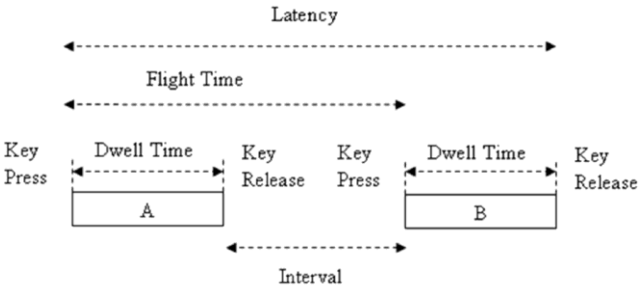


Figure 3. Features from Keystroke Dynamics

The paper is organized as follows, Section I contains the introduction of opinion mining, Section II contains the literature review, Section III contains summary of literature review and Section IV concludes review work with future directions.

II. Literature Review

Keystroke dynamics recognition has been used in the research area since 1980s and most of the works were done with fixed password text for authentication i.e on static authentication. Free-text systems do not restrict users to a particular fixed text. The users are given complete freedom in using any text of any length without any restriction. Only few researches focused on long free text i.e on continuous authentication. This section presents a review on the researches carried out on KDA systems focusing on fixed-text and on free-text.

In this paper, the authors [6] had considered typing behavior as a form of time-series and proposed an iterative real-time Keystroke Continuous Authentication (KCA). The DTW mechanism has been used in finding the similarities between the pairs of time series. The authors had compared the previous keystroke subsequence of the user with his current subsequence and identify the changes in the typing behavior. The authors had used 2 publicly available datasets VHSS and

GP dataset and 1 dataset generated by the authors namely ACB, for their verification process. The results had shown that time series-based approach to iterative KCA had much better performance than feature vector based approach. The authors have also suggested that this method can be used in applications like online assessment.

The authors [7] had proposed a continuous keystroke dynamics based user authentication based on their user-adaptive feature extraction method. Using this method the authors captured individual users' distinctive typing behavior and embedded in relative typing speeds for different digraphs. The authors had used two different languages, namely Korean and English to test 13000 keystrokes collected from 150 participants. The method was also tested with 5 different machine learning algorithms like Gaussian density elimination, Parzen window density elimination, 1-class SVM, K-NN and K-Means clustering. The results had shown that the Equal Error Rate (EER) of 0.44% and the authors had shown that the result has been improved compared to the benchmark method.

Researchers [8] had attempted to recognize the gender of the user using keystroke dynamics. Information Gain of each feature was calculated and the authors have used 5 different classifiers namely SVM, Naive-Bayes, Random Forest, Multilayer perceptron and Radial basis function network to evaluate the accuracy of the feature sets. The authors have claimed that they were able to obtain 95.6% accuracy in identifying the gender of the unknown user.

To improve the keystroke based authentication, the authors [9] had proposed using overt and linguistic context through which accuracy of user authentication can be increased. The author collected the typing data from 486 students in 2 sessions with a difference of 6 months using QWERTY keyboard. Scaled Manhattan Verifier has been used for testing the features and through the experiments, the authors had shown that they have achieved EER of 0.0232.

In this study, the authors [10] had proposed to build a robust and dynamic trust model algorithm which can be applied to any continuous authentication systems. The authors had taken two input devices namely keyboard and mouse for verifying the usage patterns of both devices. Average Number of Imposter Actions (ANIA) and Average Number of Genuine Actions (ANGA) were used as performance indicators and Genetic Algorithm was used to optimize the results.

The authors [11] had explored the usage of KDA using various input devices. The authors collected samples from 35 participants using 3 different keyboards like PC keyboard, soft keyboard and touch keyboard and their findings had shown that PC keyboard had achieved greater accuracy compared to others. In this paper [12], the authors had experimented the usage of different keyboards for feature extraction and identification. The authors have collected the samples using different text each time for 5 minutes from the participants. They have found that keyboard difference for users whose typing skills reached high level with about 900 or more letters in 5 minutes is not to be worried. The authors also claimed that the other users need to register their profile with the keyboard specific data.

Instead of taking fixed-text for keystroke based authentication, the authors [13] had taken free-text for analysis of keystrokes. The authors had combined monograph and digraph analysis for free text. The authors have also used neural network to predict missing digraphs between the keystrokes. The authors collected data in an uncontrolled environment from 53 individuals. The experiments were conducted in homogeneous and heterogeneous environments and they were able to achieve EER of 2.13% and 2.46% respectively.

In this paper [14], the performance of a continuous keystroke dynamic system was tested based on the function called penalty-reward method. The author collected data from 35 participants over a period of 6 days. EER has been used as the metric and the author had achieved the results between 41.8% to 48.0%. The most cited author [15] performed analysis on free text. Experiments were conducted by combining Absolute measure (A-measure) and Relative measure (R-measure) to calculate the distance between the typing samples and the results were promising. The authors collected 765 samples from 205 individuals over the duration of 6 months.

Researchers [16] had introduced the novel frequency feature extraction and the classifier which operates in frequency domain for the keystroke authentication system. The spectrograms are generated by short time Fourier Transform and include frequency and timing data. Frequencies which are above the threshold are used for training by the Gauss-Newton based Neural Network classifier and the same has been used to validate the attempts. 60 real attempts of the password owner and 60 fraud attacks from 12 different users were undertaken and the authors have obtained 4.1% Equal Error Rate (EER).

Clustering Based Keystroke Authentication Algorithm (CKAA) had been proposed by [17] and K-Nearest Neighbor approach has been used to classify users' keystroke dynamics profiles. From the experiments conducted, the authors had achieved FRR of 0% and FAR of 0.045%.

Profile of the participants had been collected using different categories like hand usage, gender, age etc by [18]. The authors have analyzed both static and free-text type of passwords and the accuracy rate of free-text has been found to be better than static type of passwords.

The authors [19] have used fixed text for authenticating the users using keystroke dynamics and they have achieved FAR of 4%. Mean distance has been used for classification. In [20], the authors have proposed statistical approaches like Hidden Markov Model and Gaussian model for web authentication. The authors have achieved EER of 2.54%.

Neural Network (NN) and Gaussian mixture model (GMM) have been used to classify the users in [21]. The accuracy was found to be increased when the length of the input string has increased. In this paper [22], feature subset selection in Keystroke Dynamics for identity verification has been used, and it reports the results of experimenting Ant Colony Optimization technique on keystroke duration, latency and digraph.

The authors [23] have proposed Two Component Information Set (TCIS) based on spatial and temporal features. New features have been generated like information value, Energy, Sigmoid, Multi-quadratic, Hanman Transform etc. These features are then tested using SVM, Random Forest and Convex Entropy Based Classifier. The experimental results were compared against CMU and SU datasets. The authors also tested these features on Android based Mobile Dataset and they have claimed that the results were found better than the earlier results.

Researchers [24] proposed a novel non- intrusive and privacy-aware biometric modality that utilized keystroke sound. The author used static text for authentication and collected data from 45 participants. The author had proposed a fusion of digraph statistics, histogram of digraphs, and intra-letter distances to authenticate a user. They have achieved EER of 25%.

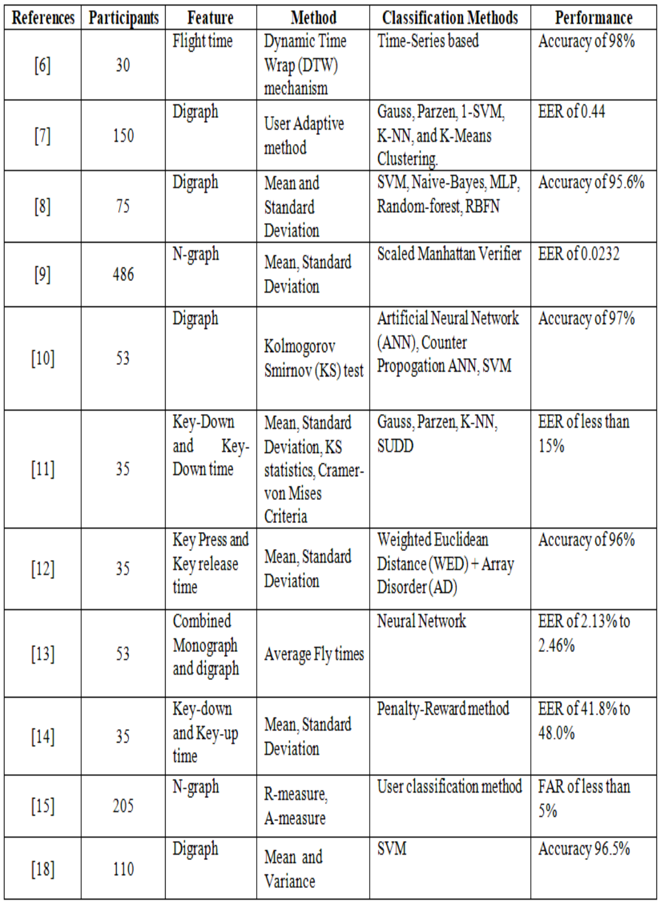
III. SUMMARY

From the literature, it has been understood that the factors listed below had influenced the result of the existing researches and have been discussed below.

- Datasets were generated using the keystrokes collected from the participants using either fixed-text (Static) or free-text (Continuous).
- Required number of keystroke samples had been collected from the participants which were recorded in a specific lab environment (Controlled) or from any place (Uncontrolled).
- Researches were carried out using traditional keyboard or laptops or mobile phones during the data enrolment and verification phases.
- Features like flight time and dwell time were most widely used for measuring keystroke patterns and monograph, digraph or n-graph of the keystrokes have been extracted from the template.
- Similarity measures have been computed using the statistical methods like Mean, Standard Deviation, and Euclidean Distance etc.
- Classification algorithms like Support Vector Machines (SVM), Neural Networks, Decision Tree, Random Forest, Time series based algorithm etc have been widely used by the researchers.
- To assess the performance of the classifiers, False Acceptance Rate (FAR), False Rejection Rate (FRR), Equal Error Rate (EER), Accuracy rate etc., have been used as the metrics.

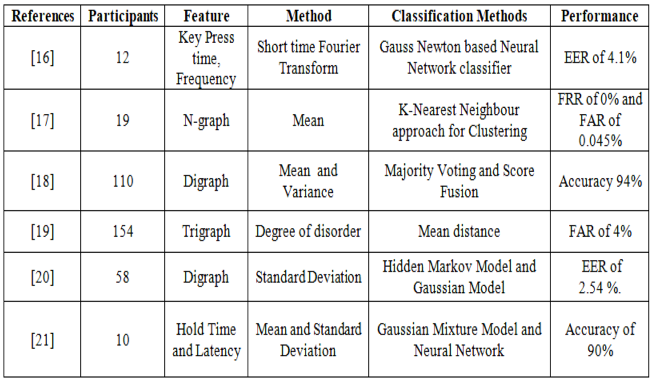
Table 1 and Table 2 summarize the data collected from different participants, features extracted, classification methods in various research articles using free-text and fixed text respectively. From these tables, it has been observed that the accuracy of the classifiers is higher when KDA is used with free-text than with fixed-text type.

Table 1. KDA using free-text



References	Participants	Feature	Method	Classification Methods	Performance
[6]	30	Flight time	Dynamic Time Wrap (DTW) mechanism	Time-Series based	Accuracy of 98%
[7]	150	Digraph	User Adaptive method	Gauss, Parzen, 1-SVM, K-NN, and K-Means Clustering.	EER of 0.44
[8]	75	Digraph	Mean and Standard Deviation	SVM, Naive-Bayes, MLP, Random-forest, RBFN	Accuracy of 95.6%
[9]	486	N-graph	Mean, Standard Deviation	Scaled Manhattan Verifier	EER of 0.0232
[10]	53	Digraph	Kolmogorov Smirnov (KS) test	Artificial Neural Network (ANN), Counter Propogation ANN, SVM	Accuracy of 97%
[11]	35	Key-Down and Key-Down time	Mean, Standard Deviation, KS statistics, Cramer-von Mises Criteria	Gauss, Parzen, K-NN, SUDD	EER of less than 15%
[12]	35	Key Press and Key release time	Mean, Standard Deviation	Weighted Euclidean Distance (WED) + Array Disorder (AD)	Accuracy of 96%
[13]	53	Combined Monograph and digraph	Average Fly times	Neural Network	EER of 2.13% to 2.46%
[14]	35	Key-down and Key-up time	Mean, Standard Deviation	Penalty-Reward method	EER of 41.8% to 48.0%
[15]	205	N-graph	R-measure, A-measure	User classification method	FAR of less than 5%
[18]	110	Digraph	Mean and Variance	SVM	Accuracy 96.5%

Table 2. KDA using fixed-text



References	Participants	Feature	Method	Classification Methods	Performance
[16]	12	Key Press time, Frequency	Short time Fourier Transform	Gauss Newton based Neural Network classifier	EER of 4.1%
[17]	19	N-graph	Mean	K-Nearest Neighbour approach for Clustering	FRR of 0% and FAR of 0.045%
[18]	110	Digraph	Mean and Variance	Majority Voting and Score Fusion	Accuracy 94%
[19]	154	Trigraph	Degree of disorder	Mean distance	FAR of 4%
[20]	58	Digraph	Standard Deviation	Hidden Markov Model and Gaussian Model	EER of 2.54 %.
[21]	10	Hold Time and Latency	Mean and Standard Deviation	Gaussian Mixture Model and Neural Network	Accuracy of 90%

IV. Conclusion

Security is an important issue in all applications which protects from malicious or accidental misuse of resources. The conventional user-id and password authentication is insufficient. To strengthen the security, keystroke dynamics has been integrated with user-id and password as multi factor authentication and protects the resources from impostors. This technique can be applied in web applications like online bank transactions, online exams, smart-phone authentication, detecting emotions of the user, intrusion detection etc. Earlier, the focus of the researchers on KDA was based on fixed text. Instead of collecting fixed text during enrolment and verification stages, free-text would take more number of keystrokes which will increase the accuracy rate. This work presented an investigation on the existing literature of KDA systems that focused on fixed-text and on free-text. Through the literatures, the accuracy rate of KDA systems when used with free-text, has been found to be considerably better when DTW mechanism and time series based approach has been used for authentication. It has also been observed from the literature that other statistical algorithms like Mean, standard deviation together with the classification techniques like KNN, ANN, SVM etc., have also been used to improve the results. The parameters extracted from the features can still be optimized to strengthen the security and integrity of the keystroke dynamics applications in future.

References

[1] M. L. Ali, J. V. Monaco, C. C. Tappert, and M. Qiu, "*Keystroke Biometric Systems for User Authentication,*" Journal of Signal Processing Systems, vol. 86, no. 2–3, pp. 175–190, 2017.

[2] A. Alsultan and K. Warwick, "*Keystroke Dynamics Authentication: A Survey of Free-text Methods,*" International Journal of Computer Science, vol. 10, no. 4, pp. 1–10, 2013.

[3] M. Karnan, M. Akila, and N. Krishnaraj, "*Biometric personal authentication using keystroke dynamics : A review,*" Applied Soft Computing, vol. 11, no. 2, pp. 1565–1573, 2011.

[4] A. Alshehri, F. Coenen, and D. Bollegala, "*Iterative Keystroke Continuous Authentication: A Time Series Based Approach,*" KI - Künstliche Intelligenz, pp. 1–13, 2018.

[5] J. Kim, H. Kim, and P. Kang, "*Keystroke dynamics-based user authentication using freely typed text based on user-adaptive feature extraction and novelty detection,*" Applied Soft Computing Journal, vol. 62, pp. 1077–1087, 2018.

[6] I. Tsimperidis, A. Arampatzis, and A. Karakos, "*Keystroke dynamics features for gender recognition,*" Digital Investigation, vol. 24, pp. 4–10, 2018.

[7] A. Goodkind, D. G. Brizan, and A. Rosenberg, "*Utilizing overt and latent linguistic structure to improve keystroke-based authentication,*" *Image and Vision Computing*, vol. 58, pp. 230–238, Feb. 2017.

[8] S. Mondal and P. Bours, "*A study on continuous authentication using a combination of keystroke and mouse biometrics,*" Neurocomputing, vol. 230, no. October 2016, pp. 1–22, 2017.

[9] P. Kang and S. Cho, "*Keystroke Dynamics-based User Authentication Using Long and Free Text Strings From Various Input Devices Keystroke Dynamics-based User Authentication Using Long and Free Text Strings From Various Input Devices,*" Journal of Information Sciences, vol. 308, pp. 72–93, 2015.

[10] Y. Matsubara, T. Samura, and H. Nishimura, "*Keyboard Dependency of Personal Identification Performance by Keystroke Dynamics in Free Text Typing,*" Journal of Information Security, no. July, pp. 229–240, 2015.

[11] A. A. Ahmed and I. Traore, "*Biometric recognition based on free-text keystroke dynamics,*" IEEE Transactions on Cybernetics, vol. 44, no. 4, pp. 458–472, 2014.

[12] P. Bours, "*Continuous keystroke dynamics: A different perspective towards biometric evaluation,*" Information Security Technical Report, vol. 17, no. 1–2, pp. 36–43, 2012.

[13] D. Gunetti and C. Picardi, "*Keystroke Analysis of Free Text," ACM Transactions on Information and System Security (TISSEC),* vol. 8, no. 3, pp. 312–347, 2005.

[14] O. Alpar, "*Frequency Spectrograms for Biometric Keystroke Authentication Using Neural Network Based Classifier,*" Knowledge-Based Systems, vol. 116, pp. 163–171, 2017.

[15] J. Hu, D. Gingrich, and A. Sentosa, "*Authentication through Biometric Keystroke Dynamics,*" IEEE International Conference on Communications, pp. 1556–1560, 2008.

[16] S. Z. Syed Idrus, E. Cherrier, C. Rosenberger, and P. Bours, "*Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords,*" Computers and Security, vol. 45, pp. 147–155, 2014.

[17] 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.

[18] C. H. Jiang, S. Shieh, and J. C. Liu, "*Keystroke statistical learning model for web authentication,*" ASIACCS '07: Proceedings of the 2nd ACM symposium on Information, computer and communications security, pp. 359–361, 2007.

[19] R. K. Das, S. Mukhopadhyay, and P. Bhattacharya, "User authentication based on keystroke dynamics," IETE Journal of Research, vol. 60, no. 3, pp. 229–239, 2014.

[20] M. Karnan, M. Akila, and A. Kalamani, "*Feature subset selection in keystroke dynamics using ant colony optimization,*" Journal of Engineering and Technology Research, vol. 1, no. 5, pp. 72–80, 2009.

[21] A. Bhatia and M. Hanmandlu, "*Keystroke Dynamics Based Authentication Using Information Sets,*" Journal of Modern Physics, vol. 8, no. 09, p. 1557, 2017.

[22] J. Roth, X. Liu, A. Ross, and D. Metaxas, "*Biometric authentication via keystroke sound,*" Proceedings - 2013 International Conference on Biometrics, ICB 2013.

[23] A. Bhatia and M. Hanmandlu, *"Keystroke Dynamics Based Authentication Using Information Sets,"* Journal of Modern Physics, vol. 8, no. 09, p. 1557, 2017.

[24] J. Roth, X. Liu, A. Ross, and D. Metaxas, *"Biometric authentication via keystroke sound,"* Proceedings - 2013 International Conference on Biometrics, ICB 2013, 2013.

Authors Profile

Mrs. M Rathi received her Master of Computer Applications (MCA) degree from Bharathiar University in 2003 and M.Phil Computer Science from Bharathiar University in 2005. She is currently pursuing part-time Ph.D in Hindusthan College of Arts and Science and working as Assistant Professor in Department of Computer Technology, Dr. NGP Arts and Science College, Coimbatore since 2016. She has 11+ years of teaching experience and has published papers in reputed international journals. Her research interests include Data Mining, User Authentication, Network Security.

Dr A V Senthil kumar is a Professor and Director of Department of Computer Applications, Hindusthan college of Arts and Science, Coimbatore. He obtained his Ph.D in Computer Science in 2009 . He has to his credit 9 Book Chapters, 165 papers in International Journals, 2 papers in National Journals, 30 papers in International Conferences, 5 papers in National Conferences, and edited 7 books published by IGI Global, USA. He is an Editor-in-Chief for International Journal titled, “International Journal of Data Mining and Emerging Technologies”, “International Journal of Image Processing and Applications”, “International Journal of Advances in Knowledge Engineering & Computer Science**”, “International Journal of Advances in Computers and Information Engineering” and “International Journal of Research and Reviews in Computer Science”**. He is a Key Member for India, Machine Intelligence Research Lab (MIR Labs).