

USER AUTHENTICATION USING ADVANCED KEYSTROKE ANALYSIS

Rosy Vinayak

PG Scholar, Computer Science & Engineering, Lovely Professional University

Phagwara, Punjab, (India)

ABSTRACT

Everybody needs to authenticate himself on his computer before using it, or even before using different applications like email. Most of the times, the adopted authentication procedure is the use of a classical couple of login and password. This method no longer provides consistent safety measures because passwords are prone to shoulder surfing and passwords can also be hacked. In this present study concept of keystroke dynamics is introduced to eradicate above said problems. This method is based on the assumption that every person types in a unique manner. Main advantage of introducing the concept of keystroke dynamics as compared to conventional system is that even the unauthenticated user cracks the password, he will be denied access as this method is based on the typing pattern of the user. In this research FAR error is 0% and FRR is also minimized and accuracy level has reached to a higher level.

Keywords: *Keystroke Dynamics, FAR, FRR.*

I. INTRODUCTION

Traditionally, authentication measures rested upon tools such as passwords and PINs. The main flaw with these measures is that we are not identifying the person, but the ability to have the information requested is evaluated. But transition is taking place in corporate sector, education sector, defense sector etc where biometric system has been introduced. Biometric authentication compares a live sample of the person with a template of previously recorded information. It identifies the person's recorded informational attributes so it doesn't depend on the person's previous knowledge. Biometric authentication is based upon what the user "is".

Two categories of Biometrics are: Physiological biometrics and Behavioral biometrics.

- **Physiological Biometrics:** It represents those traits that describe who we are based on physical attributes e.g. fingerprints, hand geometry, retinal and Iris scanning. For this additional equipment is required to be connected externally to the computer.
- **Behavioral Biometrics:** It is based on typing style, Voice Pattern and Signature recognition. Behavioral characteristics can be acquired without the need for external equipment although some attributes do require specialized hardware solutions.

II. KEYSTROKE DYNAMICS

2.1 Overview of Keystroke Dynamics

This method is based on the way a user types at a terminal and then evaluating the input given thousands of times per second, basically identifying habitual typing rhythm pattern. Keystroke dynamics features are usually extracted using the timing information of the key down/hold/up events. Two basic features used for keystroke dynamics are dwell time and Flight time.

- Dwell time is the time duration that a key is pressed
- Flight time is the time duration in between releasing a key and pressing the next key.

2.2 Advantages of Keystroke Dynamics

- The keystroke dynamics can be used by any person who knows how to use a keyboard
- Every individual type in a unique manner. Therefore typing pattern of two users cannot be same. Thus it provides more cyber security
- Compared to written signatures typing pattern cannot be reproduced. Most security systems allow limited number of incorrect attempts. After few incorrect attempts they block the account.
- Compared to physiological biometric systems such as fingerprint, Iris detection Keystroke dynamics does not require any extra hardware. Thus implementation and deployment cost is low.

2.3 Disadvantages of Keystroke Dynamics

- Sensitive to changes in keyboards and changes in typing languages.
- Affected by the user's physical condition (fatigue, illness and possible hand and fingers injuries).
- High False Non Match Rate (FRR).

2.4 Challenges in Keystroke Dynamics

- The Typing pattern of a person is inconsistent as compared to other biometric technologies. A person's hands can get sweaty or sometimes tired after typing for a long period of time. This leads to changes in typing pattern.
- Second Problem is that typing pattern depends on the type of keyboard being used.
- Another problem is that typing pattern depends on the person's posture whether he is sitting or standing and posture of person in sitting position.

III. REVIEW OF LITERATURE

Rudrapal et al. (D. Rudrapal, S. Das, and S. Debbarma, 2014) has proposed combination of different matrices and degree of disorder on keystroke latency as well as duration to generate user profile was calculated. Authentication process has been enhanced by statistical analysis on the proposed matrices. Proposed method was categorized into sub phases such as keystroke features were captured, degree of disorder was calculated, calculation of standard deviation and finally profile generation. For capturing keystroke data a registration form was created based on key pressed and release timing events. The author concluded that degree of disorder on keystroke duration is also different for different human. The result of proposed method showed FRR of 8% and FAR of 2%, which enhanced the existing authentication result using keystroke dynamics [1]

Hussain et al. (A. K. Hussain and M. M. Alnabhan, 2014) in his study presented an advanced keystroke authentication model improving users' validation strength. For each authorized user a keystroke structure had been defined, to be used in the user login attempts. The keystroke structure involved two components named firstly the user's typing time deviation thresholds, Secondly a unique user secret code which was distributed between password's characters based on time distances. Proposed method depends heavily on the amount of information distributed among typing time, and on reducing the deviation of these times. This system solved the problem of large deviations in keystroke dynamics and improved keystroke authentication level was provided. A strong authentication level had been achieved and participating users accepted this system model. [2]

Senathipathi et al. (K. Senathipathi, Krishnan Batri, 2014): A comparative analysis of Particle Swarm Optimization and Genetic algorithm has been shown by the author with respect to keystroke dynamics. Dwell time, Flight time, Digraph, Bigraph and Virtual Key Force are seven features which has been used by the author. Genetic algorithm based wrapper approach is proposed for application in keystroke dynamics based authentication as stated by the author. The author used one class SVM as base classifier and four diversity through the uniqueness of each chromosome rendering post processing unnecessary. According to PSO method emotional states are employed as a biometric along with the keystroke dynamics pitching mainly on the emotions undergone by the user while entering the text on the keyboard. An improved authentication of the user in comparison to GA method has been seen as a result of PSO method. [3]

The author Maheswari et al. (T. Maheswari and S. Anitha, 2014) has introduced a novel approach for authentication that was based on biometric characteristics i.e. keystrokes of the password entry. The author has considered three phases namely, fingerprint, login credential based on username and password and keystroke dynamics. Two stages were also considered that are Training stage and testing stage. Training stage was implemented during enrolment and testing during verification period. [4]

N. Chourasia (NANDINI CHOURASIA, 2014) has introduced an additional layer of security for the authentication of the user, Keystroke Dynamics. The security can be implemented in android phones or any other smart phones through which internet is accessible as well as online transactions can be performed. Main objective was to collect a keystroke dynamics data set to measure the performance of a range of detector sand to develop a repeatable evaluation procedure so that the results can be compared more accurately. A mathematical model was presented before implementation [5]

Ahmed et al. (A. A. Ahmed and I. Traore, 2014) presented a new approach for the free text analysis of keystrokes that combined monograph and digraph analysis. A neural network had been used to predict missing digraphs based on the relation between the monitored keystrokes. The heterogeneous experiment involved 53 users, the follow-up experiment in a homogeneous environment considered only 17 volunteers. The results obtained from this research were promising with reduced error rates. [6]

Monaco et al. (J. V. Monaco, N. Bakelman, S.-H. Cha, and C. C. Tappert, 2014) evaluated and developed a new classification algorithm with reduced error rate. The description was given about the recent developments and evaluation of a keystroke biometric system for continual computer-user authentication on short-burst input duration which was on the order of minutes. The Java applet that used the PC windows event clock which recorded the key press and release times in a millisecond format captured the keystroke data. The average and standard deviation of key-press duration times and of digraph transition times has been used for feature extraction. The vector-difference authentication model which transforms a multi-class problem into a two-class

problem has been used for the classification procedure. The performance of a biometric system and the representation of trade-off between the False-Accept Rate and False reject rate are characterised by Receiver Operating Characteristics curves. [7]

Singh et al. (K. Singh, Harjeet Kaur, 2013) has presented the study of keystroke dynamics to identify individuals based on their typing rhythm behavior with the help of Fuzzy Rule Based system. Inter stroke gap existing between consecutive characters of the user identification is being used in this approach. Several factors are there which are analyzed for the identification of the impostor and legitimate user under this system. With the introduction of the rule based system the number of the attempts of the users can be restricted resulting in better biometric as there are fixed number of attempts. A major part in the field of keystroke biometrics is played by the fuzzy rule to increase the accuracy. The identification of the imposter gets easier. The introduction of neighbor key pattern eases the identification of the imposter that made the system more reliable. But various challenges being presented by the author need to be overcome to make it more effective biometric. Thus it has an enormous potential to grow in the field of cyber security. As far as future work is concerned it includes developing Fuzzy rule based system more accurate in performance so that no legitimate user can be considered as imposters. [8]

Bajaj et al. (S. Bajaj and S. Kaur, 2013) emphasized on the importance of keystroke dynamics for user authentication. The typing rhythm of a user stored in the database was compared with the login input for authentication. The author described keystroke Dynamics as a two factor biometric security. Firstly, for a successful login password should be known and secondly, typing rhythm should match. The method involved calculation of key the pressing time, dwell time and total time of password. Laptop keyboard was used for the analysis of the results. [9]

The author Kaur et al. (M. Kaur and R. S. Virk, 2013) has used perceptron function including Feed forward propagation learning algorithm to train user typing pattern through keystroke dynamics and then testing or cross validation was applied. This approach provided more security through the use of neural network. No extra hardware was required in this study like other biometric systems. [10]

Rybnik et al. (M. Rybnik, M. Tabedzki, M. Adamski, and K. Saeed, 2013) : The main aim of this paper was efficient user authentication with keystroke dynamics using non-fixed text of various sizes. The approach had been tested on a small group of individuals, and data was gathered over Internet using browser-based WWW application and on local machines using dedicated applications. Nine individuals were participated from whom keystrokes samples were collected which corresponds to the use conditions of a computer system in a home or small business. Each individual typed a long text twice in the five sessions of more than 250 characters, ten samples for each person was collected in this way. The author stated that keystroke dynamics proved to be a promising and effective biometrics feature for authentication of individuals with analysis of only two keystroke features and with the use of simple classifier. [11]

Hassan et al. (S. I. Hassan, M. M. Selim, and H. Hala, 2013) implemented a robust keystroke dynamics system with the aim to solve the problem of samples variations and an adaptive threshold is considered in this study. The proposed system was evaluated using CMU dataset and for this study a new dataset have also been created. Four Distance based algorithms named Manhattan, Manhattan with standard deviation, Euclidean and Mahalanobis were implemented. Manhattan with standard deviation produced the best results because standard deviation of the training samples was also considered in that method. [12]

Schclar et al. (A. Schclar, L. Rokach, A. Abramson, and Y. Elovici, 2012) has provided novelty in the field of authentication of users for login. It was based on two approaches. The first approach is called Cluster representative which used a unique user as a representative from each cluster. The second approach called Inner Cluster representative which selects that user as a representative whose biometric profiles were the most similar to that of examined user. [13]

IV. METHODOLOGY

Work is divided into two phases:

A. Enrolment/Registration phase

B. Verification/Authentication phase

The following algorithms are used to create new users and for authentication of users by matching data with the previously stored data of users.

New User Creation (Fig 1)

Step 1 Administrator will create a new account of the user in which user will type his own name and a given string in presence of the admin of the system.

Step 2 Users have to enter a given phrase of text once, in order to measure the typing timing period.

Step 3 Relevant Features such as Key Hold Time, Inter Key time, No. of times Shift key used , No. of times CAPS LOCK key used and No. of Backspaces will be extracted.

Step 4 Typing pattern of the given text will be stored in the file.

User Authentication (Fig 2)

Step 1: If user has entered user id then he has to enter password and for wrong password he has given three chances and if again no correct password is entered then form disappears saying that you are not allowed, otherwise first string is enabled which has to be entered by user and this procedure go on up to five strings.

Step 2: For all these five strings keystroke features such as key hold time, inter key time, key type change time as well as number of times user have pressed SHIFT, BACKSPACE and CAPS will be calculated.

Step 3: After that their average values are calculated which were then compared with each key hold, inter key, key type change time of all trusted users list and their Euclidean distance is calculated and minimum Euclidean distance value is given as maximum token and that user is taken as suspected user which is then compared for number of SHIFT or CAPS used and if they match then number of BACKSPACE is compared if it matches then only he is granted access otherwise not.

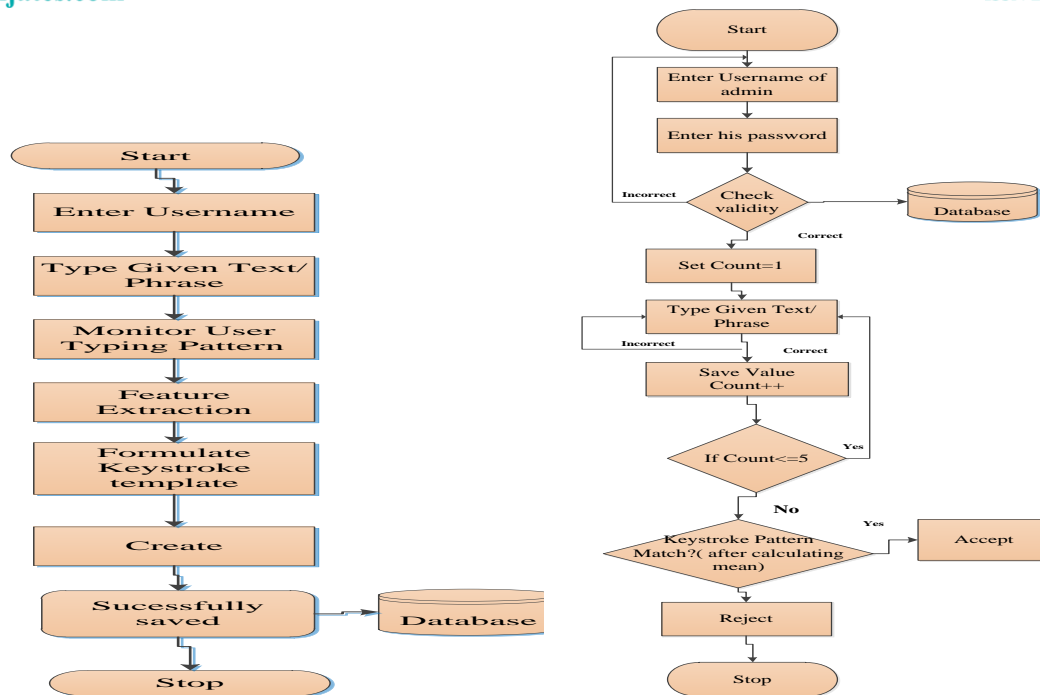


Figure 1: Flowchart for New User Creation Figure 2: Flowchart for user authentication

V. RESULTS AND DISCUSSION

The proposed methodology is implemented with the help of software called JAVA.

5.1 Parameters Used for Experimentation

In this research various keystroke features for user authentication are proposed. Euclidean distance metric has been applied between data collected from genuine users and incoming user's typing characteristics and find most suitable match using token based approach.

I have used following features:

1. **Key Hold Time:** Key hold time, refers to the time elapsed between pressing and releasing a single key.
2. **Inter Key Time:** It refers to the amount of time between pressing and releasing two successive keys.
3. **Key Type Change Time:** It is basically the Inter Key Time. Two types of key time change time are:
 - In which first key is alphabet and second is other than alphabet (numeric)
 - In which first is other than alphabet (numeric) and second key is alphabet.
4. **No. of SHIFT:** It specifies how many times SHIFT key is pressed while typing the pattern.
5. **No. of BACKSPACE:** It specifies how many times BACKSPACE key is used while typing the pattern.
6. **No. of CAPS:** It indicates number of times CAPS Lock key is used while typing the pattern.

5.2 Data Collection

The next step is to collect the data for all trusted users. For this purpose a form has been prepared which collect user details (Fig. 3). The form contains two entries to be filled by user; one is his user name and second consists of alphanumeric string. While user enters that string keystroke features such as Inter key time, Key Hold time and Key type change time, number of times SHIFT key is used and no. of times CAPSLOCK key used and



Figure 3: Registration Form

5.3 Experiment and Results

To check the validity of user an experiment has been performed. For this a form has been prepared that consists of multiple entries (Fig. 4). First one is for user name entry and then for password entry. If user has entered user id then he has to enter password and for wrong password he has given three chances and if again no correct password is entered then form disappeared saying that you are not allowed.

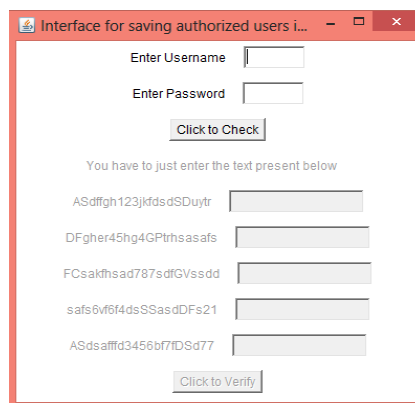


Figure 4: Verification Form

If the username and password is correct then first string is enabled which has to be entered by user and this procedure go on up to five strings. For all these five strings their key hold time, inter key time, key type change time as well as number of times user have pressed SHIFT, BACK and CAPS has been calculated. After that their average values are calculated which were then compared with each key hold, inter key, key type change time of all trusted users list and their Euclidean distance is calculated and minimum Euclidean distance value is given as maximum token and that user is taken as suspected user which is then compared for number of SHIFT or CAPS used and if they match then number of BACKSPACE is compared and some constraints are applied if all these conditions matched then only he is granted access otherwise not.

Below Fig. 5 is a screenshot of the message that indicates successful login of the user.

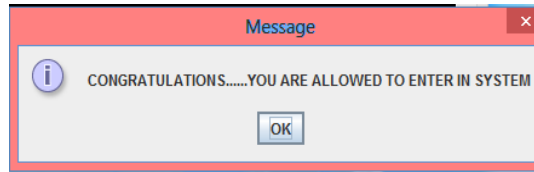


Figure 5: Screenshot on Successful Login of the User

If the user has not been registered, although he enters correct username and password he will not be allowed to enter the system because his typing pattern has not been matched with the pattern of registered or trusted users. Below Figure 6 indicates the error message that appears when impostor tries to enter into the system.

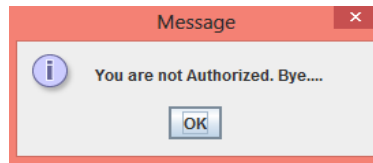


Figure 6: Screenshot Showing Unsuccessful Login of User

5.4 Results and Analysis

Previous researches have shown only a limited range of accuracy. In this present study a great accuracy level has been achieved because in this Type 1 error has been reached to 0% and Type 2 error is also minimized. More the number of attributes the more accuracy level can be achieved. Type 1 and Type 2 are explained below:

- Type 1 error or False Acceptance Rate (FAR): This means an Impostor is wrongly accepted by the system. This happens when biometric system treats two different persons to be the same person.
- Type 2 error or False Rejection Rate (FRR): This means a legitimate user is wrongly rejected by the system. This happens when a biometric system treats two different samples from the same person to be from a different person.

Table 1 Shows the Accuracy Level and Measure of Type 1 and Type 2 Error of Present Study.

Attributes	Accuracy in %	Type 2 Error in %(FRR)	Type 1 Error in % (FAR)
Key Hold+ Inter Key Time	20	45	35
Key Hold Time + Inter Key Time + Key Type Change Time	35	40	25
Key Hold Time + Inter Key Time + Key Type Change Time Alphabet to Numeric) + no. of SHIFT or CAPS used	55	35	10
Key Hold Time + Inter Key Time + Key Type Change Time + no. of SHIFT or CAPS used + no. of BACKSPACE used	70	30	0
Key Hold Time + Inter Key Time + Key Type Change Time (Numeric to Alphabet) + no. of SHIFT or CAPS used+ no. of BACKSPACE used	94	6	0

5.5 Comparisons and Improvement

Comparison of present study with the Base paper [1] is shown below in table 2.

Table 2: Comparison with the Base Paper

PARAMETER	BASE	PROPOSED
No. of Keystroke Features	2	7
Keystroke Features	Key hold Time, Inter Key Time	Key Hold time, Inter Key time, Key Type change time, Key type change Numeric type No. of Shifts used, No. of Caps used & No. of Backspace used.
FRR	8 %	6
FAR	2 %	0 %
PASSWORD COMPRISES	Only Lowercase letters	Lowercase letters ,Uppercase letters and Numeric characters.

VI. CONCLUSION

Keystroke dynamics has shown positive results in process of user authentication. It is an effective method to strengthen existing password based approach. Keystroke Dynamics is based on individuals typing pattern. It is cost effective approach as well as provide reliability. This approach is invisible to users as they only need to type on a keyboard in normal manner instead of providing their physical attributes such as fingerprint or retina scanning. Keystroke patterns are captured with no knowledge to the user. The results from this study indicate that behavioral based biometric technique i.e. Keystroke Dynamics provide a level of security with great accuracy and lower error rates. The distance metric used to compute the distance D between the template and test input is Euclidean distance algorithm. A threshold has been set. After applying various values for threshold it has been deduced that a value of 5000 ns for each parameter performs the best. In this research Type 1 error is 0% and Type 2 error is also minimized and accuracy level has reached to a higher level.

Future work can include assessing the feasibility of using keystroke dynamics on small devices such as palm tops, tablets and other touch screen devices because technology is now not limited to desktops and laptop it has been upgraded to mobile devices like cell phones. But the old QWERTY keyboard as input device has not been changed for many years. The application of keystroke dynamics needs to be applied on various touch screen devices. Majority of research on keystroke dynamics involves English language; it can be applied on other languages as well as it can lead to different results due to layout of keyboards for different languages.

Keystroke dynamics has potential to grow in the area of cyber security since it is cost effective and non intrusive biometric.

REFERENCES

- [1] D. Rudrapal, S. Das, and S. Debbarma (2014), "Improvisation of Biometrics Authentication", 10th International Conference, ICDCIT 2014, Bhubaneswar, India, pp. 287–292
- [2] A. K. Hussain and M. M. Alnabhan (2014), "Advanced Authentication Scheme Using a Predefined Keystroke Structure", Int. J. Comput. Sci. Inf. Technol., vol. 6, no. 2, pp. 163–169.

- [3] K. Senathipathi (2014), “An Analysis of Particle Swarm Optimization and Genetic Algorithm with Respect to Keystroke Dynamics”, Green Computing Communication and Electrical Engineering (ICGCCEE), Coimbatore, pp. 1 – 11.
- [4] T. Maheswari and S. Anitha (2014), “User Authentication Based on Biometrics for Convincing User in keystroke dynamics”, ISR NATIONAL Journal of Advanced Research in Computer Science Engineering and Information Technology, Volume: 1 Issue: 1 08-May-2014, pp. 63–70.
- [5] N. Chourasia (2014), “Authentication of the user by keystroke dynamics for banking transaction system”, Proceedings of International Conference on Advances in Engineering & Technology, 20th April-2014, pp. 41–45
- [6] A. Ahmed and I. Traore (2014), “Biometric Recognition Based on Free-Text Keystroke Dynamics”, IEEE TRANSACTIONS ON CYBERNETICS, vol. 44, no. 4, pp. 458–472.
- [7] J. V. Monaco, N. Bakelman, S.-H. Cha, and C. C. Tappert (2013), “Recent Advances in the Development of a Long-Text-Input Keystroke Biometric Authentication System for Arbitrary Text Input”, Eur. Intell. Secur. Informatics Conf., pp. 60–66.
- [8] K. Singh (2013), “Rule Based Approach for Keystroke Biometrics to identify authenticated user”, (IJCSIS) International Journal of Computer Science and Information Security, vol. 11, no. 7, pp. 6–14.
- [9] S. Bajaj and S. Kaur (2013), “Typing Speed Analysis of Human for Password Protection (Based On Keystrokes Dynamics)”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-3, Issue-2, July 2013, pp. 88–91.
- [10] M. Kaur and R. S. Virk (2013), “Security System Based on User Authentication Using Keystroke Dynamics”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 5, May 2013, pp. 2111–2117.
- [11] M. Rybnik, M. Tabedzki, M. Adamski, and K. Saeed (2013), “An Exploration of Keystroke Dynamics Authentication Using Non-fixed Text of Various Length,” 2013 Int. Conf. Biometrics Kansei Eng., pp. 245–250, Jul. 2013.
- [12] S. I. Hassan, M. M. Selim, and H. Hala (2013), “User Authentication with Adaptive Keystroke Dynamics”, IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 4, No 2, pp. 127–134.
- [13] Schclar, L. Rokach, A. Abramson, and Y. Elovici (2012), “User Authentication Based on Representative Users”, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, vol. 42, no. 6, pp. 1669–1678.