International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org Volume 3 Issue 9 ǁ September 2014 ǁ PP.41-48

Enhanced Knowledge-Based User Authentication Technique Via Keystroke Dynamics

[1,]Soumen Roy , [2,]Utpal Roy , [3,]D. D. Sinha
[#1,3]*Department of Computer Science and Engineering*
University of Calcutta, 92 APC Road, Calcutta -700 009, INDIA.
[*2]*Department of Computer & System Sciences,*
Visva-Bharati, santiniketan -731235, INDIA

ABSTRACT: *Some common words (name, address, E-mail ID, password, …), we press daily and we are habituated to press it in same rhythm, which is unique and can be used to segregate and distinguish people. In this paper we are considering password as well as habitual rhythm of the entered password to enhance the security level in knowledge-based user authentication. Recognizing typing style promises a parameter like biometric characteristics that may facilitate non-intrusive, cost-effective and continuous monitoring. But this technique, as is now, suffers from accuracy level and performance. In order to realize this technique in practice a higher level of security and performance together with low cost version is needed with an error to an accepted level. Hence, it is highly essential to identify all the controlling parameters and optimize the accuracy, performance and cost with new algorithms. Automated Password recovery mechanism also can be implemented by this technique. In this paper we also suggest some future plans that can be effectively implemented.*

KEYWORDS: *Behavioral biometric, Computer Security, Euclidean Distance, Keystroke Dynamics, Mahanobolis Distance, Manhattan Distance, Z Score.*

I. INTRODUCTION

Today, password and PIN are not limited in knowledge-based user authentication. Keystroke dynamics characteristics have been meshed up with password and PIN. The use of only password or PIN in Knowledge-based user authentication techniques is risky because of not only the possibility of off-line guessing attacks. The password is also risky as pressing the password in public place,it is unsafe while all around the areas; class room, office, bank, college campus, railway station are covered by video cameras or spy cameras or pressing password slowly can be traceable by friends or people those are very near to us. It is also unsafe, if we pick up a word from a relatively small dictionary for password,which may content our personal information. An attacker may collect our personal information and can check one by one until the actual result is obtained. So there is a probability of brute force attack or shoulder surfing attack. Among all password recovery mechanisms, "secret questions answers" and "password hints" are very popular. But, if the attacker knows our personal information, then he may change our password or can access the system. Another password recovery mechanism in OTP (One Time Password), here one extra account or mobile phone is required to get the verification code. In order to realize the technique in practice, a higher level of security with strong password recovery mechanism and performance together with low cost version is demanded with an error to an accepted level as may be designed.

Keystroke dynamics is a behavioral biometrics which is the method of analyzing the way a user types on a keyboard and classify him based on his regular typing rhythm. It is the study of people who can be identified by their typing rhythms; much like handwriting is used to recognize the author of a written text. User's typing pattern is unique because of the neuro-physiological factors that also make written signatures unique.Keystroke Dynamics as biometrics characteristics is not a new one. Keystroke Dynamics was first formally investigated by Bryan and Harter in 1897 as part of a study on skill gaining in telegraph operators. In 1975 Spillane suggested in an IBM technical bulletin that typing rhythms might be used for identifying the user at a computer keyboard. That bulletin described keystroke dynamics in concept. Forsen et al. in 1977 conducted preliminary tests of whether keystroke dynamics could be used to distinguish typists [1]. Gaines et al. in 1980 produced an extensive report of their investigation with seven typists into keystroke dynamics [2]. After then S. Bleha submitted his PhD thesis on Recognition system based on keystroke dynamics in 1988 [3]. R. Joyce and G. Gupta proposed an identity authentication based on keystroke latencies in 1990 [4]. F. Monrose et al. [5] proposed keystroke dynamic as a biometric for authentication in 2000. Different online and offline applications already have been done by fixed text and free text keystroke dynamics. Keystroke dynamics research has been going on for the more than thirty three years.

Many methods have been proposed during that time. Methods based on traditional statistics-such as mean times and their standard deviations-are common. Over the years, different pattern recognition methods have come into vogue and been applied to keystroke dynamics; neural networks, Fuzzy logic and support vector machines among others. They often used two features Dwell time and Flight time as biometrics features, Dwell time which refers to the amount of time between pressing and releasing a single key and Flight Time which refers to the amount of time between pressing and releasing two successive keys.A laboratory made sample password database has been used to train the system. Here system records all the key press and release timing and calculates the duration of depressed characters, latency time between various down and up key sequence latencies for each sample, then finds out the actual timing template by applying some statistical methods. Then some features mining mechanism or distance based algorithm such as Euclidean distance, Manhattan distance, Manhattan distance with standard deviation, Mahalanobis distance, Bhattacharyya Distance as per Janakiraman, R. &Sim, T. [6], or Genetic algorithm, particle swan optimization which is explained by Marcus, K. and Akila, M. [7] may be used to decide whether the user is valid. Thus we can minimize the probability of any off-line guessing attacks since rhythm of password is used, which cannot be copied even after watching it several times. The rhythm of the password as it is entered is used to validate the authenticity of the user rather than only password. It updates itself continuously by Growing Window, Moving Window or Adaptive threshold mechanism, defined by Pin, S. T.and Giot R.et.al [8, 14], which can help to recover the account and minimize Equal Error Rate (EER) in future. Here keyboard is enough to recognize our gait; no extra security apparatus is needed. It enhances the security level and can also be used to identify an individual. This technique is promising as biometric characteristics recognition which cannot be lost or stolen in addition with inexpensive, continuous monitoring, non-intrusiveness, convenient and can be easily implemented into the existing computer security system with minimal alternation and user intervention.

II. SCIENCE OF KEYSTROKE DYNAMICS

Keystroke Dynamics is a technology to segregate and distinguish people based on their typing rhythms, which is the method of analyzing the way a user types on a keyboard and classify him based on his regular typing rhythm. It is the study of whether people can be well-known by their typing rhythms, much like handwriting is used to recognize the author of a written text. A user's typing pattern may be unique because similar neuro-physiological factors that make written signatures unique. Here users are not only identified by their corresponding userID and password or PIN, but their typing style is also accounted for. In our experiment, we used fixed-text comprising of characters instead of password and habitual typing pattern of entered fixed texts. Here, user can choose any text as password from his/her own dictionary. It is very simple and nothing difficult to remember; still it enhances the security level and can be used to identify an individual. Our typing style can be easily calculated by simple program which can calculate key pressing and releasing time of each key which is defined by the table I and then generates key-hold time and sequence of down and up keys latency times. All the timing parameters are calculated in millisecond (ms).

TABLE I
(RAW DATA SAMPLE OF KEY PRESS & RELEASE TIME FOR FIXED-TEXT "kolkata123" OF A USER)

Entered Keys	Key press time	Key release time
k	1408858545370	1408858545479
o	1408858545542	1408858545651
l	1408858545745	1408858545854
k	1408858545979	1408858546103
a	1408858546119	1408858546259
t	1408858546696	1408858546821
a	1408858546852	1408858546993
1	1408858547320	1408858547429
2	1408858547539	1408858547648
3	1408858547757	1408858547866

TABLE II
(SAMPLE KEY HOLD AND KEY LATENCY TIMES WITH STANDARD DEVIATION FOR THE RECORDED FIXED-TEXT "kolkata123")

Key	Key hold time	Key hold (Sd)	Down Down key latency	Down Down key latency (Sd)	Up Up key latency	Up Up key latency (Sd)	Up Down key latency	Up Down key latency (Sd)	Down Up key latency	Down Up key latency (Sd)
k	78	16.79881	187	9.80816	213	29.003448	291	16.14311	109	16.982344
o	104	12.743626	205	6.1481705	213	19.052559	317	8.01249	101	13.046072
l	111	6.5726705	267	36.460938	291	30.518847	403	36.10263	156	31.403822
k	135	21.1849	140	19.77372	116	27.386127	252	20.918892	5	8.01249
a	111	23.17326	231	28.81662	231	30.35457	343	26.176325	119	33.86739
t	111	15.349267	127	18.42824	119	12.601587	231	15.735311	15	13.885244
a	104	7.745967	377	18.308468	351	21.624062	455	23.099783	273	21.475567
1	78	17.527122	174	11.924764	208	16.204937	286	8.01249	96	15.433729
2	112	6.387488	205	11.882761	197	16.211107	309	12.049896	93	9.81835
3	104	12.743626	-	-	-	-	-	-	-	-

After key events timing calculation we have calculate all Up Down sequence timing features by the following equations:

[1] Key Hold Time: time between key pressed and key released for a single key
[2] Key Hold Time= Release timing of a key – Press timing of same key. (1)
[3] Down Down Key Latency: time between two consecutive presses.
[4] Down Down Key Latency = Press timing of a key – Press timing of previous key. (2)
[5] Up Up Key Latency: time between two consecutive releases.
[6] Up Up Key Latency = Release timing of a key – Release time of previous key. (3)
[7] Up Down Key Latency: time between the current key release and the next key press.
[8] Up Down Key Latency = Release timing of a key – Press timing of a previous key. (4)
[9] Down Up Key Latency: time between the current key press and the next key release.
[10] Down Up Key Latency = Press timing of a key – Release timing of a previous key. (5)

III. PROPOSED MODEL

Our systems works in a four-stage process that consists of the following steps:

Capture: System collects the chosen password as well as press and release time of each key events of the person who wants to enroll. Some features are also can be captured such as key pressure, sequenced combined keys timing (di-graph, tri-graph), typing speed, finger movement style on keyboard, method of error correcting, sequence of special action keys(left right Alt, Shift, Ctrl) etc.

Extraction: System calculates all five timing features such as key duration, down down, up up, down up and up down key latencies by the equations 1 to 5 and generates a template.

Comparison: Stored password and timing template is then compared with a claim sample by some distance measurement algorithm and calculates the scores by the equations 6 to 10.
Match/non-match: The system then identifies the minimum score and corresponding user's name and decides whether the features extracted from the claim sample is match or non-match.

A. Password generation model: In our system, one or more samples are captured at the time of enrolment. Then by applying some statistical method password template will be generated. Our system takes password and typing signature then generates a new encrypted timing template along with password. So no one can predict the password string and our biometric properties; keystroke dynamics.

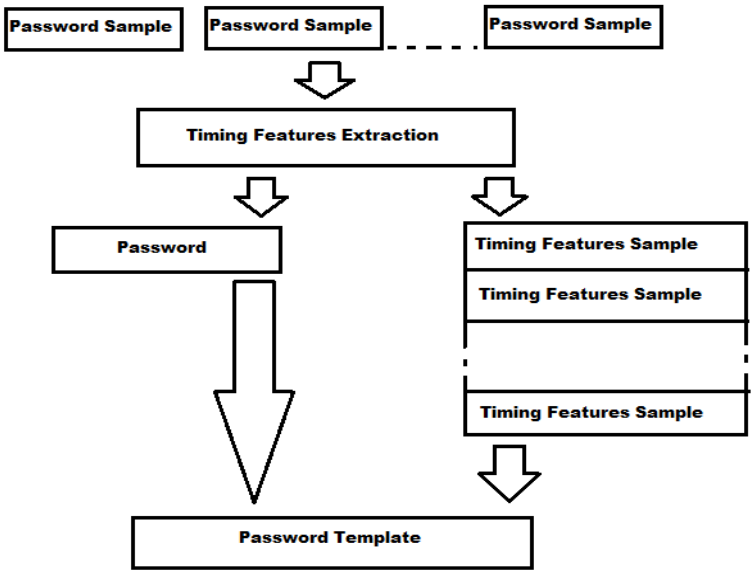

Fig. 1: Password template generation model

B. Password verification Model: Stored password templateare compared with a claim sample for that person then the system decides whether the features extracted from the claim sample are match or non-match with the password template.

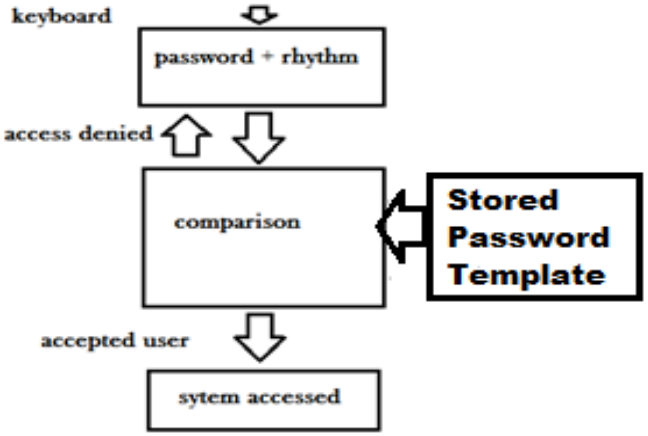

Fig. 2: Password verification model

KEYSTROKE DYNAMICS ALGORITHMS : Many classification methods have been applied in keystroke dynamics study over the last three decades, where statistical methods, features learning methods and neural network are popular. But in our experiment, following are the distance based algorithms were used to evaluate the system. Down Up key latency of two consecutive keys may be negative for key overlapping. So we have to take absolute values of the samples.

A. **Manhattan Distance:**

The score is calculated in Equation 1 which represents Manhattan distance:

$$M=\sum_{i=1}^{n}(|x_i - y_i|) \qquad (6)$$

Where $x=(x_1, x_2, x_3, ...,x_n)$ represents test vector and $y=(y_1, y_2, y_3, ..., y_n)$ represents the mean vector of the training sample.

B. *Manhattan with Standard Deviation Distance:*

The standard deviation of each feature is calculated as in Equation 2. Here αi represents standard deviation.

$$M_s=\sum_{i=1}^{n}(|x_i - y_i|)/\alpha_i \quad (7)$$

C. *Euclidean Distance:*

The score is calculated as the squared Euclidean distance between the test vector and mean vector as in Equation 3.

$$E = \sqrt[2]{\sum_i^n(|x_i - y_i} \quad (8)$$

D. *Mahanabolis Distance:*

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

$$Eh = \sqrt[2]{\sum_i^n((|x_i - y_i|)/\alpha_i)^2} \quad (9)$$

E. *Z Score Values:*

The score is calculated in Equation 5 which represents Z score in Equation 5:

$$Z= \quad \sum_{i=1}^{n}(|x_i| - \mu(|x_i|))/ \quad (10)$$

Where $\mu(x_i)$ are mean value and α_i is standard deviation.

IV. EXPERIMENTAL SETUP

We have implemented a program in java for experimental purpose, which has the capability of capturing all key pressing events and to create the database. It also can calculate different score or distance between vectors. Fifteen users are invited to press three most common passwords six times each using same keyboard. According to SplashData, who gathered data from millions of stolen passwords posted online, the top three passwords in the year 2013 are "123456," "password" and "12345678". So we can say most of the people are uninspired while choosing a healthy password because we, as people are still very lazy. It increases the probability of guessing attacks. For three different reasons, we have chosen three different passwords for experimental purpose. First password which is the combination of only digits (i.e., "123456"), second password is the combination of alphabets (i.e., "password") and third one is combination of both (i.e., "kolkata123").

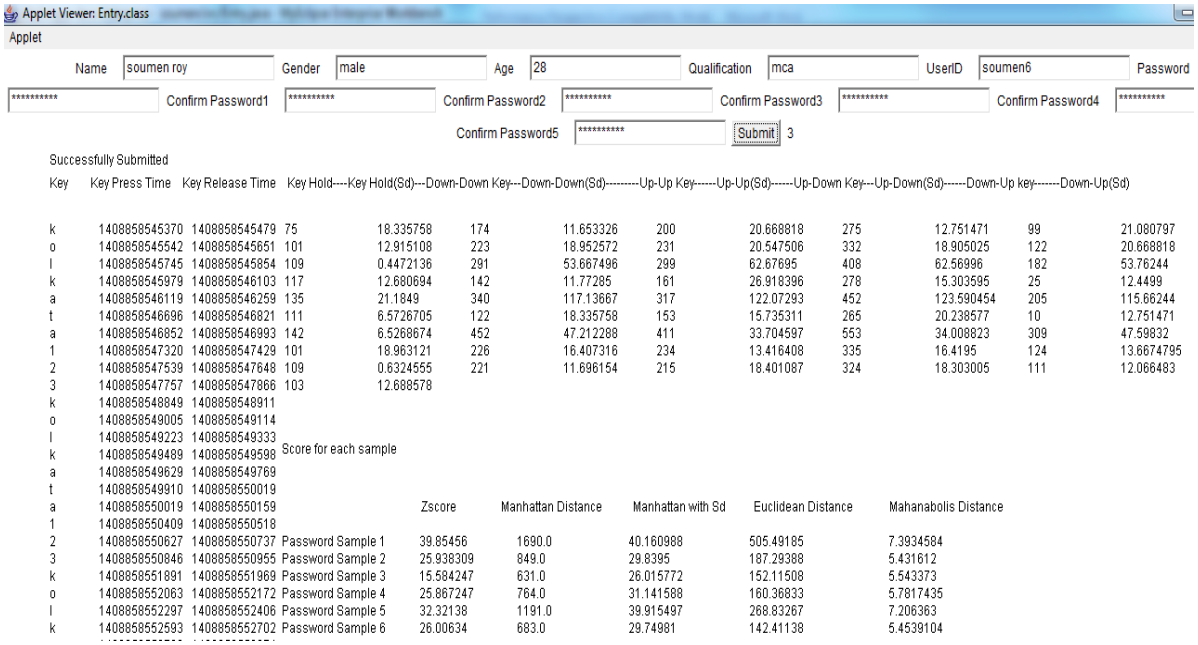

Fig.3: Experimental setup

EVALUATION AND ANALYSIS : There are a few performance measurement parameters that can be used to evaluate performance of different biometric system.
False Acceptance Rate (FAR): FAR is defined as the percentage ratio between falsely accepted illegal users against the total number of imposters accessing the system.
False Rejection Rate (FRR): FRR refers to the percentage ratio between falsely denied genuine users against the total number of genuine users accessing the system.Equal Error Rate (EER): ERR is the rate at which both false acceptance and false rejection error are equal. In our simulation program, we have recoded each key pressing and releasing time for six sample of passwords (size of password <=10) and calculated key hold time, down-down key latency, up-up key latency, up-down key latency, down-up key latency and their mean and standard deviation (Sd) which is shown in the table II. After then we have calculate Z score, Manhattan distance, Manhattan with standard deviation, Mahanabolis distance and Euclidean distance for each samples with calculated mean, which are very much similar. Calculated core is defined in the following bar chart, where score calculation by different algorithms with the fixed-text "kolkata123" for 6 same samples of a user are represented and we can see Manhattan distances may vary than z score.

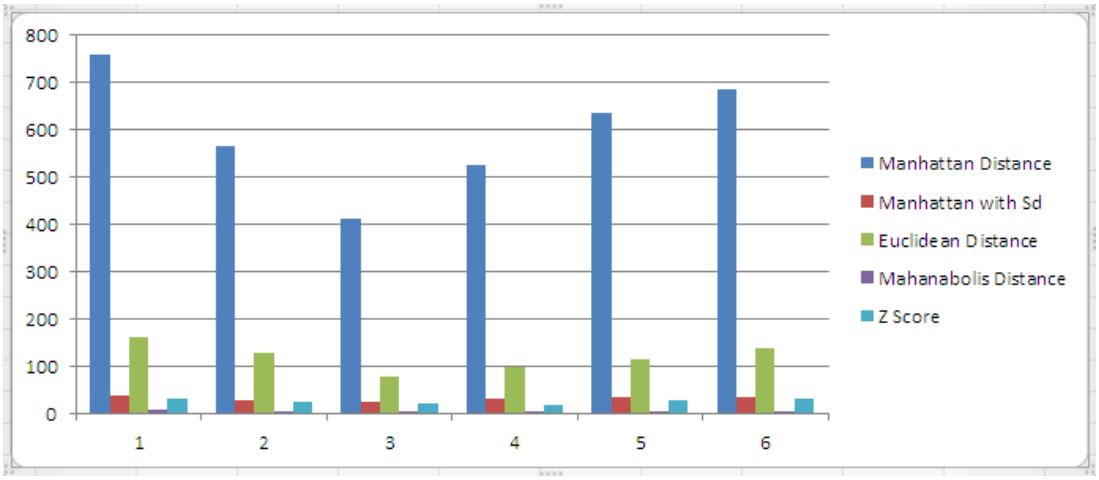

Fig.4: Score calculation by different algorithms with the fixed-text "kolkata123" for 6 same samples of a user

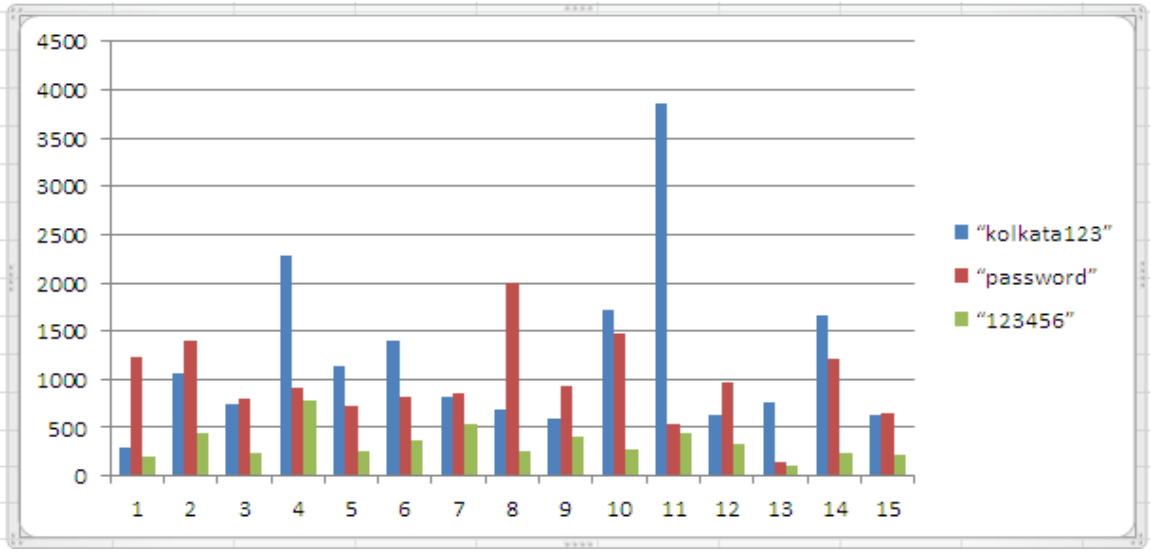

Fig.5: Euclidean distance of different set of data for the password "kolkata123", "password" and "123456"

In the avobe figure we can see "kolkata123" given the good result than others password. Here we can conclude that string length is a factors which significantly varies for different user's typing style. In our experiment, we got 0.133 EER for the password "kolkata123" where 0.4 and 0.53 EER for the password "password" and "123456" respectively. So best result obtained is fixed-text "kolkata123" because we have collected data from Kolkata. People of Kolkata are habituated to press "kolkata" and we got the excellent result.

In our experiment we collected rhythm of 270 fixed-texts from 15 users and seen that all the user's typing style are different as we see in the following line chart. Here all samples used same password string but Sample 1 and sample 2 are probably same where Sample 3 and Sample 4 are same for the two different users. Here user, who pressed password for sample 1 & sample 2 is not much habituated with the press password for that reason line chart is varied in rhythm. Good suggestion is choose the password what we press daily such as userID, name, place etc. Otherwise three factors will affect the system, finger movement time, key searching time and different keyboard.

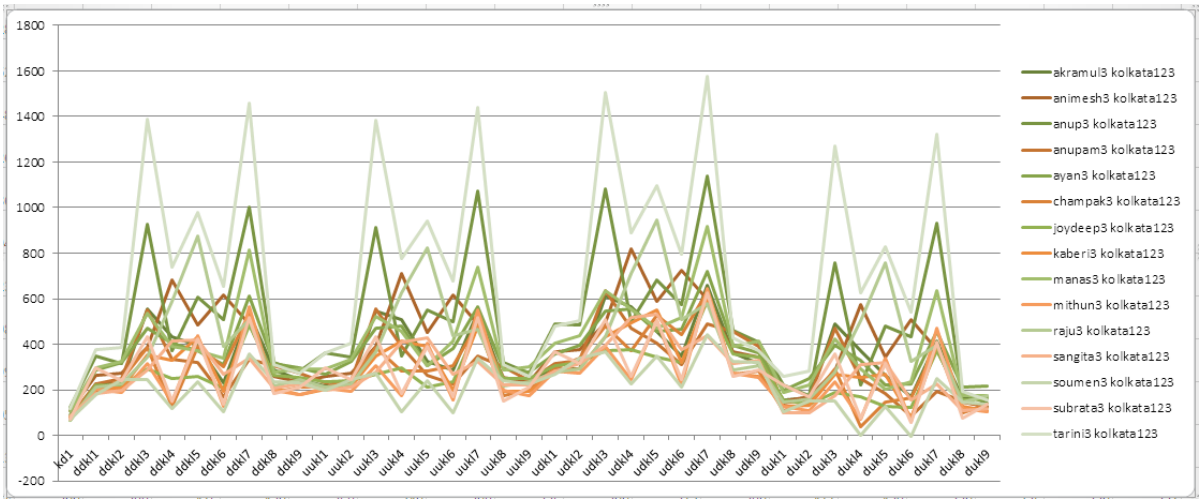

Fig.6: Line chart of 15 same sample of passwords "kolkata123" for 15 different users where all key latencies are considered

V. DISCUSSION AND FURTHER AREA OF RESEARCH

Keyboard is essential for a computer device, which can recognize our typing style and very much unique as per our experiment and cannot be copied or stolen. It can be used as safe guard of our password in any access control system. This technique can be used in online criminal investigation as the information are getting stored in the cloud or in a server which will help the investigators to identify the particular human behavior with which they can recognize the individual, back door account identification, where user having different accounts can also be recognized, online typing examination, emotion recognition, lab attendant system and many more.
Different types of keyboard such as keypad, desktop keyboard or standard keyboard, and screen touch keypad may affect the way of keystroke dynamics. Basically keypad is not changing frequently in mobile phone. So this technique can be effective for mobile security as per Trojahn, M. and Ortmeier, F. [17] otherwise, artificial keystroke dynamics or keystroke sound implemented by Roth, J. et al. and Metaxas D. [18] would be introduced.Characteristics of human may change over time. So update mechanism is needed to update template periodically after acceptation of verification or identification. Sometimes, score of different algorithms varies. It would be better if we combined all scores in a single equation like mean value calculation with given weights of all scores.

VI. CONCLUSION

This technique is easier to implement in any computer system to make the system secure and convenient to use. We can conclude by saying that do not hesitate to press password in front of your friend or public place where all areas are covered by video cameras or spy cameras to recognize the finger movements. Here, the user does not have to remember anything extraordinary or difficult combination of alphabets and numbers and no extra security apparatus is needed to recognize the user's typing style. This technique can be effectively applied in application areas such as student or employee attendance system, distance based examination, password recovery mechanism, emotion recognition, private data encryption, continuous user verification, criminal investigation, identifying backdoor accounts, free-text user authentication etc. But it has some drawbacks; person's typing speed may vary subsequently during a day or between two days depending on change in psychological or physical state of the person. Using retraining module this problem can be solved. The results from this study and others indicate that behaviour based biometrics generally and keystroke dynamics specifically provide a level of security and it can be applied in any system. This paper suggests that if we introduce the rhythm based analysis to passwords it enhances the level of the security system but if we also implement the same to the userID section, then it will create a much stronger string thus taking the security system to the next level. In any particular language, if we can identify the pause in syllables then with that we

can remove the limitations of fixed-texts where we do not have to type the password again to recover it, rather we can type a sentence or a paragraph to match the original rhythm stored in the database to identify the user.

VII. ACKNOWLEDGEMENTS

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

REFERENCES

[1] Forsen G., Nelson M., and Staron R., Jr. "Personal attributes authentication techniques", *Technical Report RADC-TR-77-333*, Rome Air Development Center, October 1977.
[2] Gaines R., Lisowski W., Press S., Shapiro N., "Authentication by keystroke timing: some preliminary results", *Rand Rep. R-2560-NSF*, Rand Corporation, 1980.
[3] Bleha S., Slivinsky C. and Hussien B., "Computer-access security systems using keystroke dynamics", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (1990) 1217–1222
[4] Joyce R., Gupta G., "Identity authorization based on keystroke latencies"', *Communication of ACM 33* (2) (1990) 168–176.
[5] Monrose F., Rubin A. D., "Keystroke dynamics as a biometric for authentication", *Future Generation Computer Systems*, Vol. 16, No. 4, pp. 351–359, 2000.
[6] R. Janakiraman and T. Sim. Keystroke dynamics in a general setting. *In Advances in Biometrics (ICB 2007)*, volume 4642 of Lecture Notes in Computer Science, pages 584–593, August 27–29, 2007, Seoul, Korea, 2007. Springer-Verlag, Berlin.
[7] Marcus K. and Akila M., "Personal Authentication based on Keystroke Dynamics using Soft Computing Techniques", *Second International Conference on Communication Software and Networks, IEEE*, 2010
[8] Pin S. T., "A Survey of Keystroke Dynamics Biometrics", *The Scientific World Journal*, Vol-2013, Article ID 408280.
[9] Killourhy K. S., "A Scientific Understanding of Keystroke Dynamics", PhD thesis, Computer Science Department, Carnegie Mellon University, Pittsburgh, US, 2012.
[10] Roy S., Roy U., Sinha D.D., "Rhythmic Password-based Cryptosystem", *2nd International Conf. on Computing and System*, University of Burdwan, West Bengal, India, 2013, 303-307.
[11] Roy S., Roy U., Sinha D.D., "Modified Knowledge-based User Authentication Technique", *7th International Conf. on Mathematical Science for Advancement of Science and Technology, MSAST, IMBIC*, Kolkata, India, Vol: 2, 2013, 236.
[12] Roy S., Roy U., Sinha D.D., "Combined User Authentication Technique", *International Conf. on Recent Trends in Science & Technology (ICRTST), College of Engineering and Management*, Kolaghat, West Bengal, India, 2013, 106-113.
[13] Shimaa I. H., Mazen M. S. and Hala H., "User Authentication with Adaptive Keystroke Dynamics", *IJCSI*, Vol. 10, Issue 4, July 2013.
[14] Giot R., Dorizzi B., Rosenberger C., "Analysis of template update strategies for keystroke dynamics," *Computational Intelligence in Biometrics and Identity Management (CIBIM)*, 2011 IEEE Workshop on , vol., no., pp.21,28, 11-15 April 2011 doi: 10.1109/CIBIM.2011.5949216
[15] Giot R., El-Abed M., and Rosenberger C., "Keystroke dynamics authentication for collaborative systems," in Proceedings of the International Symposium on Collaborative Technologies and Systems (CTS '09), pp. 172–179, May 2009.
[16] Killourhy K. and Maxion R., "Why did my detector do that?!: predicting keystroke-dynamics error rates," in *Proceedings of the 13th International Conference on Recent Advances in Intrusion Detection*, pp. 256–276, Ottawa, Canada, 2010.
[17] Trojahn M., Ortmeier F., "Toward Mobile Authentication with Keystroke Dynamics on Mobile Phones and Tablets," *Advanced Information Networking and Applications Workshops (WAINA)*, 2013 27th International Conference on , vol., no., pp.697,702, 25-28 March 2013.
[18] Roth J., Xiaoming Liu, Ross A., Metaxas D., "Biometric authentication via keystroke sound," *Biometrics (ICB), 2013 International Conference* on , vol., no., pp.1,8, 4-7.