KEYSTROKE ANALYSIS FOR USER AUTHENTICATION

[1]Beenish Shabir, [2]Sheikh Riyaz Ul Haq

[1]Research Scholar, [2]Assistant Professor
[1]Department of Information Technology,
[1]Central University of Kashmir, Tullamulla Ganderbal, Jammu and Kashmir, India

Abstract :

As we are in the twenty-first century, new challenges proliferate to safeguard against fraud and impersonations. Today we are much dependent on computers to sensitive and delicate information. Dependence on computers to store and process information has made it obligatory to secure them from unauthentic users. Biometric Security is a security mechanism used to authenticate and provide access to a facility or system based on instant verification of an individual's physical characteristics. It is the strongest and the most foolproof measure used for identifying an individual. One such behavioral biometric technique is keystroke dynamics or keystroke analysis which can make use of typing pattern of an individual can be used to enhance existing security techniques like password protecting techniques efficiently and effectively .This work is premeditated to present one such cover based on validating access to computers by recognizing certain unique and habitual typing patterns of users..

Keywords-Keystroke Dynamics, Keystroke Analysis, Biometrics, PCA (Principal Component Analysis), SVM(Support Vector Machine).

I. Introduction

Computer Systems are being used in almost every point of our life. As a result, the security menacing to computer systems have also augmented notably. The conventional password protected authentication systems are unsafe to use as passwords can be stolen or forgotten. Keystroke dynamics or keystroke analysis on the other hand is considered a substitute solution due to its non-intrusiveness. Keystroke Analysis is economical and convenient than other biometric techniques like retina/iris or fingerprints. Biometrics is the most reliable, robust, user-friendly authentication tool. Biometrics measure is individual's physical or behavioural characteristics or features to validate their identity. Common Biometrics include: fingerprints, voice, gait, Iris, face, keystroke patterns. Keystroke Analysis is also known as Typing Dynamics, Typing Biometrics etc. To achieve high level of security conventional password based authentication systems can be conjugated with keystroke analysis. An imposter trying to access a computer system of any application can be easily detected with high accuracy using the proposed system. We can secure or safeguard our systems by using this technology. Keystroke analysis is hardware independent .Only a standard keyboard is required for authenticating a specific user.

Table 1:Overview of different schemes used for user authentication.

APPROACH	ADVANTAGE	DISADVANTAGE	EXAMPLE
Knowledge	Effortless High acceptance	Forgotten Spoofed	Password PIN
Token	Simple Cheap	Lost Theft	Smartcards
Biometrics	Unforgettable Unique	Cost Trespassed	Retina Voice Keystrokes

II. CLASSIFICATION OF KEYSTROKE ANALYSIS

Keystroke dynamics is defined in [1] as “ the process of analyzing the way users type by monitoring keyboard inputs and identifying them based on patterns in their typing rhythm ”.It is based on the assumption that “Typing patterns are unique to each typist”-According to NIST(National Institute of Standards and Technology) and NSF (National Science Foundation). Keystroke dynamics is a type of behavioral biometric measure that focuses to validate or identify the human beings based on the analysis of their typing patterns or the way they type on a keyboard as every human has a certain way of typing that separates him from the others.[2] Keystroke analysis is the detailed timing information which reports exactly when each key was pushed down and when it was released as a person types on a computer keyboard. Keystroke biometrics analyzes the way in which an individual types on a keyboard. The keystrokes of an individual are measured to develop a template that is stored in a database for future authentication. Data needed to analyze keystrokes is obtained by keystrokes is obtained by keystroke logging.

The following section summarizes the keystroke verification techniques that can be used:

- **Static Verification-**Static keystroke verification authenticates or validates a typing pattern based on a known keyword, phrase or some pre-defined text. The keystrokes are captured and stored as a template in a database during enrolment phase and then compared against the keystrokes that the user types on a keyboard. Static keystroke verification authenticates a particular user on the basis of what they typed and how they typed at the initial login time i.e., authentication is done only at the login time.
- **Dynamic Verification -** Dynamic keystroke verification authenticates an individual on the basis of their typing during a logged in session. The captured details of the particular user are compared to the users profile to determine deviations.

The Dynamic keystroke verification provides some advantages over the static verification method in a way that it is able to authenticate users on the basis of any input i.e., it is not dependable on certain text or certain input. Static verification system studies the keystroke characteristics at a specific time while as the continuous verification or the dynamic verification on the other hand studies the typing behavior of an individual throughout the interaction or session time. Timing feature information can be drawn out from keystrokes captured and stored as a template in the database in numerous ways, such as studying or examining the dwell time or the time for which the key was held down(key duration),keystroke latency, frequency of error words while typing, typing rate etc. The extracted timing information is then processed through a unique learning algorithm and then the information is used for future comparison in both identification and authentication tasks. The data needed to analyze the users is obtained by keystroke logging or keystroke recorder.

III. RELATED WORK

Mariusz Rybnik et al., (2009) [3] proposed an approach, aimed at better security with keystroke dynamics using short fixed text. In this paper they present experiment on one word phrase. This can stand for a simple user password and extracted the pressing time and duration between two keys. They calculate new term overlapping between two keys (one key is pressed and last key is not released).

N. Harun et al. (2010) [4] addresses the issue of enhancing systems security using keystroke biometrics as a translucent level of user authentication. The paper focuses on using the time interval between keystrokes as a feature of individuals’ typing patterns to recognize authentic users and reject imposters.

Dr.Shaimaa Hameed et al. (2014)[8] Combined the timing features to be used to verify the user using Multi-Layer Perceptron Neural Network (MLP NN) classifier with dimension reduction using Principal Component Analysis (PCA). Results showed enhancement in system accuracy due to PCA reduction. Not only the False Rejection Rate (FRR) and False Acceptance Rate (FAR) are dropped to 24% and 6% respectively, but also the neural network Mean Square Error (MSE) and training time are decreased with PCA deployment.

L. K. Maisuria et al.[6] Keystroke was classified based on an MLP approach and K-means cluster algorithm. Both the MLP and K-means gave an 84% and 85% acceptance
rate and a 69% and 85% impostor rejection rate.

Shimshom et al (2010)[5] The proposed method was evaluated on 21 legitimate users and 165 attackers. The results were encouraging and suggest that the detection performance of the proposed method is better than that of existing methods. Our method is divided into training and verification phases. In the training phase we build a verification model that consists of a multiclass classifier and a mapping function, for the user u based on all users' sessions.

S. Bleha et al[9] Two types of passwords were considered: phrases and individual names. A fixed phrase was used in the identification system. Individual names were used as a password in the verification system and in the overall recognition system. All three systems were tested and evaluated. The identification system used 10 volunteers and gave an indecision error of 1.2%. The verification system used 26 volunteers and gave an error of 8.1% in rejecting valid users and an error of 2.8% in accepting invalid users. The overall recognition system used 32 volunteers and gave an error of 3.1% in rejecting valid users and an error of 0.5% in accepting invalid users.

Pavaday et al [10] compares applications of neural networks to the field of hardened password mechanism in a typical workplace environment.

R-Giot et al[11] proposed a new method based on the Support Vector Machine (SVM) learning satisfying industrial conditions (i.e., few samples per user are needed during the enrollment phase to create its template). In this method, users are authenticated through the keystroke dynamics of a shared secret (chosen by the system administrator). They used the GREYC keystroke database that is composed of a large number of users (100) for validation purposes. They compared the proposed method with six methods from the literature (selected based on their ability to work with few enrollment samples). Experimental results show that, even though the computation time to build the template can be longer with our method (54 seconds against 3 seconds for most of the others), its performance out performs the other methods in an industrial context (Equal Error Rate of 15.28% against 16.79% and 17.02% for the two best methods of the state-of-the-art, on their dataset and five samples to create the template, with a better computation time than the second best method).

Table 2:Comparison between research works with Static Authentication mode

STUDY	DATA SIZE	LATENCY	INPUT REPITITION	INPUT FREEDOM	METHOD	FAR (%)	FRR (%)	EER (%)	INPUT SESSIONS	DEVICE FREEDOM
[9]	26	Flight Time	30	YES	Bayesian and min Distance Classifier	2.8	8.1	-	Once	NO
[10]	100	Dwell Time Flight Time	-	NO	Multilayer Perceptron	1	8	-	Once	-
[11]	100	Dwell Time Flight Time	12	NO	Support Vector Machine	-	-	15.28	YES	NO
[12]	51	Dwell Time Flight Time	50	NO	Manhattan Distance	-	-	7.1	YES	NO
[13]	10	Dwell Time Flight Time	20	YES	Inductive learning	9	10	-	-	NO

IV. METHODOLOGY TO BE ADOPTED

The proposed methodology is illustrated in the figure 1 where the users will be asked to register in the system or simply get enrolled in the system. Each user will be asked to enter the username and password. The user will be asked to type password x=20 times to build the database which will be needed to verify him each time he tries to have access in future. The timing features will be then extracted from the password transparently while the user is typing it. After 20 times of entries, username and password are appended to the feature vectors and saved in the database or repository. PCA (Principle Component Analysis) is performed on the features to reduce them before they are served to the learning algorithm. PCA reduces high dimensionality to low dimensionality in order to overcome the problem of over-fitting. The outputs from the PCA will be served as inputs to the SVM (Support Vector Machine) Algorithm to be trained. SVM is a supervised machine learning algorithm which can be used for both classification and regression. However, it is primarily used for classification problems. SVM is best known for its high dimensionality input space. It looks at the data and sorts it into one of the two categories.

The built model including the SVM and the feature templates will be ready to be used to classify the users When the user claims to grant access to the system . He will be asked to type his username and password for verification, the same features will be extracted from the password after checking the validity of the user and then comparing them to the enrolled users' templates. These typing features are then fed to the trained SVM to classify the claimed users as authentic or fake.

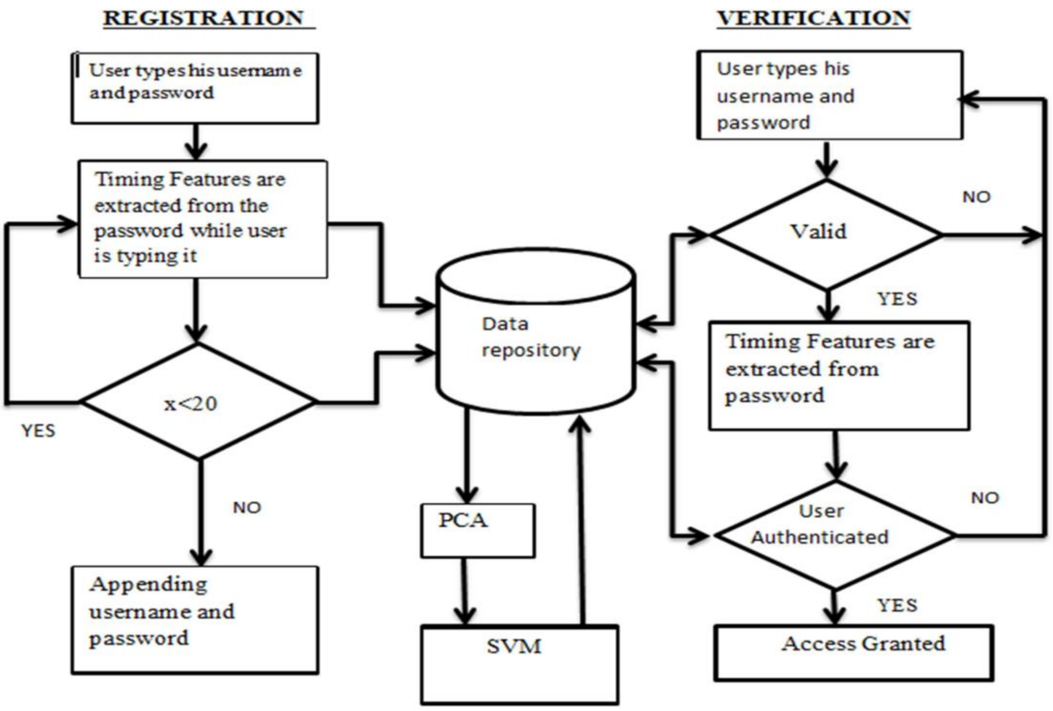
REGISTRATION
VERIFICATION
User types his username and password
Timing Features are extracted from the password while user is typing it
x<20
YES
NO
Appending username and password
Data repository
PCA
SVM
User types his username and password
Valid
NO
YES
Timing Features are extracted from password
User Authenticated
NO
YES
Access Granted

Figure: 1

V. CONCLUSION

Keystroke Analysis has a strong psychological basis which should be explored to gain a good knowledge of the users behavior during typing. Using the gained knowledge, models could be built to better understand the process involved in typing. In the proposed model, three time features will be extracted and combined to be used for user verification. These features are: Key Duration or Dwell Time, Flight Time and Total Typing (Total time to type the whole string). Using Keystroke Analysis Technology we can secure our passwords. The framework may provide emphasis on security that is growing in demand in web-Based applications over internet.

REFERENCES:

[1]. Patrick Elftmann,"Secure Alternatives to Password-based Authentication Mechanisms", Diploma thesis, Laboratory for Dependable Distributed Systems, RWTH Aachen University, Aachen, Germany, October 2006.

[2]. Wikipedia [Online] Available : http://en.wikipedia.org/wiki/Keystroke_dynamics.

[3]. Mariusz Rybnik, Piotr Panasiuk & Khalid Saeed "User authentication with Keystroke dynamics using Fixed Text" University of Bialystok Marii Sklodowskiej Curie 14, 15 097,978-07695-3692- 7/09 $25.00 © 2009 IEEE.

[3]. N. Harun, W. L. Woo and S.S. Dlay, "Performance of Keystroke Biometrics Authentication System Using Artificial Neural Network (ANN) and Distance Classifie rMethod",International Conference on Computer and Communication Engineering, 11-13 May 2010, Kuala Lumpur, Malaysia,2010 IEEE.

[5]. TomerShimshon, Robert Moskovitch, LiorRokach, Yuval Elovici, "Continuous Verification Using Keystroke Dynamics", International Conference on Computational Intelligence and Security, 2010 IEEE.

[6]. L. K. Maisuria , C. S. Ong and W. K. Lai, " A comparison of artificial neural network and cluster analysis for typing biometrics authentication", International Joint Conference on Neural Network, IJCNN'99, vol.5, pp 3295-3299, (1999).

[7]. F. Monrose, a .D . R. "Keystroke Dynamics as a Biometric for Authentication. "Future Generation Computing Systems (FGCS), 12(12), pp. 351-359, (2000).

[8]. Dr. Shaimaa Hameed shaker, Dr. Riyadh Jabbar Saydani, Mina Khidhir Obaid ,"Keystroke Dynamic Authentication based on Principle component analysis and neural networks", International Journal of Scientific and Engineering Research, Vol 5,Issue 6,(June-2014).

[9] S. Bleha, C. Slivinsky, and B. Hussien, "Computer-access security systems using keystroke dynamics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 12, pp. 1217–1222, 1990.

[10] N. Pavaday and K. M. S. Soyjaudah, "Investigating performance of neural networks in authentication using keystroke dynamics," in *Proceedings of the IEEE AFRICON 2007 Conference*, pp. 1–8, September 2007.

[11] R. Giot, M. El-Abed, B. Hemery, and C. Rosenberger, "Unconstrained keystroke dynamics authentication with shared secret," *Computers and Security*, vol. 30, no. 6-7, pp. 427–445, 2011.

[12] K. Killourhy and R. Maxion, "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.

[13] J. A. Robinson, V. M. Liang, J. A. M. Chambers, and C. L. MacKenzie, "Computer user verification using login string keystroke dynamics," *IEEE Transactions on Systems, Man, andCybernetics A*, vol. 28, no. 2, pp. 236–241, 1998.

[14] R. Joyce and G. Gupta, “Identity authentication based on keystroke latencies,” *Communications of the ACM*, vol. 33, no. 2, pp. 168–176, 1990.

.