User Authentication Through Keystroke Dynamics Using KNN

Amol Borgaonkar, Ankita Salunke, Nikita Gupta, Vikrant Sharma

Student, Department of Computer Engineering, Sinhgad Institute of Technology, Lonavala, SPPU, Pune, Maharashtra, India

***Abstract* — In most cases, where security is the concern, many computer systems use the common username/password scheme for user verification and identification. But this line of security can be broken easily using methods like dictionary attack and brute force attack. Some advanced biometric tools like fingerprint or retina scan are used for the same purpose but the cost of implementation is high. In this paper we propose a novel approach for user verification and identification using keystroke dynamics as a biometric tool. KB authentication system secures the authentication process by verifying the claimant's typing pattern. This paper presents a KB authentication system which uses KNN as the clustering algorithm, which is a non-parametric algorithm for classification. KB Authentication System does not require any additional hardware for its implementation and provides an extra layer of security, thus it is less costlier and secure than the traditional authentication methods.**

Keywords— Security, Keystroke Biometric, clustering Algorithm , KNN.

I. INTRODUCTION

Biometric technologies are mainly used for verification and identification of a person which takes his / her physical and behavioral characteristics into account. Some Authentication systems uses biometric tools like retina or fingerprint scans while some don't.

Authentication by password is the most common and usual approach to assure the security and access control of a computational system. But, on the other hand, this is the weakest and unreliable approach to secure an authentication system where security is really a big concern, like banking systems or online examination systems. If someone knows the username-password of a authentication system, he/she can easily get into it without any complications. Moreover, these authentication systems are vulnerable to attacks like brute force, phishing or dictionary attack. Using biometric systems like fingerprint or retina scan with a typical username / password based authentication system clearly makes the system more secure, but the cost of implementation of these systems are high.

Biometrics can be broadly classified into two types, which are behavioral and physiological biometrics. Physiological biometrics are related to the physiological traits of a person such as DNA, fingerprint, iris, retina etc. While behavioral biometrics are related to the behavior pattern of a person like signature, voice, keystroke, gait etc. Biometric system based on physiological traits provides high accuracy but the cost of implementation is high. Biometric systems which uses behavioral traits as a medium for authentication are still under research because the accuracy is little bit low as compared to the biometric system based on physiological traits. Keystroke Biometric Authentication System falls under the behavioral biometric systems which uses the concepts of keystroke dynamics to create a template of typing pattern of a person.

Researchers have found that one every has a different typing pattern. Certainly, no one can copy someone else's typing pattern because the timing information of keystrokes is calculated at nanosecond level and typing is a person's behavioral characteristic, which are difficult to duplicate. Taking both the facts into consideration, we can say that it's almost impossible for an imposter to copy the typing pattern of a genuine user making the KB Authentication System more secure.

Information like Flight Time (FT) and Dwell Time (DT) are extracted during the process of creating the template of typing pattern. This template is then used against the user for his authentication.

Keystroke Biometric Authentication system has been developed using various methods which can be broadly categorized into statistical approach and machine learning approach.

Because of its higher accuracy, ease of implementation and simplicity, Statistical approach is the most common and popular approach to implement a keystroke dynamic system. The inclusive statistical measures encompass k-nearest neighbor, standard deviation, statistical t-test, mean and median.

The core idea behind machine learning approach is to classify and identify a pattern and make the correct conclusion based on the provided data. Sub domain of machine learning includes fuzzy logic, neural networks, evolutionary computing and decision tree. Better results can be produced with the help of neural networks however, it also requires the intruder's keystroke pattern to train the network. It's almost impractical to obtain intruder's keystroke pattern at the early enrollment stage.

Taking the above factors into account, we propose a KB authentication system which uses K-Nearest Neighbor as classifier. A demo Banking application has been developed which uses KB system for authentication of users.

II. Literature Survey

A novel approach was proposed by Hosseeinzadeh et al. [1] which uses up-up keystroke latency feature. In comparison with the prevalent key down-down and hold-down features, the up-up keystroke latency (UUKL) features proved to be more beneficial. The comparison was done using a GMM based verification system.

Classification of keystroke feature vectors as heterogeneous and homogeneous was done by Balagani et al. [2]. He then concluded that for short reference text, higher discriminability can be achieved if the features are classified as heterogeneous vector instead of aggregate vector. However, the difference between the discriminability of aggregate vectors and heterogeneous vectors tends to decrease as the text size increases.

Purgason et al. [3] proposed a method which uses the time to transition from one finger to another as the main timing feature. For analysis purpose feed forward neural network was used.

Analysis of keystrokes of free text which incorporates the examination of digraphs and monographs was done by Ahmed A. Ahmed et al. [4]. He used ANN to anticipate the missing di - graphs .The anticipation was based on the correlation between the monitored keystrokes.

Wangsuk et al.[5] used features as inter key time, hold time, and latency time and used trajectory dissimilarity technique to verify user's typing behavior on a username as extra identification token.

III. System Architecture

A typical Keystroke dynamic system creates a template of typing pattern based on the calculated timing information. We propose to use KNN as a classifier on this timing information for template creation.

All keystroke dynamics assessment involves some common steps like

(1) Providing a typing task to a user

(2) Taking down the keystroke timing details

(3) Feature extraction as per the classifiers requirements.

(4) Training the classifier

(5) Testing the classifier

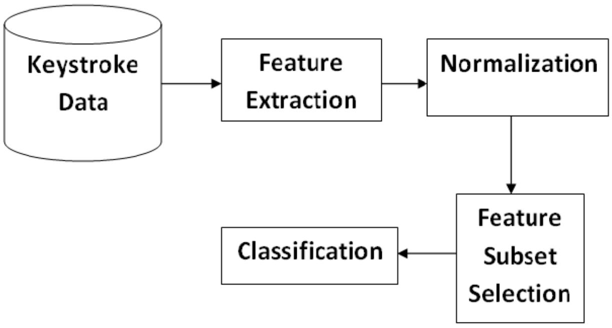


Fig. 1. *Stages for template creation by typing pattern.[7]*

While typing, timing information like the duration for which a key is pressed, the time interval between pressing one key and next key and at what time a key is pressed can be calculated by a computer. Dwell time is the time at which a key is pressed and Flight time is the time measured between key up and the key down. From these raw data four features can be extracted which are -

(1) Key Hold time (KH) – the time span between pressing and releasing a key, also known as the dwell time.

(2) Key Press Latency (KPL) – the time interval between pressing two successive keys.

(3) Key Release Latency (KRL) – time interval between releasing two successive keys.

(4) Key Interval (KI) – time span between releasing one key and pressing another, also known as flight time.

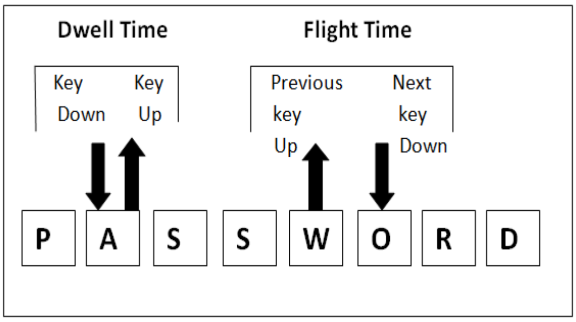


Fig. 1. *Dwell time, Flight Time and Inter Key Latency.[7].*

Additional timing information like total time taken to write a word, n-graph (di-graph or tri-graph) can also be calculated. Di-graph falls under Key Press Latency category. N-graph contains n successive keystrokes, this explains di-graph (containing two successive keystrokes) and tri-graph (containing three successive keystrokes).

The extracted features are then processed to get the template of typing pattern of a user using KNN.

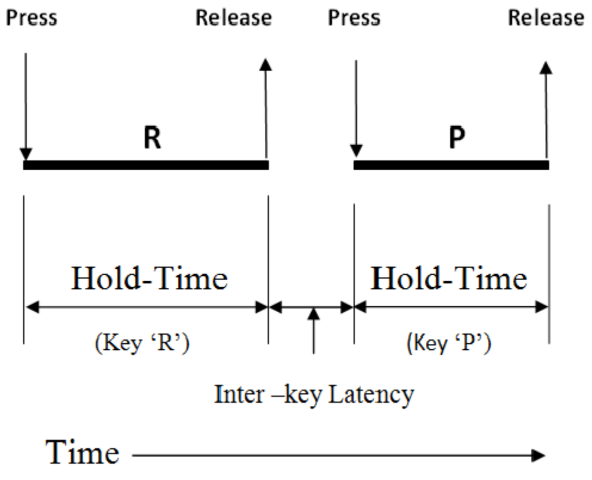


Fig. 2. *Dwell time, Flight Time and Inter Key Latency Contd.[7].*

A. Traditional Matrices for Keystroke Dynamics -

Due to the presence of many classifiers, these models are ratified based on the following security metrics -

1. False Acceptance rate (FAR) – the percentage of imposters accepted genuine user.

2. False Rejection rate (FRR) – the percentage of genuine users rejected from using the system.

3. Equal Error rate (EER) – The ratio of FAR divided by FRR. Represents overall effectiveness of the system.

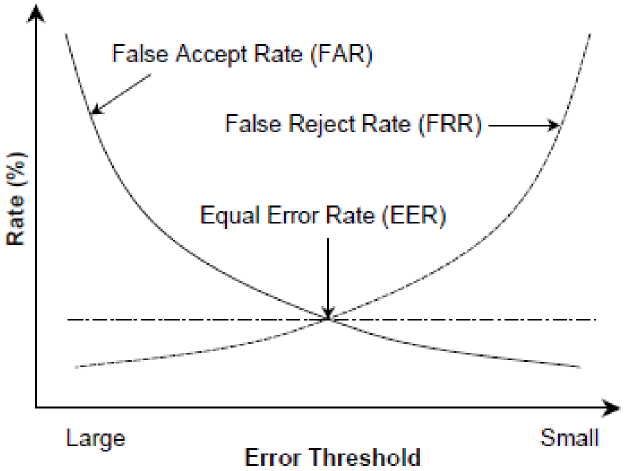


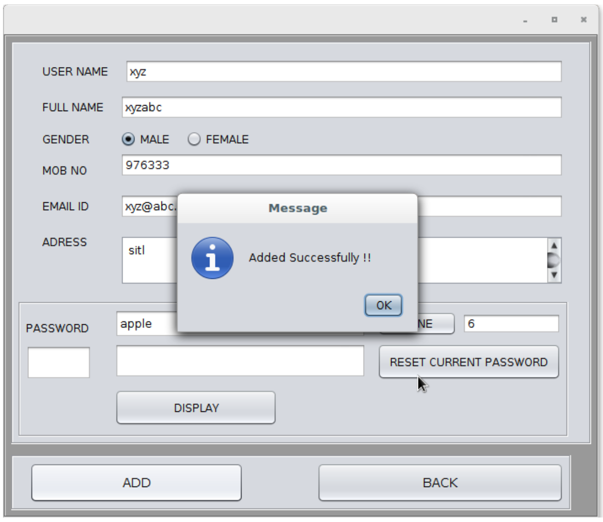
Fig. 3. *Relation between FAR, FRR and EER.*

IV. Proposed Work

The proposed system is a banking application which has a KB Authentication system working underneath the login system. The system calculates the timing information, dwell time and pressing time, and passes it to the KNN classifier for further operations.

A. Adding New User

Like any other banking application, user has to create a new account. The main difference between the traditional system and this system is that user has to enter the password 4 to 6 times to train the system. The accuracy of the system directly depends on the number of times user enter the password. In fig. 5 We entered the password 6 times to train the system.



USER NAME xyz
FULL NAME xyzabc
GENDER ● MALE ○ FEMALE
MOB NO 976333
EMAIL ID xyz@abc.
ADRESS sitl

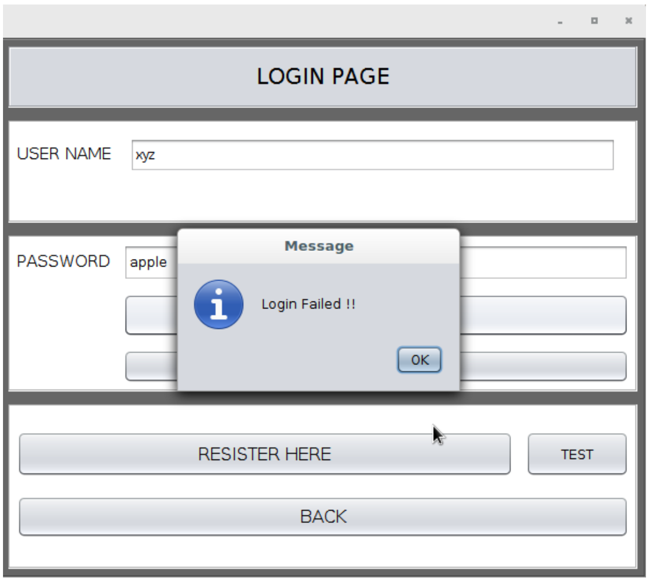
Message
Added Successfully !!
OK

PASSWORD apple
NE 6

Fig. 4. *The Add-User Window*

B. Login and Testing -

The test case for this system would be providing correct password but by an imposter.



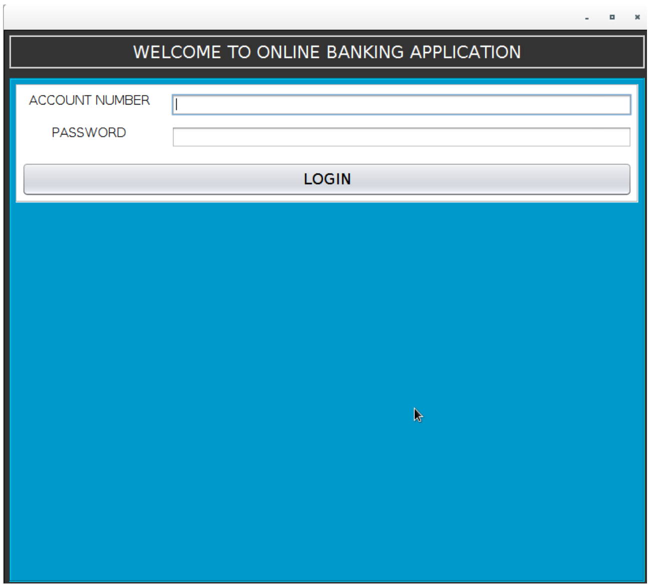
LOGIN PAGE
USER NAME xyz
PASSWORD apple

Message
Login Failed !!
OK

Fig. 5. *Login Window*

Fig 6. shows that even when we provide the correct password, which is 'apple' in this case, the system is not authorizing the claimant because he is an imposter.

A. Banking Application -


WELCOME TO ONLINE BANKING APPLICATION

If user passes the first layer of security, which uses keystroke dynamics for authentication, then only

he/she can proceed to the next step which is the actual login to the system using the account number and password.

B. Storage Database and Retraining

MySql database has been used to store the user's information and keystroke's timing information.

The feature of retraining is also added to provide more accuracy to the system. The system retrain istself every time the user login which keeps the system aware of user's current typing pattern.

V. Application Area

To gain the access to a computational system via biometric security system, behavioral or physiological characteristics are needed which clearly adds an extra level of security to the authentication system since it proves the claimant's identity physically. Keystroke dynamics is a behavioral biometric which is not used much in the field of security. The KB Authentication system can be further categorized into static authentication system and dynamic authentication system. Thus a large category of applications can also benefit from these authentication systems. KB system can be useful in some well known issues like online attacks, banking applications and online examination system.

VI. Conclusion and Recommendation

Our work focuses on the importance of keystroke dynamics for laptops, desktops and mobile devices. Due to the cost effectiveness and compatibility (no additional hardware is needed) of KB Authentication system, it is obviously a much better choice to provide extra security to the computational systems rather than using the traditional methods like retina or fingerprint scan. The main issue in the field of keystroke dynamics is there are no standard protocol for keystroke system calculation which would be helpful for producing more accurate results.

VII. References

[1] Danoush Hosseinzadeh and Sridhar Krishnan,‖ Gaussian Mixture Modeling of Keystroke Patterns for Biometric Applications‖, IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 38, no. 6, 2008,pp 816-826.

[2] Kiran S. Balagani , Vir V. Phoha , Asok Ray and Shashi Phoha ,‖ On the discriminability of keystroke feature vectors used in fixed text keystroke authentication‖, in Pattern Recognition Letters 32, 2011, pp 1070–1080.

[3] Benjamin Purgason and David Hibler, ―Security through Behavioural Biometrics and Artificial Intelligence‖, Procedia Computer Science, vol 12, 2012, pp 398 – 403.

[4] Ahmed A. Ahmed and Traore Issa, ―Biometric Recognition Based on Free-Text Keystroke Dynamics", IEEE Transactions on cybernetics, 2013.

[5] Kasem Wangsuk and Tanapat Anusas-amornkul,‖ Trajectory Mining for Keystroke Dynamics Authentication.‖, in Procedia Computer Science 24, 2013, pp175 – 183.

[6] Kasem Wangsuk and Tanapat Anusas-amornkul,‖ Trajectory Mining for Keystroke Dynamics Authentication.‖, in Procedia Computer Science 24, 2013, pp175 – 183.

[7] Amar L. Renke and Rohit A. Patil, Keystroke Dynamics for User Authentication and Identification by using Typing Rythm, in International Journal of Computer Applications (0975 – 8887) Volume 144 – No.9, June 2016

[8] Khandaker A. Rahman , Kiran S. Balagani , Vir V. Phoha ,‖Snoop- Forge-Replay Attacks on Continuous verification with Keystrokes‖, IEEE Transactions on information forensics and security, vol. 8, no. 3, pp 528-541, 2013.