User Verification by the Keystroke Dynamics System for Security in Online System

Sameer Dass[1], Suresh Kumar[2]
[1, 2] *Dept. of Computer Science, AIACTR*
(E-mail: dass.sameer2@gmail.com)
(E-mail: sureshkumar@aiactr.ac.in)

***Abstract*—**Biometrics software which works on typing dynamics is a comparatively profitable and practical method for providing immunity against attack, loss and theft and provide protection. Keystroke dynamics which relies on the typing style of the login access, is an effective way of figuring out valid or spoof login. In today's world, there are many problems which one faces while computing. Specially those utilized in online banking, online shopping, cloud computing, e-learning, multi-usage of computer and distinctive offerings over the internet. Keystroke dynamics provides immunity to the password in this volatile era. This technology is primarily and totally on human conduct and the manner in which one type password. Since, Computers are used in each and every field, hence it becomes quite relevant to keep user credential and private records secure.

Keywords*—Bioinformatics;, Keystroke Dynamics,; Manhattan distance;, Mahalanobis Distance;, Biometrics.*

I. INTRODUCTION

Authenticating a client's identification is important part of business in today's world. Many of those systems save rather touchy, private, business, personal or economic records which leave people vulnerable. Unauthorized get right of entry to such information will lead to loss of cash or unwanted disclosure of pretty special data that threats the safety of Information. Biometric technology are described as automatic method for verifying and spotting method to identification of a living character that are primarily based on physiological or behavioral trends. Biometrics strategies are specially used for consumer/user authentication. "Biometrics" means "biological" but the time period is normally associated with the usage of specific physiological characteristics to understand an identity of user.

We proposed a dynamic system which takes the typing pattern of user, in which we collect the certain parameters like flight time, dwell time and latency.

We possibly can broaden a model that captures probably unique characteristics that may be used for the identity of an individual. To facilitate the improvement of the model of the way the user enters their password. [1]Gaines became the primary and first researcher to file the results of a well-controlled environment. They have studied about the area of keystroke dynamics thoroughly. He examined the typing varieties of 7 expert typists – with the intention of figuring out if there had been unique typing styles that could be used to differentiate between the users. Commonly, the keystroke dynamics of the person are extracted in the course of login and in comparison with a reference model that changed into constructed based on the user's keystroke dynamics and similar functions and parameters of different users.

Keystroke dynamics evaluate to some different biometric techniques that are different from every other biometric method. additionally Keystrokes of user are vary from their previous typing pattern, due to various factors. But overall features which are discussed above are same. The essential terms applied in keystroke is that there we required software program and keyboard is needed for input. Using keyboard enter gadget we take a look at the human behavior to kind there password, what shortcuts. typing techniques, specific keys, characters utilized by man and woman.

The Physiological biometrics based totally completely authentication structures that use hardware, and therefore more high priced and time consuming to increase whilst keystroke dynamics does not need additional hardware and therefore it's far a lot much less priced.

Terms used for Keystroke Dynamics[2].

1. Press time.
2. Key release time.
3. Latency.

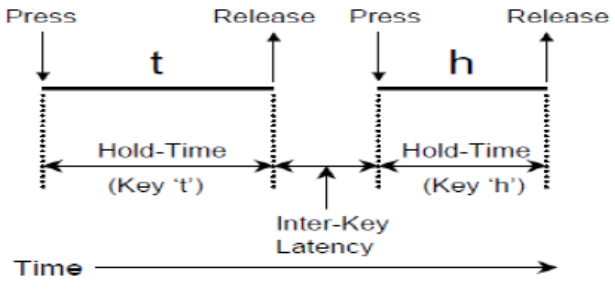

Fig. 1 Timing of keys.[3]

The predominant benefit of biometrics is that it also includes more tough to copy the biometric tendencies of an man or woman than maximum of various authentication strategies

inclusive of passwords or tokens. Many attacks exist due to which researchers and developers are offering more strong systems. Fingerprints are the maximum used biometric modality although it is quite smooth to copy the fingerprint of an person.

II. LITERATURE REVIEW

In this section we have discussed the evolution of keystroke dynamics applications. The keystroke/typing dynamics for validation and identity of a person has a protracted data and can be dated back to the 1970's. There are several of research papers, patents of software, and thesis have been published to address this trouble,signal processing, patter recognition, and tool mastering. This is a new arena which is developing and attracting research interest as with the coming of digital era, the attacks as well as the concerns have arisen. The keystroke dynamics abilities are based totally on the timing information of the key's down/maintain/up events, even though some custom business keyboard can acquire strain information. This state of affairs has changed dramatically with the recognition of the mobile clever gadgets, which can be normally embedded with a rich suite of superior sensors and might be unique within the subsequent phase. The hold time or press time of character keys, and the latency among keys, i.e., the time intervals program language period amongst the release of a key and the pressing of the subsequent key, moreover known as flight time .

Xue and Zhang [4] proposed the algorithm which maximizes the class performance and minimizes the range of functions. They used multi objective PSO-based function choice approach used to better discover the Pareto front of non-dominated answers in function choice troubles. These algorithms achieve the good set of feature sets.

Mudhafar et al. [5] proposed a paper on an anomaly detector for keystroke dynamics authentication, primarily based on a statistical diploma of proximity, evaluated through the empirical data set. A password typing rhythm is used to stumble on the real and unauthenticated person. They delivered information regarding the per word information which can be achieved by system. They calculate key down/up time of every key and latency time among keys.

Cho et al. [6] proposed an algorithm which reduces the redundant outlier values and minimizes the cluster area. Keystroke period values offers high-quality characteristic subset results at the same time as compared with different feature values. Better standard performance is finalized with keystroke duration characteristic.

Gunetti et al [7] extends further prolonged the idea of free text by means of incorporating all n-graphys. They also proposed an identity primarily based method for the authentication. However this approach isn't always sensible for a massive database due to its scalability trouble. The proposed identification method wants to examine an input with each schooling pattern of every man or woman profile. The verification process grows exponentially with the dimensions of the database.

Dowland et al. [8] took the typing samples of five clients by tracking their computer activities consistently, without any exact checks being forced on them along with telling clients to type pre-decided set of words like password. They decided on the 2-graphs only which occurred least amount of time across the gathered typing samples. They build profiles of person by means of computing the general deviation of two-graphs latency. Feature subset requirement is vital for an optimization problem that chooses the maximum reliable or near pinnacle of the road characteristic with apprehend to the overall performance measures. Since the goal is to acquire the most class accuracy and decrease the kind mistakes.

III. DETECTOR IMPLEMENTATION

A. Clustering for data sets

Clustering a hard and fast or grouping of objects into organizations is typically moved via manner of the goal of identifying internally homogenous corporations according to a selected set of dataset. In order to carry out this purpose, the area to start is computing a matrix, referred to as dissimilarity matrix, which includes facts about the dissimilarity of the determined devices. According to the nature of the discovered variables (quantitative, qualitative, binary or combined kind variables), we can define and use awesome measures of dissimilarity. Cluster evaluation itself isn't one particular algorithm, however the trendy project to be solved. It may be finished through several algorithms that range considerably of their perception of what constitutes a cluster and the manner to efficaciously locate them. So we used the hierarchal clustering it uses the interrelation of nearby data sets, which is required by need to calculate the measure of the data set. There is need of hierarchal structure for clusters.

a) Manhattan distance

The Manhattan distance is easy in computation and easy decomposition. Most importantly, it is extra sturdy to the have an impact on outliers in contrast to better order distance metrics along with Euclidean distance and Mahalanobis distance. The Manhattan distance has a statistical interpretation. It is associated with the log chance of the multivariate Laplace distribution with an identity covariance matrix. It is similar to the Manhattan detector except outliers are filtered from the training records. In the guidance phase, the mean vector of the timing vectors is calculated, and the identical antique deviation for each function is calculated additionally. Any timing vector detail this is greater than 3 widespread-deviations above imply is eliminated, and a more robust imply vector is computed without those intense values. In the check phase, the ambiguity score is calculated as the Manhattan distance amongst this robust mean vector and the take a look at vector.

It is equation is |x1-x2|+|y1-y2|

Manhattan has performed well over Euclidean distance, Vector Cosine distance.

b) Mahalanobis distance

This Mahalanobis distance is based on the covariance of variable facts to correct the heterogeneity and non-isotropy positioned in most real records. Mahalanobis distance is improved version of Euclidean distance to account for correlations and capabilities[9]. In the training phase of the data set, both the endorse vector and the covariance matrix of the timing vectors are calculated. In statistical literature, the Mahalanobis distance is associated with the log probability underneath the concept that data observe multivariate Gaussian distribution this is an less expensive approximation of maximum realistic information.

The Mahalanobis distance two feature set of vector calculation x and y

$$||x-y||^2 = (x - y)TS^{-1}(x - y)).$$

IV. USING THE TEMPLATE

Dwell time: It is total time of taken by a user in pressing and releasing the single key.

Flight time: It is total time taken by a user in pressing and releasing the successive two keys.

Pressing time: It is the total holding time to release the key.

Release time: It is the time at which key is released.

Latency of keys: It is the difference of timing between two consecutive key presses or key releases toolbar.

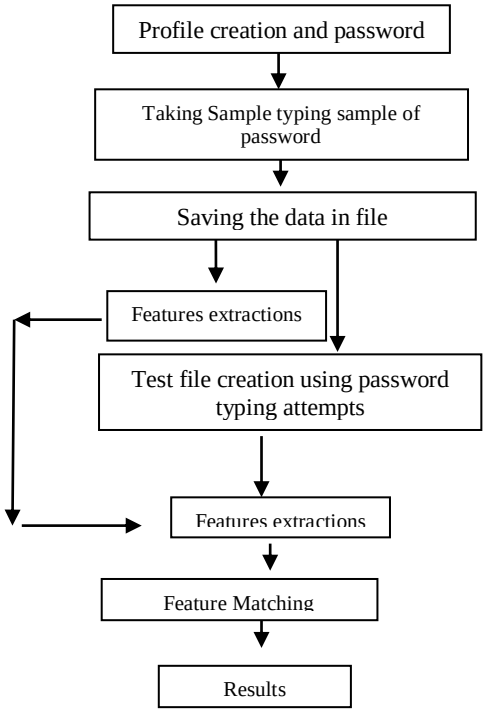

Fig. 2: Flow chart of keystroke dynamics system.

V. PROPOSED SYSTEM

This proposed system is used to find the keystroke dynamics for user verification as a second authentication in windows applications. We created this system on java on language using netbeans and fxml language. Netbeans software used to develop the java software which has its own IDE, IDE helps us to find the result at any instance. This application can run on any type of platform of operating system.

A. User Start Profile Page.

Figure: 3 Creating of password and user profile.

When a user start an application they need to choose the option of record keystroke option . Then we need to fill the user name which have to store the typing speed of keystroke of password. And click on next button . Repeat the step 5 times. Add more User and their typing pattern or speed As show in Figure 3.

Click on the Generate data button which generate the whole data with their typing speed and name of user. After that click on write to file on option then select to store the data file on desired location.

B. Test cases or User trial side

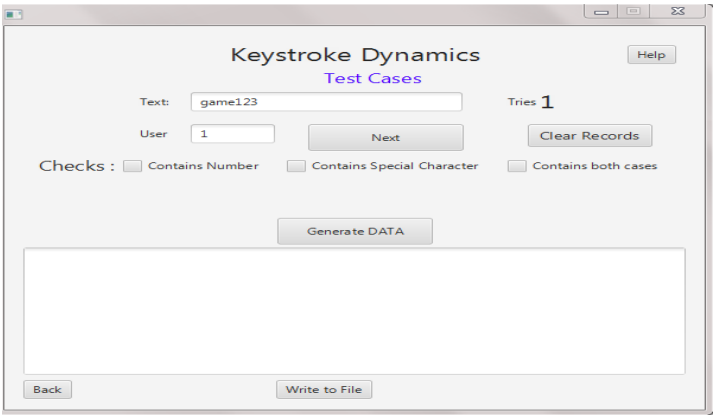

Figure 4: Generating test cases for different and anonymous user.

After saving the typing speed of user , Now we have to test the data for different user and same user for verifying the user. Fill

the test case or user name and tries the their typing speed in test button .After that click on next button.

Test with different names and different user typing speed by typing the password on text field. Now we generate the data as like profile page. Click on the generate the data option which generate the data of all user who tries the typing of password in the application . Test file is generated as in profile page. Simultaneously feature extraction is need done by automatic function after saving the file.

C. Featuring Matching.

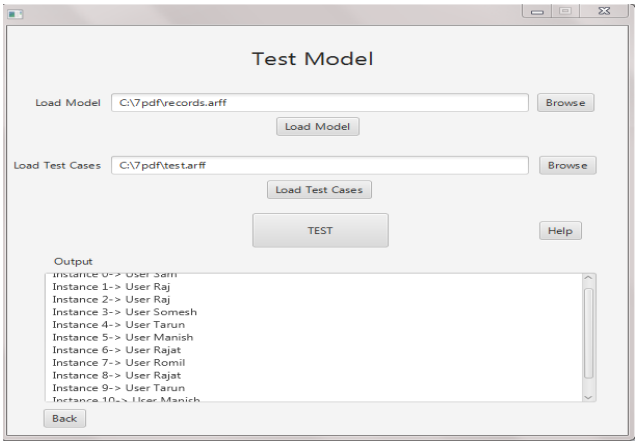

Figure 5: Verifying the users.

Now we have to evaluate the both files feature extraction by opening respective file and choosing the browse option. After that we have get all log of all people who tries to access the system, with their current instances.

VI. PERFORMANCE ANALYSIS OR RESULTS

For performance of each algorithm and our application we have decide to use the three factors.

- Feature Reduction Rate
- Error authentication rate
- Latency Domain

Feature reduction rate: Feature preference is likewise mentioned as variable choice or feature reduction or attributes choice or variable subset desire. This is the method of choosing a subset of relevant talents for constructing sturdy mastering fashions. Subset choice searches the set of viable features for the most desirable subset. The feature reduction rate may be calculated for the price of bargain of the whole functions. This rate of rejection may be calculated with the help of the entire variety of functions and the variety of competencies decided on for authentication. The feature reduction rate may be calculated using the following components.

Feature reductions rate = (Total no. features – No. of feature selected) / Total no. of features.

Here the no. of features are the timing of key presses of the text or password. And the selected features are also the time we have selected for clustering purposes and its algorithm.

Error authentication rate: Rate of Rejection of Valid Users or False Non-Match Rate is a danger that a valid patron's keystroke will incorrectly be considered a non-match for his/her reference template. Acceptance of invalid customers or false healthy rate is the possibility that an individual's template will wrongly be taken into consideration a healthful for a one-of-a-kind character's keystroke sample. The errors charge may be discovered from the technique as follows:-

Error rate = (Total no. of errors / Total no. of samples)* 100.

Here total no error is numbers of wrong answer given by the application during providing the keystroke samples.

Latency Domain: In this factor we have to calculate the latency of timing of keystroke. This latency method create a specify the domain or range of particular password or keystroke of text. This domain figure helps in finding the nearest keystroke timing similarity password.

Latency Domain is remain less than 50 ms .

Latency factor = Latency of keystroke * No of keystrokes of text.

Latency Domain = Original Latency factor – Latency factor.

Less than 50 value of Latency domain is acceptable.

Method	Total no. of features	Feature Selected	Feature reduction Rate
Manhattan distance	8	5	37.5
Mahalanobis distance	8	6	25

Table: 1 Feature Reduction Rate Table.

Method	Total no. samples	Total no Error	Error rate
Manhattan distance	50	6	12
Mahalanobis distance	50	4	8

Table 2: Error Authentication Rate.

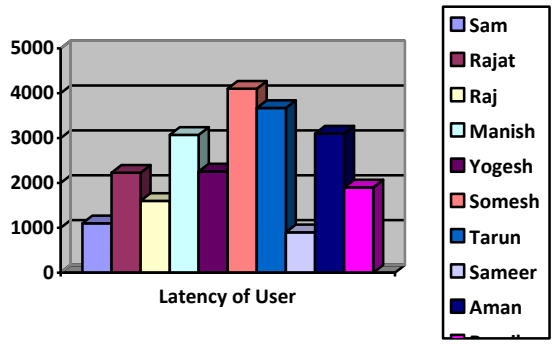

Fig 6: Graph of Latency of User

VII. CONCLUSION

In this paper we deal with the practical importance of the use of keystroke dynamics as a biometric for authenticating get admission to workstations. Keystroke dynamics is the technique of analyzing the way customers kind via way of tracking keyboard inputs and authenticating them primarily based on habitual forms of their typing rhythm. We have used the timing of key pressed like dwell time, flight time, pressing time , releasing time and latency .The evolution and research into biometric errors testing fake refuse and fake stay for has been of enthusiastic interest to biometric software developer. We have developed a system which can easily verify the user on the basis of user typing pattern. This application is developed in java language. The main drawback of this system is that sometime genuine user uses the new or different keyboard effect the typing pattern, sometime other physical environment also effects the typing of user like sickness or angriness. But we will work in this project and will find out the answer for that and the new features which can find the mood and emotions of user while typing the passwords using sensors which may be the better advancement of this keystroke dynamics. If we make all keyboards of equal fashion having equal capabilities then it offers higher/greater effectiveness.

REFERENCES

[1] Gaines, R. Lisowski, W., Press, S., & Shapiro, N. ,Authentication by keystroke timing: Some preliminary results. Rand Report R-256-NSF. (1980).

[2] S. J. Shepherd, “Continuous authentication by analysis of keyboard typing characteristics”, Proc. IEEE European Convention on Security and Detection, vol. 16-18, May 1995, pp. 111- 114.I.

[3] Heather Crawford “Keystroke Dynamics: Characteristics and Opportunities” Department of Computing Science Sir Alwyn