IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 2, Ver. III (Mar-Apr. 2016), PP 54-60
www.iosrjournals.org

Identifiying the Authorised User by Typing Speed Comparison

S.Margarat Sheeba[1], G.Sahana[2], M.Sneka[3], M.Sowmiya[4], T.Sathish Kumar[5]
Computer Science And Engineering, Saranathan College Of Engineering, Trichy

Abstract: *Computers have become a ubiquitous part of the modern society. Online attacks on companies resulted in the shutdown of their networks and compromised the passwords and personal information of millions of users. Since we depend so much on computers to store and process sensitive information, it has become all the more necessary to secure them from intruders. For user authentication and identification in computer based applications, there is a need for simple, low-cost and unobtrusive device. A user can be defined as a person who attempts to access information stored on the computer or online using standard input device such as the keyboard. Use of biometrics such as face, fingerprints and signature requires additional tools to acquire the biometric which leads to an increase in costs. Use of a behavioral biometric which makes use of the typing pattern of an individual can be obtained using existing systems such as the standard keyboard, making it an inexpensive and extremely attractive technique. One of the major advantages of this biometric is that it is non-intrusive and can be applied covertly to augment existing cyber-security systems. In this paper we implement new process for password text analysis of keystrokes that joins monograph and digraph analysis, and uses a neural network to predict missing digraphs based on the relation between the monitored keystrokes.*

Keywords*: Behavioral biometrics, Keystrokes, Typing patterns, Classification, Neural network, Monograph, Digraph*

I. Introduction

All internet applications require the user to use an authentication scheme to make sure that only the genuine individual can login to the application.

In security systems, authentication and authorization are two complementary mechanisms for determining who can access information resources over a network. Authentication is process of verifying user identity for accessing the system. Authorization is the process of giving individuals access to system object based on their identity. This information can be broadly subdivided into three categories namely knowledge, token, and biometrics-based authentication.

1.1 Biometric Based Authentication:

Biometrics provides the metrics related to person characteristics and traits. Biometrics authentication (or realistic authentication) is worn in computer science as an outline of identification and access control. It is also used to classify individuals in groups that are under surveillance. Biometric identifiers are the characteristic, measurable characteristics used to make and describe individuals. Biometric identifiers are frequently categorized as physiological versus behavioral characteristics.

Physiological characteristics are related to the form of the body. Examples include, but are not partial to fingerprint, palm veins, face recognition, iris recognition, DNA, palm print, geometry, odour /scent. Behavioral characteristics are correlated to the pattern of behavior of a person, including but not limited to typing rhythm, voice and gait. Some researchers comprise coined the term behavior metrics to explain the latter class of biometrics. More traditional means of access control include token-based identification systems, such as a driver's license or passport, and knowledge-based identification is a password or personal identification number. Since biometric identifiers are single to individuals, they are more reliable in verifying identity than token and knowledge-based methods; however, the collection of biometric identifiers raises privacy concerns about the final use of this information. The biometric system is shown in fig 1.

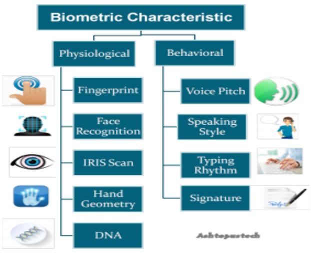


Fig 1. Biometric system

II. Objectives

Our objective is to collect a keystroke-dynamics data set, to develop a repeatable evaluation procedure, and to measure the performance of a range of detectors so that the results can be compared soundly. We collected data from 100 users typing 400 passwords each, and we implemented and evaluated 10 detectors from the keystroke dynamics and pattern-recognition literature. The three top-performing detectors achieve equal-error rates between 8.6% and 9.2%. The results along with the shared data and evaluation methodology constitute a benchmark for comparing detectors and measuring progress.

III. Related Work

A. On keystroke dynamics

Fabian Monrose, et,al,… [1] addressed the sensible importance of using keystroke dynamics as a biometric for validating access to workstations. Keystroke dynamics is the procedure of analyzing the way users type by monitoring keyboard inputs and authenticating them based on habitual patterns in their typing rhythm.

Fabian Monrose, et.al,… [2] presented a narrative approach for hardening passwords by discovers the keystroke dynamics of users. This approach enables the generation of a long-term secret (the hardened password) that can be tested for login purposes or used for encryption of files, entrance to a virtual private network, etc.

Yong Sheng,et,al,… [3] proposed a Monte Carlo approach to achieve sufficient training data, a splitting method to improve effectiveness, and a system composed of parallel decision trees (DTs) to authenticate users based on keystroke patterns. For any research involving pattern recognition techniques, one should conduct an experiment to collect two sets of data. One is for training or reference, and the other is for testing or verification.

Ramaswamy palaniappan, et, al,… [4**]** proposed as a biometric for verification of the identities of individuals in a small group. The approach is based on a novel two-stage biometric authentication method that minimizes both false accept error (FAE) and false reject error (FRE). It would be useful to study which are the most discriminatory features and using these would circumvent the requirement of using PCA to reduce the number of features.

Rajkumar Janakiraman, et, al,… [5**]** studied the feasibility of using Keystroke Dynamics as a biometric in a more general setting, where users go about their normal daily activities of emailing, web surfing, and so on. Ht). We define Held Time as the time (in milliseconds) between a key press and a key release of the same key. This is defined as the time in milliseconds between two consecutive keystrokes. A Sequence can be of any length, the minimum being two. From the samples of the Sequence appearing in the keystroke data, we estimate the probability density function (pdf) for each element in the feature vector Ft. This information is stored as a normalized histogram for the Sequence, which in turn will be used for classification.

IV. Existing System

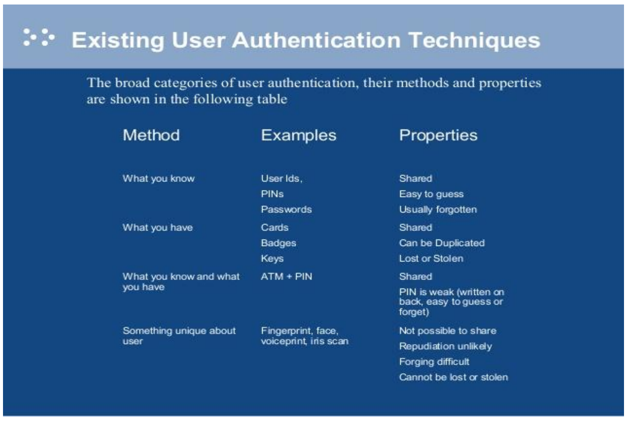
Keystroke dynamics features are typically extracted using the time information of the key down and up events. The hold time or dwell time of each keys, and the latency between both keys, i.e., the time interval between the free of a key and the pressing of the next key are typically exploited. Digraphs, which are the time latencies between two consecutive keystrokes, are commonly used. Trigraph, which are the time latencies between each three consecutive keys, and similarly, n-graphs, have been investigated as well. In their background on keystroke analysis using free text, investigated the efficiency of digraphs and more generally n-graphs for free text keystroke biometrics, and finished that n-graphs are discriminative only when they are word-specific. As such, the digraph and n-graph features carry out the dependence on the word context they are computed in.

The use of keystroke dynamics for confirmation and identification functions was first investigated back in the 1970's. In traditional did a beginning study on keystroke dynamics based verification using the T-test on digraph textures. Then extracted keystroke features using the mean and variance of digraphs and trigraphs. By means of the Euclidean distance metric with Bayesian-like classifiers, they reported a correct identification rate of 90% for their dataset containing 50 users.

After that proposed to use the relative array of duration times for dissimilar n-graphs to take out keystroke features that were found to be extra robust to the intra-class variations than total timing. They demonstrated that the new relation feature, when combined with features using total timing, improved the authentication presentation using free text.

Over the years, keystroke biometrics research has utilized lots of existing machine learning and classification techniques. Different distance measures, such as the Euclidean distance, Mahalanobis distance, and the Manhattan distance, have been explored. Both classical and superior classifiers have been used, including K-Nearest Neighbor (KNN) classifiers; support vector machines (SVMs), K-means method, Bayesian classifiers, Fuzzy logic, and neural networks. A huge range of performance numbers has been published. However, it is not probable to make comparison of a variety of algorithms directly because of the use of dissimilar datasets and evaluation criteria across the studies. To address this matter, keystroke dynamics databases including standard results of popular keystroke biometrics algorithms have been published to supply a standard experimental platform for progress assessment.

A large data are collected and published a keystroke dynamics benchmark dataset containing 51 subjects with 400 keystroke dynamics collected for each subject. Furthermore, they assessed fourteen available keystroke dynamics algorithms on this dataset, including Neural Networks, SVMs, K-means, Fuzzy Logic, KNNs, Outlier Elimination etc. Various distance metrics, including Euclidean distance, Manhattan distance and Mahalanobis distance were worned. This keystroke dataset beside with the evaluation methodology and state of the art presentation provides a benchmark to impartially measure the progresses of novel keystroke biometric algorithms. The Existing system user authentication techniques are shown in fig 2:



Existing User Authentication Techniques

The broad categories of user authentication, their methods and properties are shown in the following table

Method	Examples	Properties
What you know	User Ids, PINs Passwords	Shared Easy to guess Usually forgotten
What you have	Cards Badges Keys	Shared Can be Duplicated Lost or Stolen
What you know and what you have	ATM + PIN	Shared PIN is weak (written on back, easy to guess or forget)
Something unique about user	Fingerprint, face, voiceprint, iris scan	Not possible to share Repudiation unlikely Forging difficult Cannot be lost or stolen

Fig 2: Existing User Authentication Techniques

V. Proposed Framework

The aim of this section is to implement the monograph and digraph techniques to analyse the free text keystroke. The neural network classifiers are implemented to find the behavioural patterns.

5.1 Monograph

Monograph is an event consisting of only a single key at a time. Monograph modeling consists of capturing monograph timings and prepare sorted time mapping table for monograph data. Monograph timing of a key code can be computed as the time difference between key DOWN event and key UP event for that key code corresponding to the key which was pressed and released.

5.2 Digraph

Digraph is an event consisting of two consecutive key events. For capturing digraph data the raw data is taken, and the difference in timing of UP and DOWN key events gives the digraph time. Using monograph and digraph techniques we calculate the monograph and digraph technique.

5.3 Flight time

It is the latency between consecutive keystrokes or the time between two consecutive keys. The inter stroke interval between the keys is measured in milliseconds.

5.4 Dwell time

It is the time interval between the key down event and the key up event or in other words the time for which each key is being pressed.

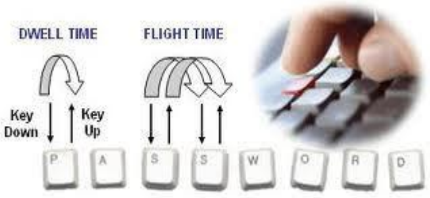


Fig:3 Dwell and flight time calculation

5.5. Neural network classifiers

Artificial neural network are computational model inspired by biological neural network are used to approximate functions that are generally unknown. In this system we use the multi layer net approach.

Multi layer net is a net with one or more layers of nodes which is called hidden units, between the input units and output units. The common type of neural network consist of three groups; a layer of “input” unit is connected to a layer of hidden units which is connected to the layer of “output units”.

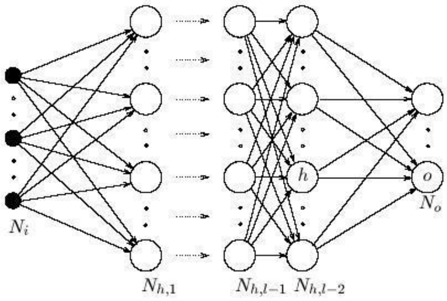


Fig:4 Neural network classifier

Raw data collected from a specific user’s sessions are processed and converted to monograph and digraph formats. Also, at this stage, outliers are removed from the mono and digraph sets before passing the data to the next stage. Outlier removal occurs only on the enrollment data; no outlier removal happens for the test data. We use Peirce’s criterion for the elimination of outliers. Peirce’s criterion is a rigorous method based on probability theory that can be used to eliminate outliers in a rational way.

It consists of an iterative process which starts with calculating the mean and standard deviation for the data set. In the first iteration an assumption is made that the data has only one doubtful observation. Based on that assumption and using the Peirce’s table, the maximum allowable deviation is calculated and all entries with greater deviations are then eliminated from the dataset. The number of doubtful observations will be incremented by one in each of the next iterations. In each of these iterations, the maximum deviation will be recalculated while maintaining the same mean and standard deviation of the original dataset. The process will be repeated until no more entries are eliminated.

After removing outliers, generated monographs and digraphs are sent as a batch to the monograph and digraph sorting modules. Each of those modules will process the data and calculate a mapping table. Calculated monograph and digraph mapping tables are considered part of the user’s signature and stored for future use. We use neural networks to model the user behavior based on the encoded sets of monographs and digraphs. Although the neural network architecture remains the same for all users, the weights are user specific. The proposed neural network architecture and provide a visual representation of the monograph and digraph signatures produced in order to illustrate the similarity or dissimilarity in behavior for different users’ sessions. The proposed framework is plotted in fig 3.

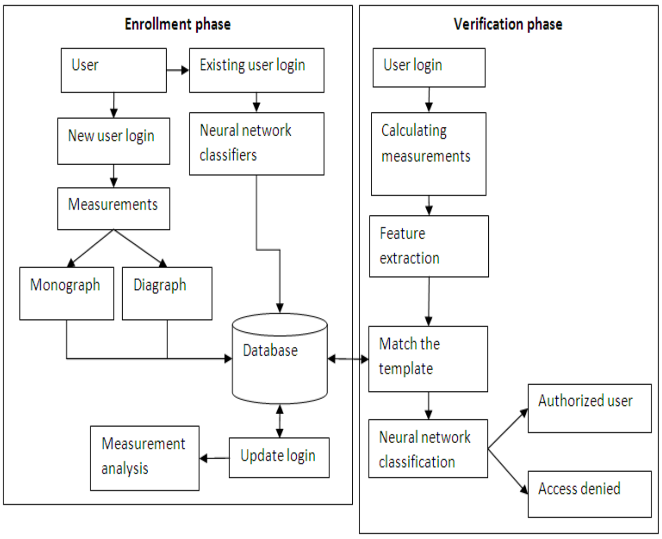


Fig:5 Proposed Framework

VI. Benefits Of Our System

- **overcoming traditional password drawbacks:**
  - Hesitation in public/spy camera- There is a chance that your password can be trapped in public places using cameras. But our system requires more than just a password to authenticate.
- **Continues verification:**
  - Verification only at the time of login does not guarantee the security for complete session. This system provides continues verification through the session by analyzing the user behavior in the session.
- Software only method(No additional hardware expect keyboard)
- Simple to deploy and (Username & passwords)- Universally accepted
- Cost effective
- No end-user Training
- Can be used over the internet
- It provides a simple natural way for increased computer security.

VII. Application

Keystroke dynamics can be used for authentication, then it is used mostly together with user ID / password credentials as a form of multifactor authentication**.**

Another use is as a very specific form of surveillance**.** There exist software solutions which, often without end-users being aware of it, track keystroke dynamics for each user account. This tracking, historization of keystroke dynamics is then used to analyse whether accounts are being shared or in general are used by people different from the genuine account owner. Reasons for such an implementation could be verification of users following security procedures (password sharing) or to verify that no software licenses are being shared (especially for SAAS applications).

Companies which develop software products applying keystroke dynamics are:

- ID Control is a dutch company developing strong but affordable authentication solutions, some of which use keystroke dynamics. Their software integrates with MS Windows logon, Citrix, VPN and many others.
- BehavioSec is a swedish company specialized in continuous authentication systems, this is software which monitors activity on a computer to make sure that it is the genuine account owner who is using the computer. Behaviosec uses not only keystroke dynamics but also mouse dynamics and the general way in which the user interacts with the computer.

VIII. Experimental Results

On performance metrics such as false positive rate, false negative rate. Our proposed work provide reduced rate at the time of key stroke dynamics. Performance chart is showed in fig 4.

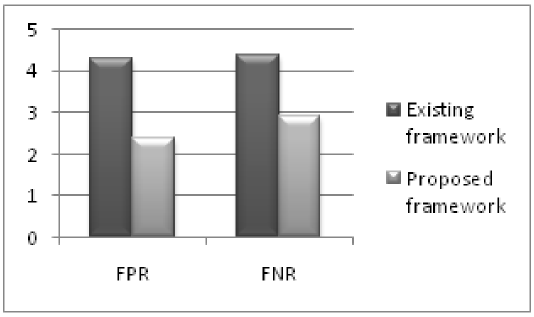


Fig:6 Performance chart

IX. Implementation Of Our Project

After reviewing the keystroke dynamics literature studies, below are some of the suggestions and potential areas that can be explored by researchers in the keystroke dynamics domain.

9.1 Features Of Quality Measures And Enchancement

One of the immediate approaches to enhance performance of keystroke dynamics recognition is by focusing on introducing new detector or classification algorithm. However, another potential route that may be looked into is by providing these detectors with higher quality feature data. A bold approach which introduced the use of artificial rhythm and cues to increase uniqueness of typing feature is a preliminary step forward in this aspect. Feature quality may also be boosted by fine tuning timing resolution, dynamic feature selection, data filtration, and feature data fusion.

9.2 Mobile Platform And Touch Screen Devices

As technology evolution grows, mobile and portable devices have been ubiquitous in human’s daily life. Smart phone and tablet have ever increasing memory and processing power as compared to few years ago. Furthermore, the introduction of advance and sensitive miniature hardware sensors such as multi-touch screen, pressure sensitive panels, accelerometer, and gyroscope has the potential of unleashing new feature data. This improved hardware is now readily available and paves a way for future keystroke dynamics research study on this platform.

9.3 Dynamic Authentication

As compared to static one-off authentication mode, keystroke dynamics research on dynamic or continuous authentication is still rather inadequate. Several research works in the literature have laid the foundation on continuous authentication on free and long text input. Potential untapped area would be continuous authentication on foreign languages such as Korean, Chinese, Italian, and non-English word (informal short abbreviation). Additionally, experimental platform should be accentuated on web browser-based authentication since the computer usage trend has be shifted from operating system-based application to browser-based cloud services. Therefore, continuous and uninterrupted validation of user identity throughout the session of accessing these services within the online platform is in high demand.

9.4 Retraining Mechanism Evaluation

Keystroke dynamics biometrics are sub-domain of behavioral biometrics that have the possibility of evolvement over time. More extensive studies need to be conducted particularly on update mechanism if keystroke dynamics are to be used as a long-term security enhancement tool. Result evaluation and the effectiveness of a retraining algorithm or framework should be assessed in stages across a longer period of time (e.g., 6–12 months) to allow time for accommodating the gradual change of typing pattern.

9.5 Benchmark Dataset

In long term, keystroke dynamics research community should be encouraged to come up with a shared benchmark dataset wherever possible. Development of homemade dataset may cater to individual experimental needs; however, experiment result cross-comparison between different methodologies employed may not be conclusive. Furthermore, some researchers may not have the resource to develop a proper dataset for experiment. We would recommend the community to produce 3 types of dataset with both free and fixed text from keyboard input as well as numerical input data from mobile phone. These would be sufficient to cater keystroke dynamics research across the 3 major platforms. A sample size of at least 100 or more should be an initial aim. Dataset owner is encouraged to share the data collection tool if possible, so that others may help contribute to the data collection process. At such, not only can the benchmark sample size increases gradually over time but also the opportunity to collect keystroke typing samples from diverse communities across the globe.

X. Conclusion

In this paper examined the ability of keystroke dynamics authentication systems for their application to real world systems. We can observe through this learn that, even with quite straightforward methods from the state of the art, the acquired results are almost correct but require yet to be improved. The neural network approach per users for the training seems to give better results. The proposed system present some helpful and easy methods in order to get better quality of a keystroke dynamics system without adjusting the algorithms: it is up to the users of the system to add breaks in their mode of typing (they can be helped by the software with visual cues). Maybe these good qualities with practices of composing keystroke dynamic based passwords could be better acknowledged than the good practice of traditional passwords

Majority of the keystroke dynamics research works from the last three decades have been summarized and analyzed in this paper. It is by no means to be an exhausted archive of all research works in the keystroke dynamics domain, but it was collected with the resource available and to the best of our knowledge at the point of writing. The aim of this review paper is to provide a reference for researchers to further look into others work to identify promising research direction for further study. We believe that this will also significantly lower the entry barrier especially for novice researchers who are interested in keystroke dynamics.

The literature study suggested that keystroke dynamics biometrics are unlikely to replace existing knowledge-based authentication entirely and it is also not robust enough to be a sole biometric authenticator. However, the advantage of keystroke dynamics is indisputable such as the ability to operate in stealth mode, low implementation cost, high user acceptance, and ease of integration to existing security systems. These create the basis of a potentially effective way of enhancing overall security rating by playing a significant role in part of a larger multifactor authentication mechanism.

References

[1]. F. Bergadano, D. Gunetti, and C. Picardi, “User authentication through keystroke dynamics,” ACM Trans. Inform. Syst. Security, vol. 5, no. 4, pp. 367–397, Nov. 2002.

[2]. M. Brown and S. J. Rogers, “User identification via keystroke characteristics of typed names using neural networks,” Int. J. Man-Mach. Stud., vol. 39, no. 6, pp. 999–1014, Dec. 1993.

[3]. P.Dowland, S.Furnell, and M.Papadaki, “Keystroke analysis as a method of advanced user authentication and response,” in Proc. IFIP TC11 17th Int. Conf. Inform. Security: Visions Persp., May 7–9, 2002, pp. 215–226.

[4]. D. Gunetti and C. Picardi, “Keystroke analysis of free text,” ACM Trans. Inform. Syst. Security, vol. 8, no. 3, pp. 312–347, Aug. 2005.

[5]. F. Monrose and A. Rubin, “Authentication via keystroke dynamics,” in Proc. Fourth ACM Conf. Comput. Commun. Security, pp. 48–56, Apr. 1997.

[6]. D. Polemi. “Biometric techniques: Review and evaluation of biometric techniques for identification and authentication, including an appraisal of the areas where they are most applicable,”

[7]. S. Ross, “Peirce’s criterion for the elimination of suspect experimental data,” J. Eng. Technol., vol. 20, no. 2, pp. 38–41, Oct. 2003.

[8]. M. Villani, C.Tappert, N. Giang, J. Simone, St. H. Fort, and S.-H. Cha, “Keystroke biometric recognition studies on long-text input under ideal and application-oriented conditions,” in Proc. 2006 Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW’06), June 2006, p. 39.

[9]. L. Ballard, D. Lopresti, and F. Monrose, “Forgery quality and its implication for behavioral biometrics security,” IEEE Trans. Syst. Man Cybernet., Part B, vol. 37, no. 5, pp. 1107–1118, Oct. 2007.