Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 10, October 2015, pg.239 – 245

RESEARCH ARTICLE

Multi-Level User Authentication via Keystroke Dynamics

Abhaysingh Saste [#1], Mangesh Bedekar [*2], Pranali Kosamkar [#3]

#Computer Department, Pune University

[1] abhaysingh.saste@gmail.com

[2] mangesh.bedekar@mitpune.edu.in

[3] pranali.kosamkar@mitpune.edu.in

Abstract— Behaviour biometric can provide a better alternative for traditional password methods and also for physiological biometrics due to the fact that they are very economical and easy to integrate with existing system. Various approaches have been made towards enhancing and improving the performance of keystroke dynamics system. The aim of writing this paper is to deal with all the possible uncertainties which lead to degrade the overall system accuracy and performance. We are proposing a multi-level authentication system which has identification module, deviation smoothening module, continuous verification and alternate authorization module that all will make the system more convenient, secure and accurate.

Keywords— Keystroke dynamics, enrolment, keystroke template, diagraph, adaptive learning.

I. INTRODUCTION

As the dependency of human being is increasing on computer systems, it becomes necessary to guarantee the user's authentication. The User ID and password is the most commonly used method of authentication. However the users are required to remember the long complex passwords and they are not very much confidential all the time, so it becomes necessary to improve the authentication mechanism. A token also provides a good solution but once the token is lost, it becomes major problem for the owner [1].

Biometrics is proved to be an effective alternative for the user authentication. Biometric features are unique to each individual such as a fingerprint, eye scan, voiceprint, or Signature. However, if a biometric is compromised or a document is lost, they are not as easily replaceable as passwords or tokens. Biometrics are further categorized into physiological and behavioral features. In general, Physiological features have been more successful than behavioral features to implement authentication systems [15]. This is because the stability of data over period of time in physiological features is more. The physiological features essentially do not vary along time, whereas behavioral features such as signature and keystroke dynamics may change greatly even between two consecutive samplings. On the other hand, the physiological biometric techniques require specific tools to sample the corresponding biometric feature. This increases the overall cost of implementing the system. Another drawback comes when implementing the system on remote connections [4].

Keystroke dynamics is part of behavior biometrics in which users are uniquely identified based on their typing rhythm. The keystroke dynamics system measures typing rhythm of the user and develop a unique keystroke template for future authentication. The duration for which the key remain pressed (Dwell time) and the time between two successive key presses (Flight time) is collected for the user session. The recorded keystroke timing data is then processed to obtain unique signature which will be used for future comparisons to determine the user's identity.

Keystroke technology can be easily integrated with existing technology environments. The technology is also designed to be scalable to operate across the Internet and throughout the enterprise.

II. RELATED WORK

The reason behind increase in the research work in keystroke dynamics is due to the fact that it is very economical and can be integrated with existing security system easily. User authentication can be done only once during initial login (static) or can be carried out during entire user session (continuous) [1].

Keystroke features are based on time durations between the keystrokes, time for which the key remain pressed, time between two successive key press, overall typing speed, frequency of errors (use of backspace), use of num-pad, order in which user presses shift key to get capital letters and possibly the force with which keys are hit for specially equipped keyboards [2].

The key-presses are captured in the form of monographs (single word), digraphs (pair of words) ….n-graphs. The average key duration and latency time is considered for preparing users signature. Any new user who wants to enter into system, his/her data is compared with the existing profile. According to Gunetti and Picardi, if the mean distance between two samples is less, they are assumed to be belonged from the same user (GP method). They achieved False Alarm Rate of about 4% and an Impostor Pass Rate of less than 0.01% as their experimental results [4]. As a part of modification to GP method, k-nearest neighbor is used to classify user profiles [6]. Experiment have demonstrated the same level of FAR and FRR as that of GP method and as high as 66.7% improvement of the authentication speed has been achieved. The major difference between the Clustering Based Keystroke Authentication Algorithm and the GP method is that the verification process of the CKAA is within a cluster while the GP method needs to go through the entire database.

Various other distance based classifiers mainly Euclidean, Weighted probabilistic, non-weighted approaches [13] are compared for their performances when used during identification process for comparing two keystroke templates but Bayesian Frameworks are found superior when identification rate is considered [5].

Mahalanobis distance is susceptible to the outliers that are abundant in keystroke dynamics data due to the frequent pauses during typing. On the other hand, Manhattan distance is shown to be more robust to outliers but it is not able to correct for the adverse interactions and redundancies between keystroke features. The New Distance Metric is formed combining the advantages of both [9]. Apply the principle of Mahalanobis distance to decorrelate and normalize the keystroke dynamics feature variables. Once the data are normalized and decoupled, then compute the Manhattan distance between two data points. This new distance metric ensures not only that the undesirable correlation and scale variations are accounted for, but also suppress the influence of outliers for improved performance.

As compared to static keystroke dynamics, less work has been done on free-text (continuous) keystroke dynamics. In [3], entire process is divided as enrollment and identification. The Neural networks are used to model the user behavior based on the encoded sets of monographs and digraphs. Each time when user sign-up in the system, user signature is computed based on the typing rhythm. During verification, it is found that high deviation of the session points occurs from the non-self-reference signature and the closeness to the self-reference signature for the two users. Free text analysis with monograph and diagraph analysis is carried out with the possible algorithms including Neural Network, Support Vector machine algorithms, Genetic algorithms etc. [8].

Instead of using any single method, an approach is made to uses fuzzy logic, neural networks, statistical techniques, and the combination of these approaches to learn the typing behavior of a user [11]. The combination of all these techniques improves the overall performance.

Similar kind of work has been carried out in enhancing the authentication process [12]. If the keystroke template does not match with the user profile, alternate email verification is proposed in there system. Along with alternate email verification module, there is updating module which keeps the track of user's current behavior and updates the profile accordingly.

This all system guarantees the user's legitimacy only during entry point in to the system. There is possibility that malfunctions can occur during the session. So there is a need to continuously monitor the user during its entire session.

III. PROPOSED SYSTEM

Overall system can be broadly categorized in to two parts, enrolment and identification phase [16]. If the user is new to the system, all the necessary information along with the keystroke pattern is captured and unique user profile is formulated. Whenever the user logins next time, the claimed identity is compared with the existing profile and decision about the access is carried out accordingly. Addition to these two modules there are several other modules which perform their respective functionalities which are elaborated in below section.

System Modules -

A. Enrollment module
B. Identification module/ Matching module
C. Adaptive learning
D. Additional verification
E. Continuous verification

A. ENROLLMENT PHASE –

When the user is interacting with the system for the first time, all the required credentials are stored as a user profile in the database. User information includes the username, password, security question and its answer. When user is typing all this data, his typing behaviour is captured, processed and stored as a reference keystroke template for future comparisons during identification process. The format of user profile is shown in Table 1.

User id	*System generated*
User name	*User specified*
Password	*User specified*
Security question/answer	*User specified*
Keystroke template	*System generated*

Table.1 Format of user profile

a. Parameters used-

Keyboard is the only hardware required to implement the keystroke dynamics authentication system which is generally available with all computers. To analyse the user behaviour, his typing patterns are captured and analysed. While doing these following are the main parameters that are considered-

1. **Press-press time (PP)** - Time between first key press and second key press.
2. **Press-release time (PR)** - Time between first key press and same key release (hold duration).
3. **Release-press time (RP)** - Time between fist key release and second key press.
4. **Release-release time (RR)** -Time between first key release and second key release.

And for evaluating the system performance following parameters are used-

- **False Rejection Rate (FRR)** - The percentage of times that a valid user is labeled as an adversary and denied access; also known as the false negatives rate.
- **False Acceptance Rate (FAR)** - The percentage of times that an adversary gains access as a user; also known as the false positive rate [10].

b. Keystroke template formulation-

The above mentioned parameter values are generated while monitoring the user's typing pattern. These values are processed to get the keystroke template. The format of template that is stored in database is-

No. of samples	Mean	Standard Deviation

Table 2. Format of keystroke template

When user is typing on the keyboard all the values of parameters (PP, PR, RP, RR time) are also calculated. If t1,t2…tn are these timing data, the mean μ is calculated as

$$\mu = \frac{1}{n}\sum_{i=1}^{n} ti$$

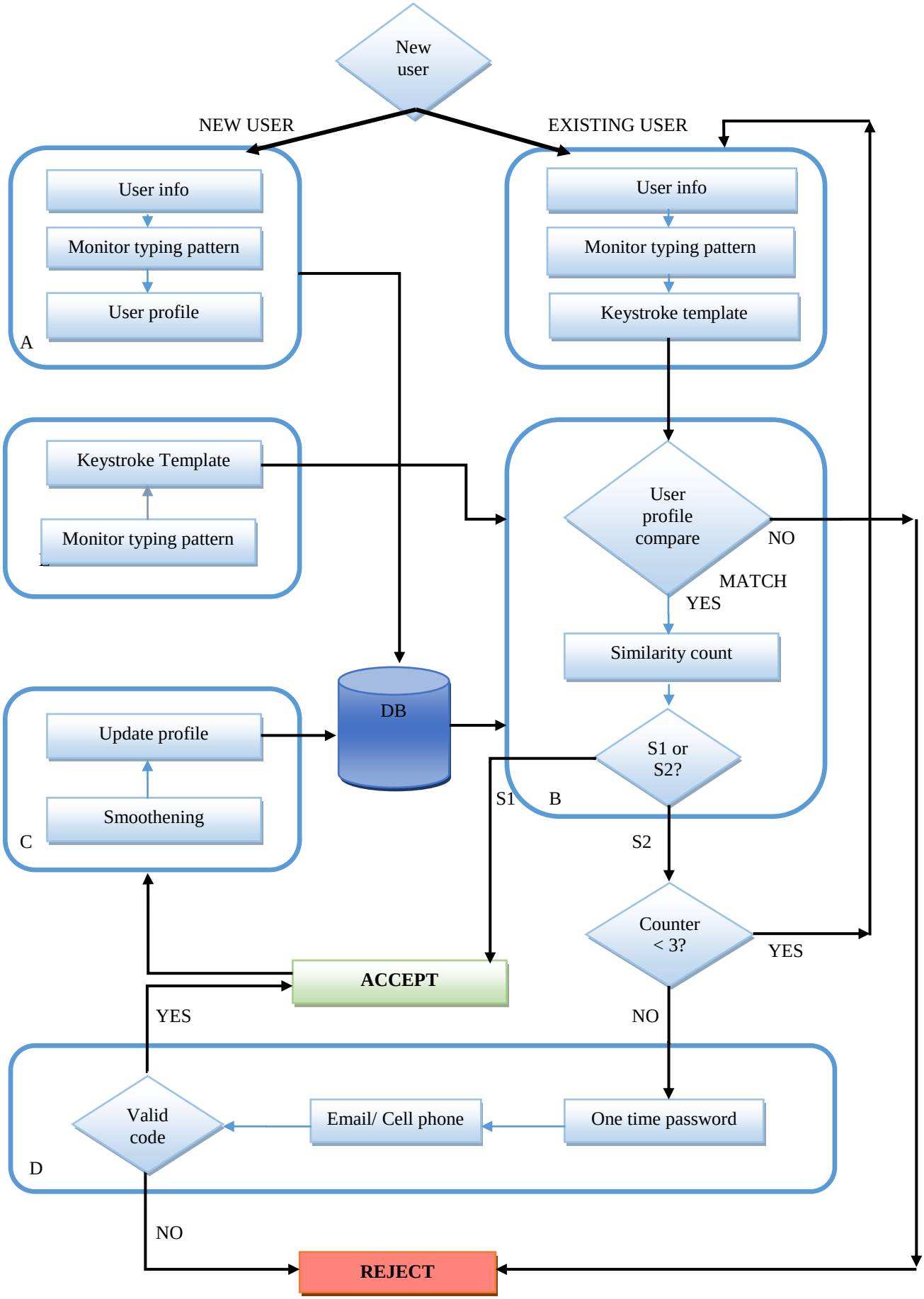

Fig. 2 Flow chart of proposed system

Similarly the Standard Deviation σ is calculated as

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(ti-\mu)^2}{n}}$$

B. IDENTIFICATION MODULE

When the user registers into the system for first time, user profile is created and stored in the database. Identification module compares the new credentials with the reference profile from the database. If the match found system allows the access otherwise denies the access.

i. MATCHING MODULE-

The main task of matching module is to compare the claimed data with existing data and take appropriate decision. Username, password and security questions can be directly compared with the data available in the respective user profile. For comparing the keystroke template we require an algorithm which finds the similarity count between the new template and reference template.

Gaussian probability density function (GPD)- This is an algorithm which is used to compare two templates. It returns a similarity count between which is between [01] indicating the similarity found between two templates which are under testing.

$$S_{GPD} = \frac{\sum_{I=1}^{K} e^{-\left(\frac{(ti-\mu i)^2}{2\sigma_i^2}\right)}}{k}$$

Where t_i indicates the timing data, k indicates the number of parameters considered, μ is the mean and σ is the standard deviation. Threshold value is set for the similarity count. If similarity is found above 80% (S1), accept the user else if similarity count is less than that then re-initiate the identification process invoking the security question.

Following is the complete flow of activities happening in the matching module -

1. Compare the Username and Password. If does not match, deny the access.
2. If match found, compare the keystroke template.
3. If deviation is d1, accept the user and grant the access.
4. If the access is granted, implicitly invoke the adaptive learning module in background.
5. If deviation is d2, re-invoke the identification process with security question.
6. In counter reaches to 3, One-time password is generated and is send to Email/Cell phone.
7. If code is valid access is granted. Otherwise the access is denied.

C. ADAPTIVE LEARNING

Considering the changing behaviour of the user, it is necessary to maintain the copy of keystroke template in the database that reflects the user's recent behaviour. Along with this age is also one factor due to which the human behaviour is keeps on changing. So the keystroke template needs to be updated to accommodate all the changes. So the values of mean and standard deviation are recalculated and updated in the existing profile. This module is only invoked when the user successfully logins in to the system.

From both the keystroke template, the claimed and the reference the mean and standard deviation are recalculated as follows-

$$\mu_u = \frac{\sum_{i=1}^{n} ti + tn_u}{n_u}$$

Where t_i is timing data of old template and t_u is timing data of claimed template. n_u is the total number of samples. This gives a new mean value μ_u. Similarly standard deviation is calculated using the new mean value.

$$\sigma_u = \sqrt{\frac{\sum_{i=1}^{n_u} t_i^2}{n_u} - \mu_u^2}$$

This module should work in background to maintain the low overhead. So whenever the template will be updated, it ensures the most recent behaviour of the user which helps to reduce the False Rejection Rate.

D. ADDITIONAL VERIFICATION-

If user fails to prove his identity there is additional mechanism by which the user can enter into the system. One-time password is generated and sent to registered email id/ mobile number. If that code matches then user gets access into the system.

E. Continues verification -

If the user's legitimacy is verified only at the time of entry, it is possible that account can be hacked or misused by other person. Our system should be responsible for providing authentication during entire user session. This can be done by monitoring the further user session and gather more samples. If the samples collected are less, an approximation method can be applied and regenerate the missing samples. Then the same approach is used to match the keystroke templates. This complete process occurs at background without user's intervention. If the user is not found to be a genuine, then users session is interrupted.

Benefits of our system:

- **Overcoming traditional password drawbacks-**
  - Password length/difficulty- Passwords are generally consists of alphanumeric characters, special symbols etc. to increase the difficulty level. It becomes hard to remember such complex passwords.
  - Hesitation in public/ spy camera- There is chances that your password can be trapped in public places using spy camera. But our system requires more than just a password to authenticate.
  - Password recovery- It becomes very convenient to recover the forgotten password just by identifying the typing pattern.
- **Physical injuries-** we are dealing with the typing rhythm of the individual but what if hands are injured? System must be implemented considering all such possibilities. Since typing speed is going to get change due to injured hands, the keystroke template matching is surely going to fail. To authenticate the legitimate user, our system includes email/ phone based one-time generated code for which can be used for login if multiple tries are failed.
- **Changing human behavior**- One of the major drawback of keystroke dynamics is the changing behavior of human as compared to physiological data (ex. Fingerprint) which is constant over period of time. Considering this, the system must accommodate the changes and reflect the keystroke template based on the current behavior of user. Our system has Adaptive Learning Unit which is responsible for updating the user profile to reflect the changes.
- **Dependency on mood and emotions**- Typing speed is greatly influenced by the emotional state of the user while typing. Moreover, lots of distractions are possible. The system is flexible enough so that users are not rejected if keystroke template is not matched. Security questions are provided so for little deviation system should allow the user after answering the question.
- **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.

IV. CONCLUSIONS

The entire authentication system is divided into several modules each is responsible for improving the overall system performance. Providing keystroke based verification in addition with the password, helps to reduce the False Acceptance Rate (FAR). The security question and answer makes the system more flexible to access in case of unmatched keystroke behaviour. Adaptive learning module takes care of the changing nature of human and reflects the most recent behaviour during matching. This helps to reduce the False Rejection Rate (FRR). If the user is unable to generate the matching keystroke template due to some psychological or physical problems, alternate authentication can be done using One-time passwords generated on Emails/Cell phones. The user's session can be later monitored to maintain the authenticity during the entire session. So this paper is an approach

made towards providing a more reliable and complete solution for implementing user authentication in public environment using keystroke dynamics.

ACKNOWLEDGEMENT

We are extremely grateful to those who have helped, contributed and supported us during the project. Our deepest thanks to our advisor Prof. M. Bedekar for his continuous encouragement throughout the research. It was our pleasure to work under his guidance.

REFERENCES

[1] Pin ShenTeh, Andrew Beng Jin Teoh, ShigangYue, "A Survey of Keystroke Dynamics Biometrics",Hindawi Publishing Corporation, The ScientificWorld Journal Volume 2013, Article ID 408280, 24 pages.

[2] Roman V. Yampolskiy ,VenuGovindaraju , "Behavioural biometrics a survey and classification", Int. J. Biometrics, Vol. 1, No. 1, 2008.

[3] Ahmed A. Ahmed and IssaTraore, "Biometric Recognition Based on Free-Text", IEEE TRANSACTIONS ON CYBERNETICS, VOL. 44, NO. 4, APRIL 2014.

[4] Francesco Bergadano Daniele Gunetti Claudia Picardi, "User authentication through keystroke dynamics", ACM Transactions on Information and System Security (TISSEC), Volume 5 Issue 4, November 2002 Pages 367-397.

[5] Fabian Monrose, Aviel D. Rubin "Keystroke Dynamics as a Biometric for Authentication", Future Generation Computer Systems, Volume 16, Issue 4, February 2000, Pages 351–359.

[6] J.Hu, D. Gingrich, A. Sentosa, "A k-Nearest Neighbor Approach for User Authentication through Biometric Keystroke Dynamics", Communications, 2008. ICC '08. IEEE International Conference on May 2008, ISBN: 978-1-4244-2075-9, Pages 1556 - 1560.

[7] Fabian Monrose, AvielRubin,"Authentication via Keystroke Dynamics", 4th ACM conference on Computer and communications security 1997, ISBN:0-89791-912-2, Pages 48 – 56.

[8] M.Kanimozhi, M.Savitha, KavyaPuvirajasingam,"Keystroke analysis as a method of biometric free text recognition for user authentication", Taraksh Journal of Information Systems 2014 / Page No. 32 / Volume 1 Issue 1.

[9] Yu Zhong, Yunbin Deng, Anil K. Jain "Keystroke Dynamics for User Authentication", IEEE Computer Society Conference on June 2012, ISSN:2160-7508, Pages: 117 – 123.

[10] Edmond Lau, Xia Liu, Chen Xiao, and Xiao Yu "Enhanced User Authentication ThroughKeystroke Biometrics", Computer and Network Security Final Project Report Massachusetts Institute of Technology, December 9, 2004.

[11] S. Haider, A. Abbas, and A. K. Zaidi. "A multi-technique approach for user identification through keystroke dynamics", IEEE Int'l Conf. on Systems, Man and Cybernetics, pp. 1336–1341, 2000.

[12] Pin Shen Teh, Andrew Beng Jin Teoh, Connie Tee, Thian Song Ong "Keystroke dynamics in password authentication enhancement", Expert Systems with Applications 37 (2010) 8618–8627.

[13] Soumen Roy, Utpal Roy, D. D. Sinha, "Enhanced Knowledge-Based User Authentication Technique Via Keystroke Dynamics", 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.

[14] Lawrence O'GormanAvaya Labs, Basking Ridge, "Comparing Passwords, Tokens, and Biometricsfor User Authentication", Proceedings of the IEEE, Vol. 91, No. 12, Dec. 2003, pp. 2019-2040.

[15] M. Karnana, M. Akilab, N. Krishnaraj' "Biometric personal authentication using keystroke dynamics: A review", Applied Soft Computing 11 (2011) 1565–1573.

[16] Lívia C. F. Araújo, Luiz H. R. Sucupira Jr., Miguel G. Lizárraga, "User Authentication Through Typing Biometrics Features", IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 53, NO. 2, FEBRUARY 2005.

